

**The Moral Status of Artificial Intelligence:
A Jewish Ethical Perspective**

Mois Navon

Department of Jewish Philosophy

Ph.D. Thesis

Submitted to the Senate of Bar-Ilan University

Ramat Gan, Israel

October 2023

This work was carried out under the supervision of

Prof. Hanoch Ben Pazi

Department of Jewish Philosophy
Bar-Ilan University.

Acknowledgements

Without a doubt technology is entering a bold new age, an age referred to as “the fourth industrial revolution,” it is the age of artificial intelligence (AI).¹ While this revolution will advance the world in many positive ways, it will also bring to bear many ethical questions.

Thus began my thesis proposal in February 2019. At the time, people still asked me, “what is AI?” Today, only a mere four years later, ChatGPT has taken the world by storm and not only has practically everyone become aware of AI, practically everyone has become aware of the ethical dilemmas engendered by AI.

Now, I must admit that while I have long been aware of AI (spending my entire professional career designing computer systems), I was not aware of the direction it was taking. Indeed, the idea for this thesis came about when someone asked me, “So, when do you think AI will become sentient?” I thought the question was, well, ridiculous. Nevertheless, I went home and did a quick Google search to find that over a million sites did not think the question was ridiculous at all. And thus my thesis was born.

The technical question, “When will AI become sentient?” led me to the ethical question, “What is the moral status of AI?” And this entirely new ethical question, “What is machine?”, led me to the oldest existentialist question, “What is man?” My thesis, then, encompasses the two great fields of philosophy – existentialism and ethics – fields that have intrigued me ever since the day when I was about seven years old and discovered that people die. It was then that I asked myself, “Why are we here?” and “How are we to live?”

To address these questions within the broader context of AI ethics, this thesis is truly interdisciplinary, incorporating an extremely diverse spectrum of fields, including computer science, biology, bioethics, moral philosophy, philosophy of mind, philosophy of science, philosophy of technology, Jewish philosophy, Halacha, Kabbalah, and more. Accordingly, this thesis represents the fruits of my labor, not only of years specifically

¹ Schwab 2017.

researching the subject of AI ethics, but a lifetime of learning. So, while I would like to acknowledge all those who have had an influence on my thinking, that would mean listing every teacher I have had since kindergarten. Therefore, I will have to limit myself to the following.

First and foremost, I must acknowledge the Creator of the universe, whose guiding hand has graciously brought me to fulfill this dream of writing a doctoral thesis in the field of Jewish ethics and philosophy. I thank my mother, Becky Navon-Gellmann, who always made it clear that both Jewish values and education were a priority. I thank my early teachers in Jewish thought: Ray Eskenazi, R. Yitzchok Adlerstein, Dr. Steve Bailey, and *lehavdil bein hayim lehayim*, R. Shlomo Schwartz, z"l. I thank my semicha rav, R. Murray Goldenhirsch, and my teachers in Jewish philosophy: Prof. Dov Schwartz, R. Dr. Tzachi Herskovitz, Dr. Yehuda Halperin. Special thanks to my Golem teacher, Dr. Naama Ben Shachar; my Philosophy of Mind teacher, Prof. Alon Chasid; my Moral Philosophy teacher, Dr. Efrat Tiktin; and my Philosophy of Technology teacher, Dr. Galit Wellner.

All of this learning would not have resulted in this thesis if not for my thesis advisor, Prof. Hanoch Ben Pazi, who is also due no small thanks for being my Jewish Philosophy teacher throughout the years of my Masters and Doctorate degrees. Worthy of note here is how, at the very beginning of my work, I expressed certain opinions about how I thought my thesis would conclude. Prof. Ben Pazi just smiled. His smile was, of course, prescient as I changed my opinions radically over the course of my research. Thanks is thus due to his openness, experience and good nature.

And last, but certainly not least, to my wife Deena, who not only pushed me to follow this dream, but supported me throughout and sacrificed of herself in no small part so that I could achieve it.

Mine and yours are hers.

Technical Preface

This dissertation combines six separate chapters previously published as standalone essays in various journals. Together, they provide a comprehensive response to the ethical dilemmas that this thesis seeks to address. The interdependency of the essays is explained in the Introduction and Overview sections that follow, as well as in the final Conclusion section. To ensure that the text reads as a coherent whole, the six essays have been uniformly formatted (i.e., modified from their original publication format). In addition, references from one essay to another have been modified to refer to the corresponding internal chapter rather than to the original external publication.

Table of Contents

Abstract.....	i
The Dilemmas.....	ii
The Paradigms.....	iii
1. Introduction.....	1
1.1 The Mindless Robot Dilemma.....	1
1.2 The Mindful Robot Dilemma (1).....	4
1.3 The Mindful Robot Dilemma (2).....	5
2. Thesis Structure.....	0
3. The Thesis.....	5
Ch. 1: Article #1 - Finding Virtue in a Law on Slaves.....	6
Ch. 2: Article #2 - The Virtuous Servant Owner.....	41
Ch. 3: Article #3 - To Make a Mind.....	72
Ch. 4: Article #4 - Polemics on Perfection.....	92
Ch. 5: Article #5 - Eudemonia of a Machine.....	115
Ch. 6: Article #6 - “Let Us Make Man in Our Image”.....	138
4. Conclusion.....	164
5. Bibliography.....	169
6. Appendix: The Moral Status of a Soul.....	207
7. Appendix: Other Ethical Issues Inherent in AI.....	209
Hebrew Abstract.....	⌘

Abstract

The past decade, known in technology circles as “the third AI summer,”² has witnessed the increasing power and proliferation of artificial intelligence (AI) throughout society, powering technological applications that were, in the not-so-distant past, considered the stuff of science fiction. AI is driving everything from vacuum cleaners to automobiles, smart phones to smart bombs, virtual assistants to virtual reality. And while the list of AI-based innovations goes on and on, it is the Social Robot that is to my mind the most interesting, if not the most perplexing. For, while Social Robots, like just about every technology since the plow,³ are designed to relieve humanity of its burdens, they bring with them moral dilemmas radically different from every other technology since the plow. And this because, a robot that looks like us, acts like us and speaks like us, also challenges us in ways never before encountered. For the first time in the history of humanity we are to confront a being with the capacities of “intelligence and speech” (*de’ab v’dibbur*);⁴ capacities that have, until now, been the exclusive province of that pinnacle of creation: the human being.

Not all robots, however, are created equal and, correspondingly, neither are the dilemmas they engender. On the one hand, we have the robots of today, which are driven by what is known as “weak AI,” or technology that employs artificial neural networks (ANNs) to analyze, decide, and perform tasks based purely on mathematics.⁵ Such robots’ cognitive processing is purely functional, lacking the subjective understanding attendant in human-level consciousness. They are what I call “mindless” robots – essentially highly sophisticated computing systems in human-like form. On the other hand, we have the robots of tomorrow, to be based on “strong AI,” or technology that seeks to engender

² See, e.g., Kautz 2022.

³ See Burke 1978; also commentaries on Gen. 5:29 (Rashi, Radak, et al.).

⁴ Rashi (Gen. 2:7).

⁵ Domingos 2018; Boucher 2019; Brand 2020: 207; Coeckelbergh 2020a: 83-94.

and employ human-level consciousness to analyze, decide and perform its tasks.⁶ Such robots, with cognitive faculties on par with those of human beings, complete with human-level consciousness, can be said to have a “mind.” They are, then, “mindful” robots – human-like in the deepest sense, but not of the human species. Each of these humanoid robots – the mindless and the mindful – poses its own unique moral dilemmas and, accordingly, requires its own unique moral approach.

Worthy of mention is that, while both types of AI can be implemented, theoretically, in a myriad of physical forms (e.g., watches, phones, laptops, servers), what is of particular interest in this thesis is their embodiment in human-like form, colloquially known as “robots.” For it is when a computer system looks, acts, and speaks like a human that we humans become confused, enamored, and ultimately engaged with them as if they were “human.” That said, any embodiment that engages us as human-like – be it as simple as ChatGPT, Siri, or Alexa, or as sophisticated as the virtual girlfriend “Caryn.AI” – is of relevance to the analysis herein.

Interestingly, it is precisely these latter incarnations of AI (e.g., ChatGPT) that have given greater urgency to this work. For, while my initial focus was on social robots (SRs), the proliferation of LLMs (Large Language Models), like ChatGPT, has preceded the rise of SRs. Consequently, the practical application of my proposed ethical approaches to the SRs that are sure to come,⁷ can already be applied to the LLMs that are already here.

The Dilemmas

Now, as said, the ethical dilemmas inherent in the mindless robot are entirely different from the ethical dilemmas inherent in the mindful robot. Whereas the mindless robot engages us as human-like while it is not, the mindful robot engages us, in a sense, as machine-like while it is not. Accordingly, as will now be explained, the focus of the

⁶ While I describe the position that Artificial General Intelligence (AGI) requires human-level consciousness, and thus, can only be achieved with “Strong AI,” it should be noted that some believe human-level intelligence is not contingent on *consciousness*, and thus, AGI can be achieved with “Weak AI.”

⁷ Some of the more popular robots in the works are: Atlas (Boston Dynamics), Optimus (Tesla), Ameca (Engineered Arts), Sophia (Hanson Robotics). For more see Biba 2023.

mindless robot dilemmas is *us* humans, whereas the focus of the mindful robot dilemmas is *them* humanoids.

To be specific, while all technologies give rise to numerous dilemmas, there is one prominent dilemma presented by the mindless robot, what I call the “Virtue Authenticity Dialectic” (VAD).⁸ That is, given that a social robot looks, talks, and acts like a genuine human being, we naturally respond and interact with it as such. It behooves us, then, to relate to it in the same virtuous manner we are to strive for when interacting with all human beings, lest we come to treat our fellow humans as machines. That said, by engaging with the robot as if it were a living, conscious being, we lose our appreciation for authentic human relationships. We become trapped in a perceptual mode of relationships in which the robot is perceived *as if* it were a live, conscious being. The humanoid robot is then seen as a suitable, even preferable, substitute for humans who notoriously entail far more complicated interrelations. Our virtuous relationships with machines can thus result in the demise of our authentic relationships with humans, leading to the unraveling of the very fabric of human society.

In contradistinction to the mindless robot that appears to be conscious, but is not, the mindful robot appears to be a machine in the sense of having been created to serve our needs, but is actually conscious. Such a being gives rise to two primary dilemmas centered around its very creation: (1) is it ethically permissible to create a conscious being whose sole purpose in “life” is to serve us? Is this not slavery? And (2) if we do, accordingly, reject the creation of mindful machines to do our bidding, what about creating them to be our children, our friends, our life partners? Is there anything ethically untoward in the creation of synthetic conscious beings? Would human society be better or worse in welcoming non-human humanoids into the family?

The Paradigms

While not a few modern thinkers have grappled with these dilemmas using various moral approaches, it is my thesis that Jewish thought affords us ancient yet profound paradigms

⁸ There is another closely related, but different, dilemma called the “Anthropomorphism Dehumanization Paradox” or “Dilemma of Empathy and Ownership” (see herein Ch. 2 “The Virtuous Servant Owner”).

from which to formulate meaningful moral responses to these ultramodern dilemmas. These paradigms, anchored as they are in Jewish tradition, provide unique philosophical, ethical and legal perspectives. To begin, Jewish thought looks at the world as purposive – as having a beginning, an end and a goal to be achieved – and this informs what it means to live “a good life.” Furthermore, such a purposive good life both entails and is defined by the ethical values of a 3300-year-old tradition kept alive by a people that, according to U.S. President John Adams, “has done more to civilize men than any other nation.”⁹ Finally, those ethical values are not merely suggestions for good conduct but are codified as legal norms (i.e., halacha), thus adding the element of obligation to said ethical conduct. Accordingly, Jewish ethical paradigms are charged with telos and bolstered by obligation. And while these paradigms, and the ethical approaches they engender, are firmly grounded in Jewish thought, they transcend religious boundaries and are presented here in a way that all peoples can appreciate and apply.

The first paradigm to be employed is what I call “the Virtuous Servant Owner,” derived from my reading of Maimonides’ last Law on Slaves (9:8). At first glance, the text appears to be a random collection of quotes and anecdotes brought to encourage merciful interaction with one’s slave. A closer reading, however, reveals a deliberate and decisive logical proof that both defines and demands nothing less than complete moral perfection. I apply this paradigm as an ethical approach to the VAD dilemma engendered by the mindless robot (as well as other mindless AI entities, e.g., LLMs and virtual people). Whereas some have opted to “resolve” the dilemma by opting for virtue at the cost of authenticity, and others for authenticity at the cost of virtue, I argue that the Virtuous Servant Owner can maintain both his virtue and his authenticity.

The second paradigm I employ is what could be called the “Jeremiah Glory of Man” paradigm, which, as explained by Maimonides, tells of the *summum bonum* of man. The prophet Jeremiah proclaims, “Thus saith the Lord: ... let him that glorieth glory in this, that he understandeth, and knoweth Me, that I am the Lord who exercise mercy, justice, and righteousness, in the earth ...” From this, Maimonides explains that the Glory of Man, the *summum bonum* to which everyone is to strive, is composed of knowing God (i.e., developing one’s intellect) and emulating His ways (i.e., developing one’s ethical

⁹ Adams 1809.

dispositions). Thus, one must have the freedom to develop, to strive to fulfill one's *summum bonum*. And while it must be admitted, as Maimonides does, that this telos was held by "the ancient wise people" (e.g., Aristotle), it was articulated long before them by the prophet Jeremiah. Accordingly, I apply this "Jeremiah Glory of Man" paradigm to address the moral propriety of programming a mindful robot to relieve us of our burdens, demonstrating that such programming would render that being's life meaningless and thus constitute an ethical violation of the highest degree. And though the dilemma has been discussed using secular moral approaches, this paradigm offers insights that even those with secular perspectives can appreciate.

Finally, the third paradigm I employ is that of "the Golem," a synthetic humanoid conjured from the dust of the earth. The Golem appears in Jewish literature throughout history and is variously referred to as (a) a being animated by a "vitality" – i.e., a power to allow for mobility but no phenomenal experience; (b) a being animated by an animal soul – i.e., having sentience; or (c) a being with human-level consciousness. It is this last Golem, of course, that is of the greatest interest in addressing the moral propriety of creating a mindful robot. That said, while I do employ it in the discussion, of greater interest is the very first Golem to appear in Jewish literature (San. 65b) referred to as a "Gavra" (humanoid). This Gavra, though it was a category (b) creature (i.e., lacking human-level consciousness) it was, nevertheless, created with the intent to have human-level consciousness. Through my comprehensive analysis of the Gavra narrative, I demonstrate that the Talmud is categorically opposed to the creation of a humanoid with any level of consciousness. Furthermore, I argue that this position is of both ethical and legal import; indeed, it is halachically forbidden to create mindful robots.

1. Introduction

My thesis, “The Moral Status of Artificial Intelligence,” seeks to address, as described in the abstract, two types of artificial intelligence – i.e., mindless and mindful – that engender three prominent dilemmas to which I apply three Jewish paradigms. In the following three sections, I provide an overview of my application of these paradigms to the dilemmas.

1.1 The Mindless Robot Dilemma

As noted in the abstract, the mindless robot engages us *as if* it is human and thus draws us to respond *as if*, indeed, it is human. How can we maintain our humanity, our virtuous behavior toward humans, if we act toward this seeming human as a machine? Not being virtuous toward humanoids puts us in jeopardy of not being virtuous toward humans. Conversely, how can we maintain our appreciation for authenticity if we act toward a machine as if it were human? By acting virtuously with mindless humanoids, we run the risk of losing our sensitivity to the infinite value of conscious human beings. This is the Virtue Authenticity Dialectic (VAD). It is one that poses a dilemma that transcends the individual, as it fundamentally concerns all of our relationships and interactions, thus presenting a profound challenge not only for individuals but for society as a whole.

In response, I propose the Virtuous Servant Owner paradigm described in Maimonides’ last Law on Slaves.¹⁰ This law teaches, in profound yet concise detail, how to behave in a

¹⁰ This approach finds halachic support in the comments of R. Yosef Engel (Krakow, 1859-1919) in his glosses to the Gemara (Gilonei Hashas, San. 19b, s.v. *sham maale*). There he writes that a synthetically created being (i.e., via Sefer Yetzirah) should be treated as a “Canaanite Slave.” Now, while he does not distinguish between the three types of Golems (i.e., those with mere motor “vitality,” versus those with animal sentience, versus those with human consciousness – see Chapter 6), we will apply his paradigm to the mindless machine and argue later that it should not apply to conscious machines. It is acknowledged that the Golem paradigm is not a “perfect fit” for the mindless machine in that the Golem (and presumably the Canaanite slave) did not have the intelligence of today’s AI, and conversely, today’s AI does not have the feelings of a Canaanite slave. Nevertheless, they both engage us human-like and

virtuous way with one's servant. And while many ethicists call for virtuous behavior with mindless robots, none provide a method to simultaneously maintain a distance, an awareness, that these servants are not to be engaged with for more than the work they are to perform. By applying the laws of slaves, we are immediately awakened to the fact that this being is, well, our slave – a being that is in our midst to perform our labors and not to be engaged with beyond that.

Maimonides' last Law on Slaves, then, serves as a kind of user's guide, an ethical addendum to a technical instruction manual for the mindless robot. To better understand this proposition, the in-depth analysis of Maimonides' ruling (as provided in Chapter 1), is essential to highlight the precise basis for our virtuous behavior with the mindless robot. Maimonides goes to great lengths, not only to make and support his claims for virtuous relations, but to address them to any and all – regardless of which moral stage “on life's way” one finds oneself: aesthetic, ethical, or religious.¹¹ And it is here, in the distinction of the stages, that we discern a distinction in the basis of virtuous behavior.

Beginning with the “ethical” stage, Maimonides' arguments (what I refer to in Chapter 1 as claims (2) and (3)) are based on our biological and spiritual likeness with our slave. His claims delineate ethical interactions that follow Maslow's pyramid of *human* needs and are thus clearly inapplicable to a mindless robot made of silicon.¹² Accordingly, there is no place for the “ethical” – i.e., virtuous interactions based on the “humanity” of the other – in our interactions with a mindless robot. To do so would be fallacious, foolish, and – if we read this as implicit in Maimonides' legal ruling – forbidden.

confront us in an “uncanny” relationship. It is for such a relationship that Maimonides' law, as will be explained in this introductory text, provides ethical guidance.

¹¹ As will be elaborated in Chapter 4, these are the three stages of moral development according to Kierkegaard and according to which I delineate Maimonides' claims.

¹² It is important to note here, as I do in Chapter 1 (fn. 59) itself, that the frameworks provided by Kierkegaard and Maslow are employed as hermeneutical tools to help structure Maimonides' law and do not imply anything about the interrelatedness of the thought of these three thinkers.

That said, there is a strong basis for virtuous behavior with a mindless robot to be found in the “religious” stage, where “religious” means ethical as demanded by divine authority.¹³ It is here that Maimonides (referred to in Chapter 1 as claims (4) and (5)) grounds his claims not on the intrinsic nature of the slave, but on the intrinsic value of moral virtue. Accordingly, “religious” conduct – i.e., virtuous behavior as demanded by God – applies to all of one’s relationships, regardless of the nature of the being with whom one interacts, even if it be a mindless silicon-based humanoid.

Finally, Maimonides makes the plea for virtuous behavior (specifically, “mercy”) based on the “aesthetic” – i.e., the self-centered concern for one’s own needs. The logic, according to Jewish theology (as elaborated in Chapter 1), is that God bestows mercy on those who show mercy to others.¹⁴ As in the “religious” claim for virtuous behavior with a Canaanite slave – or in our case, a mindless robot – it is not made based on the value of the slave or robot, but on the needs of the master himself (specifically, mercy).

From this “aesthetic” perspective it might be claimed that mercy cannot be exhibited toward a mindless machine, it having no subjective experience to appreciate such affections. Indeed, the Talmudic principle teaching that divine mercy is granted one who exhibits mercy states, “Whoever is merciful *to the creations (briot)* will receive mercy” – implying mercy towards humans and animals.¹⁵ Nevertheless, the mindless humanoid *does* evoke a sense of interacting with “creations (*briot*)” – for, as will be explained in Chapter 2, once an entity exhibits enough human-like qualities, we see it as “human.” Accordingly, an individual could be credited with exhibiting genuine mercy toward a mindless robot. But more significantly, Maimonides did *not* quote the above Talmudic principle *in toto* but left out the proviso “*to the creations (briot)*”, writing only, “Whoever is

¹³ While Kierkegaard defines “religious” as following the divine imperative against the ethical, I have explained elsewhere that Jewish thought sees the religious as divinely ethical and not counter to natural morality (Navon 2014).

¹⁴ If one is merciful, even though he may not otherwise merit heavenly mercy, he is nonetheless accorded it undeservedly (Maharam Shik, Shabbat 151b); for God will so grant mercy in order that the individual will become genuinely merciful (Hevruta, Shabbat 151b).

¹⁵ See, e.g., Ayelet HaShachar (Shabbat 151b, s.v., *kol hamerachem*) who explains that “creations” implies that this principle applies to humans and by extension to animals.

merciful [...] will receive mercy.” For Maimonides, I suggest, it is the expression of the moral disposition (i.e., mercy) by the agent that is of import, not the nature of the recipient.¹⁶

In summary, then, the three stages “on life’s way” demonstrate that virtuous behavior is demanded, but only because of the value of the moral disposition of the human involved. The mindless robot, in and of itself, demands nothing more than the ethical treatment due a toaster. This approach allows for what might be called “informed virtuosity,” as opposed to what might be called, “blind virtuosity,” that urges virtuous behavior without regard for authenticity. My “informed virtuosity” approach, as derived from Maimonides’ Virtuous Servant Owner, is developed in Chapter 2: *The Virtuous Servant Owner A Paradigm Whose Time has Come (Again)*.

1.2 The Mindful Robot Dilemma (1)

In contradistinction to the mindless robot, the mindful robot introduces an ethical quandary of an entirely different nature. For, while the mindless robot forces us to look at ourselves and ask how we, as humans, should act for the sake of maintaining our own moral integrity, the mindful robot demands that we look at the humanoid and ask how we, as humans, can morally create a conscious being to serve our needs. Of course, this question then reflects back on the moral integrity of humans and human society. For, even if we justify creating conscious servants, what kind of individuals will we have become and what kind of a society will we have created?

To appreciate the gravity of creating such a being, one must understand the role of consciousness among the fundamental ontological features of human beings. To this end, Chapter 3, entitled *To Make a Mind*, provides, as per its subtitle, *A Primer on Consciousness Machines*. There, after delineating a long list of ontological features from various and varied sources, I show that consciousness (specifically, second-order phenomenological consciousness) is held to be the uniquely defining feature of human beings. Furthermore, I also explain how what is called in modern terminology

¹⁶ Further support for this position can be seen in Maimonides explanation of the biblical commandment against cursing the deaf (Sefer HaMitzvot, Neg. 317).

“consciousness” is nothing other than what was once referred to as “soul.” Human-level consciousness or soul, then, is that which will confer human-like status on a synthetic humanoid.¹⁷

This understanding has broad consensus, and consequently, most find it morally repugnant to build conscious humanoids to serve humans. That said, there is an argument to be made for making such servants, if we could ensure that they had a “good life.” Many have argued against such a possibility from the three classic moral approaches of deontology, consequentialism and virtue ethics. In Chapter 5, *Eudemonia of a Machine*, I bring their arguments as well as culling new sources from within the classic approaches to further the position against creating conscious robot servants. However, my primary contribution to the discussion is to argue that the “good life” must be seen in light of Aristotle’s *eudemonia* and Maimonides’ *summum bonum*. To understand what exactly is Maimonides’ *summum bonum* – a most contentious issue amongst scholars of Jewish philosophy – I provide Chapter 4, entitled *Polemics on Perfection*. With this understanding in hand, I argue that making conscious robots to serve would be ethically prohibited, as it would thwart their possibility of attaining the perfection – *summum bonum* – to which all conscious beings (should) aspire.

1.3 The Mindful Robot Dilemma (2)

But if we cannot make conscious beings to serve us, the question is then, can we make them for the reasons we make humans – as children, friends, lovers, etc.? Here there is far less reticence to stop such seemingly innocuous efforts at advancing science and technology. Indeed, some even argue that we may be obligated to make such beings instead of humans! Julian Savulescu (2001) offers a bioethical approach that argues for “programming” dispositions in our offspring. He calls it the principle of procreative beneficence (PPB), which claims that we have a moral obligation to bring forth beings that will enjoy “the best life possible.”¹⁸ He does not, however, propose we engineer

¹⁷ For a discussion on the Jewish approach to determining moral status, see below: Appendix: The Moral Status of a Soul.

¹⁸ For a discussion on Savulescu’s approach, see Liao (2014: 106-108), who refers to it as “The Perfectionist View.”

dispositions toward a specific vocation, but that we instill dispositions (e.g., intelligence, emotional regulation, etc.) to ensure, according to some accepted definition, “the best life possible” (2001: 419-421).

In contrast to this, Johannes Grossl (2020), among others (see, e.g., Holland 2016, Saunders 2015 & 2016, Overall 2011 in Danaher 2019b), argues that applying Savulescu’s principle will negatively impact moral reasoning, freedom of will, and more. John Danaher (2019b: 30-33) explains that the principle, as applied to human persons, is quite problematic, demanding (among other things) that children be sired only through IVF after genetic testing/manipulation.¹⁹ These problems are absent, however, when applied to bringing about artificial persons (i.e., conscious humanoids). Accordingly, taking Savulescu’s principle to its logical conclusion could demand that anyone seeking to bring a new life to the world be morally obligated to do so via a conscious humanoid, that being the surest way to bring about “the best life possible.”

Such arguments notwithstanding, the fundamental question is, “Should we be making conscious humanoids?”

In the final chapter of this thesis, “*Let Us Make Man in Our Image*”, I address this question based on the paradigm of the Golem in Jewish literature. I first bring a mystical Midrash that tells of the only human-level conscious Golem. Already here we learn that Jewish literature is very uncomfortable with the idea of creating a conscious humanoid, as the Golem itself exhorts its creators against the propriety of making it.

But this mystical Midrash does not carry the normative weight of the original Talmudic Golem – the Gavra. Accordingly, I analyze in depth the short (35-word) but seminal Talmudic narrative to demonstrate that it argues emphatically, despite different readings over time, for the categorical prohibition of creating a conscious humanoid. My reading finds the prohibition a deontological value, it being adduced neither from consequences

¹⁹ Danaher notes that this demand puts an undue burden on the would be mother. Importantly, R. Bleich (1998) notes that Jewish law does not obligate woman to employ procedures outside natural pregnancy.

nor virtues but from the actions of great teachers – in rabbinic parlance, “*maaseh rav*.”²⁰ And, I argue, it is precisely the fact that this teaching is sourced in rabbinic action that gives it not only ethical value but legal value. Accordingly, this deontological prohibition presents us with a legally binding ethical duty.

For those who believe in the sanctity of these texts and their values, this paradigm provides what could be called divine direction; for those who do not, the paradigm nevertheless serves as a precedent to be analyzed on its philosophical merits.

²⁰ Note that it is entirely irrelevant whether this “*maaseh*” (act) ever actually took place or not, for it is brought as if it did.

2. Thesis Structure

To develop a cogent and comprehensive response to these dilemmas, my thesis encompasses the following six essays, referred to as “chapters”:

Article #1: *Finding Virtue in a Law on Slaves – An Analysis of Maimonides’ last Law Concerning Slaves*

Maimonides’ last Law on Slaves (Hil. Avadim 9:8) has been quoted far and wide as an example par excellence of his great creativity and broad vision, expressing, in the most eloquent of terms, the obligation to exercise mercy and compassion. This obligation, despite its appearance in the laws of slaves, is not merely a call for mercy and compassion towards one’s slave but towards every “other” with whom one engages. Accordingly, the ethical interrelations delineated here apply no less to the mindless “others” in our midst – e.g., social robots and LLMs. Nevertheless, while the subtext of this law calls for ideal ethical interrelations with all others, the law *qua* law-of-slaves recognizes that there is a distinction between one’s slave and one’s friend. It is precisely this fine distinction, between encouraging ethical relations on the one hand, and maintaining social/emotional distance on the other, that is at the heart of the paradigm I extract from Maimonides’ last Law on Slaves. It is a paradigm I call the Virtuous Servant Owner, developed here in Chapter 1 and applied to our modern mindless servants in Chapter 2.

Publication: “Finding Virtue in a Law on Slaves – An Analysis of Maimonides’ last Law Concerning Slaves,” *Tradition: A Journal of Orthodox Jewish Thought* 56.4 (2024), <https://doi.org/10.54469/DHZM677HG>.¹

Article #2: *The Virtuous Servant Owner: A Paradigm Whose Time has Come (Again)*

Artificial Intelligence (AI) – whether in the form of anonymous LLMs, embodied online as virtual people, or offline as Social Robots – poses a huge moral dilemma.

¹ Note: due to publisher constraints, the published version contains significantly fewer footnotes and references compared to the more comprehensive version presented herein.

On the one hand, these machines are just that, machines. Accordingly, some thinkers propose that we maintain this perspective and relate to them as mere machines. Yet, in treating them as such, we deny our own natural empathy, ultimately inculcating vicious as opposed to virtuous dispositions. Many thinkers thus apply Kant’s approach to animals – “he who is cruel to animals becomes hard also in his dealings with men” – contending that we must not maltreat AI lest we maltreat humans. On the other hand, because we innately anthropomorphize entities that behave with autonomy (let alone entities that exhibit beliefs, desires and intentions), we become emotionally entangled with them. Some thinkers actually encourage such relationships, but there are problems here also. To begin with, many maintain that it is imprudent to have “empty,” unidirectional relationships for we will then fail to appreciate authentic reciprocal relationships. Furthermore, such relationships can lead to our being manipulated, to our shunning of real human interactions as “messy,” to our incorrectly allocating resources away from humans, and more. In this chapter, I review the various positions on this issue and propose an approach that applies the “Virtuous Servant Owner” paradigm developed in Chapter 1. It is an approach that allows us to take the middle ground between the extreme of treating these new “others” as mere machines versus the extreme of accepting them as having human-like status.

Publication: “The Virtuous Servant Owner—a Paradigm Whose Time Has Come (Again),” *Frontiers in Robotics and AI* 8 (September 22, 2021), <https://doi.org/10.3389/frobt.2021.715849>.

Article #3: *To Make a Mind: A Primer on Consciousness Machines*

The dream of making a conscious humanoid – whether as servant, guard, entertainer, or simply as testament to our own creative prowess – has ever piqued the human imagination. However, while past attempts to create such beings were performed by magicians and mystics, today scientists and engineers are doing the work to turn myth into reality. In this chapter, I seek to provide a primer of the fundamental concepts that define human consciousness, as well as the fundamental approaches that define machine consciousness. This chapter thus establishes both the technical and the philosophical foundations to address the dilemmas

surrounding conscious AI discussed in Chapters 5 and 6. In addition, within the context of the various machine consciousness approaches, I offer what could be called a theological contribution to the discussion, introducing a novel understanding of the biblical references to blood and soul as they relate to consciousness in general and machine consciousness in particular.

Publication: “To Make a Mind – a Primer on Conscious Robots,” *Theology and Science* (Jan., 18 2024), <https://doi.org/10.1080/14746700.2023.2294530>.

Article #4: *Polemics on Perfection: Maimonides’ Last Law on Slaves Resolves the Debate*

“What is the *summum bonum*?” This question is, without exaggeration, the ultimate existential question. Of no less import is its corollary: how is one to achieve this *summum bonum*? Maimonides’ *Guide for the Perplexed* endeavors to provide the answers, yet the precise intent of his answers has generated no less perplexity than the original questions themselves. Indeed, modern Jewish philosophers have spared no ink in trying to ascertain Maimonides’ intentions on human perfection. In this chapter, I seek not to add yet another perspective on Maimonides’ intentions, but to provide compelling support for existing positions; support found—worlds apart from the ivory towers where philosophers spill their ink—in the private quarters of a Canaanite slave. Applying the ideas developed in Chapter 1, Maimonides’ *summum bonum* is disclosed in light of his last Law on Slaves. This analysis reveals a powerful paradigm, which could be called the “Jeremiah Glory of Man” paradigm, to address the mindful robot dilemma discussed in Chapter 5. As will be demonstrated there, a conscious being, regardless of whether the substrate supporting consciousness is biological (i.e., carbon) or technological (e.g., silicon), must be allowed to aspire to its *summum bonum*. To thwart such an aspiration by designing a being for instrumental use – i.e., to relieve us of our burdens – is to thwart the very essence of that being and would thus be an unconscionable ethical wrong.

Publication: “Polemics on Perfection – Maimonides’ Last Law on Slaves Resolves the Debate,” *Review of Rabbinic Judaism*, Volume 27, Number 2 (Sep. 3, 2024), <https://doi.org/10.1163/15700704-20240007>.

Article #5: *Eudemonia of a Machine: On Conscious Artificial Servants*

Henry Ford once said, “For most purposes, a man with a machine is better than a man without a machine.” To this, engineers today propose an addendum – “and a man that *is* a machine is best of all” – which they have made their goal. The world over, engineers are working to make the ultimate machine, “the holy grail of artificial intelligence,” a *conscious* humanoid. On the one hand, such a “machine” will be capable of relieving us of all our burdens. On the other hand, in so doing, will we not have “birthed,” as it were, a new class of slaves? In this chapter, I seek to summarize the various arguments made in this debate, bring to bear moral positions from the philosophy of technology, philosophy of law and philosophy of religion, as well as demonstrate the moral impropriety of such an endeavor from each of the classic moral approaches (i.e., virtue ethics, consequentialism, Kantian deontology). Finally, given that the debate centers around what is the “good life” for human or humanoid, I expand upon Aristotle’s Eudemonia and Maimonides’ *Summum Bonum* (as developed in Chapter 4) to argue that life is precious in its affordance to allow conscious beings, human or humanoid, to aspire to the best life possible.

Publication: “Eudemonia of a machine,” *AI Ethics* 5, (2025), <https://doi.org/10.1007/s43681-024-00553-z>.

Article #6: *Let Us Make Man in Our Image”: A Jewish Perspective on Creating Conscious Robots*

The dream of making conscious humanoid robots is one that has long tantalized humanity, yet today it seems closer than ever before. Assuming that science can make it happen, the question becomes, “Should we make it happen?” Is it morally permissible to create synthetic beings with consciousness? In this chapter, as opposed to Chapter 5, I question the moral propriety of creating a synthetic conscious being – i.e., a mindful humanoid – not as a servant, nor for any specific purpose other than to join the family of human-level conscious beings. While a consequentialist approach may seem logical, attempting to assess the potential positive and negative consequences of such a revolutionary technology is highly speculative and raises more questions than it answers. Accordingly, some turn to ancient and not-so-ancient stories of “automata” for direction. Of the many

automata conjured throughout history, if not in matter then in mind, the Golem stands out as one of the most persistent paradigms employed to discuss technology in general and technologically engendered life forms in particular. In this chapter, I introduce a novel reading of the Golem paradigm to argue not from consequentialism, but from a deep-seated two-thousand-year-old tradition, the ethical implications of which are wholly deontological.

Publication: “Let Us Make Man in Our Image’ - A Jewish Ethical Perspective on Creating Conscious Robots,” *AI and Ethics* (Sep. 12, 2023), <https://doi.org/10.1007/s43681-023-00328-y>.

As can be seen, though each chapter is a stand-alone essay that introduces some novel idea, the chapters are closely interrelated: Chapter 1 provides the background for Chapter 2, Chapter 3 the background for Chapters 5 and 6, and Chapter 4 the background for Chapter 5.

In addition, while Chapters 1 and 4 do not explicitly refer to AI, they provide, as noted, an essential understanding of the paradigms to be applied to the AI dilemmas. But their importance goes beyond mere application, since, given that this thesis is written under the aegis of the Jewish Philosophy department, no less important than employing Jewish paradigms to address modern dilemmas is the need to elucidate the paradigms in an innovate manner that contributes to Jewish philosophy itself.

3. The Thesis

Ch. 1:

Article #1 - Finding Virtue in a Law on Slaves

An Analysis of Maimonides' last Law Concerning Slaves

Introduction

In his seminal work, *Ethics of Responsibility*, Rabbi Walter Wurzburger analyzes Maimonides' philosophy of morality to demonstrate that Jewish ethics appeals not only to moral law (i.e., "the ways of wisdom"), what he refers to as "the ethics of obedience," but to moral piety (i.e., "the ways of piety"), what he refers to as "the ethics of responsibility."¹ There is, however, a seeming contradiction between these two approaches (i.e., each enjoining a differing directive for a given situation), thus leading to a "vast literature dealing with the seeming contradiction" (Wurzburger 1994: 82).² Wurzburger rejects the contradiction and demonstrates that they are both essential to Maimonides' complete moral approach – what he refers to as "Covenantal Ethics" and what I will refer to, for the sake of clarity, as simply the "dual-moral approach."

In this chapter, I argue that not only does Maimonides consider the two moral approaches essential, not only are they the way to moral perfection, but they are the way to human perfection; their realization – together – a normative obligation. Maimonides encapsulates all this in his last Law on Slaves (Hil. Avadim 9:8) wherein he explicates one's moral obligations to his slave; for if moral equity is to be pursued in the interrelationship with one's slave – i.e., the "other" at the bottom of the social ladder – then all the more so is it to be pursued with one's contemporaries, let alone with one's superiors. Maimonides himself makes this very argument: "This [kindness] we owe to the lowest among men, to the slave; how much more, then, must we do our duty to the freeborn, when they seek our assistance?" (Guide 3:39).³

¹ Wurzburger 1994. See also: Wurzburger 2008, Blau 2000, Shatz 1996, Shatz 2005.

² Here are some of the many sources dealing with this issue: Rawidowicz (1954 in Wurzburger, 1994); Cohen (1972 in Wurzburger, 1994); Weiss and Butterworth (1975); Schwarzschild (1990); Weiss (1991); Kreisel (1992); Shatz (2005); Strauss (2013); Shapira (2018). See also Ravitsky (2014) and sources in fn. 2 therein.

³ It is important to note that neither Maimonides, nor the Torah upon which he expounds, extols slavery. Rather, given the fact that humanity at the time would not accept abolition, at least the institution could be made a bit more kind. For sources, see fn. 11, 12 further herein.

Before beginning our inquiry, worthy of note is the fact that this last Law of Slaves, despite its seeming lack of relevance to modern society, has garnered lavish praise for its creativity, vision, and far-reaching boldness of thought:

“As is his systematic wont, Maimonides dedicates a section in his Code to the laws governing slaves but ends, as is also his custom in a number of other sections, with his own creative, rhetorical, non-halakhic sentiments which finally collapse the distinction between Hebrew and non-Hebrew slaves” (Diamond 2016).

“*Laws Concerning Slaves* 9:8, which ends the Book of Acquisition, contains one of the finest and most complete expressions of Maimonides’ conception of the duty to act mercifully and compassionately” (Shapira 2018).

“This remarkable passage is noteworthy for its extreme language as much as its grand vision. Maimonides, the halakhic voice who is sometimes majestic but rarely extreme, consciously resorts to immoderate terminology and expansive expression when ending the Book of Acquisition in his code” (Korn 2002).

Yet for all this, I have found no thinker who has addressed the myriad questions demanded by this extraordinary text. R. Yosef Karo (*Kesef Mishna*, Hil. Avadim 9:8:8) comments that “these are the words of the master and they are appropriate for him,” thus informing us that Maimonides has absolutely no precedent for his claims. But if so, how then does Maimonides dare make these claims in his *legal* corpus? Are his claims really “non-halakhic sentiments,” as Diamond asserts? What exactly are Maimonides’ claims outside of the obvious calls for mercy beyond the letter of the law? On what basis does he support his claims? Why does Maimonides quote the support texts that he does? Are these quotes simply a laundry list of all the biblical and rabbinic texts that speak to the claims of mercy? Is there any structure to be found in Maimonides’ arguments, or is he simply making every claim he can in an attempt to win sympathy for extralegal pleas?

It is my thesis that every sentence, indeed, every word, is hand-picked and hand-placed to build a monumental appeal for the dual-moral approach, an appeal so firmly grounded in biblical and rabbinic thought that it is more than an appeal, it is a normative demand. This claim should come as no surprise as “the care with which Maimonides chose his

words and [source] texts is well known.”⁴ In addition, the Mishnah Torah is known to be an extraordinary text, as Isadore Twersky puts it, an “unprecedented, comprehensive, meticulously arranged”⁵ text that is grand and profound, full of allusions and subtleties, powerfully compressed with a felicity of style.⁶ And if this were not enough, Twersky explains that “a central pillar of the elaborate multidimensional Maimonidean structure”⁷ consists in the expression of both law and philosophy in one and the same text.⁸ In consonance, Menachem Kellner explains that Maimonides, in his Mishnah Torah, addresses two types of readers: “Traditionalists” – i.e., Talmudic scholars well versed in Jewish legal tradition; and “Philosophers” – i.e., Talmudic scholars who have added to their Talmudic studies, “logic, physics and metaphysics without in any way giving up their prior allegiances.”⁹

Accordingly, the “Traditional” reader will note, in this law of slaves, the appeals to appropriate behavior and find the inspiration to modify his actions, no matter at what stage of life he may be. The “Philosophical” reader, on the other hand, will discern – in addition to the appeals to appropriate behavior – the appeals to appropriate dispositions, and will thus be inspired to follow the path to human perfection.

To explicate these varying readings, I first review the sections of the text, numbered below (1-6), to gain a general understanding of the material. I then explain how the text calls the traditional reader to virtuous behavior in the section entitled, “The Traditional Reader: Virtue for All.” This is followed by the section entitled, “The Philosophical Reader: Chiasmus,” in which I analyze the structure of the text as guide to human perfection.

⁴ Kellner (1990: 45). See also Soloveitchik (2016 :124).

⁵ Twersky (1980: 359).

⁶ *ibid.*: 508-509.

⁷ *ibid.*: 360.

⁸ *Ibid.*: 507-509. Hartman (1986: 26) concurs, though he notes that Strauss does not.

⁹ Kellner (2009: Ch. 3). Kellner (1990: Ch. 3) calls them “halakhists plain and simple” versus “halakhists who have perfected themselves in natural science and philosophy” (*ibid.*: 17).

I herein bring the original next (in *italics*) interleaved with my preliminary clarifications.

Hilchot Avadim (9:8) Explained

*It is permissible to work a beathen slave relentlessly.*¹⁰

Biblical law often promulgates rules in concert with ancient custom while nevertheless seeking to provide a moral improvement on the accepted state of affairs.¹¹ As such, the strict letter of law allows for slavery but with various moral restraints.¹² The law, however, is seen as a starting point, a floor and not a ceiling, to use the words of Rav Soloveitchik.¹³ Wurzburger explains this to mean that “Jewish piety involves more than meticulous adherence to the various rules and norms of religious law; it also demands the cultivation of an ethical personality... We are commanded to engage in a never-ending quest for moral perfection, which transcends the requirements of an ‘ethics of obedience’... [T]he halakhic [i.e., Jewish legal] system serves merely as the foundation of Jewish piety.”¹⁴

¹⁰ See Hil. Avadim 1:6 which defines “relentless” (*parech*) as: assigning tasks with no useful purpose or end. See Kesef Mishna (Hil. Avadim 9:8) on how this is the Biblical law.

¹¹ See, e.g., Lamm (2007), Rabinovitch (2003: esp. 9), Korn (2002), Lichtenstein (2002: 16-17). On the changing attitudes to slavery within Jewish thought, see, e.g., Shmalo (2012). On slavery as a concession to prevailing human attitudes, see, e.g., Shatz (2012: 11), Blau (2002: 56) and the impassioned argument to this effect in David Einhorn’s Civil War sermon (Saperstein, 2012: 212-213).

¹² For example, killing a slave entails capital punishment (Ex. 21:20, Rashi ad loc.), a slave is set free if injured (Ex. 21:26-27, Kid. 24a), a slave rests on the Sabbath (Ex. 20:9); a runaway slave is not to be returned (Deut. 23:16). On the differences between ancient slavery versus that of the Torah, see Beasley (2019: esp. fn. 3).

¹³ Wurzburger (1994: 32).

¹⁴ Wurzburger (1994: 3-4). As an aside, this appears to follow Kant’s ethical approach, as explained by Loudon: “Kant’s notion of action *aus Pflicht* [“out of duty”] means in the most fundamental sense not that one performs a specific act for the sake of a specific rule which prescribes it ... but rather that one strives for a way of life in which all of one’s acts are a manifestation of a character which is in harmony with moral law” (1986: 485).

Accordingly, Maimonides starts with the legal “floor” only to show that we should – and must – rise far above it.¹⁵ To this end, he constructs an argument with the precision of a geometrical proof, complete with hypothesis and assertions supported by axioms in the form of biblical verses. His proof is structured in six statements as follows. The first statement (1) is the hypothesis that claims one is to treat his slave with piety (*bemidat hassidut*),¹⁶ beyond the letter of the law, and with justice (*betzedek*),¹⁷ according to the letter of the law yet charitably. Maimonides then brings four statements (2-5), which expand on and support these assertions, finally closing with a statement (6) that rounds out the claims, each supported by biblical verse.

The complete “proof” is built in chiasmic form:

- (1) Hypothesis Claimed
- (2) Support for Piety
- (3) Support for Justice
- (4) Support for Justice
- (5) Support for Piety
- (6) Hypothesis Conclusion

Let us now analyze the text according to this outline.

(1) Hypothesis Claimed

Though this is the law (din), the quality of piety (midat hassidut) and the ways of wisdom (darkei hochma) demand of a human being to be compassionate (rachaman) and pursue justice (tzedek), and not make heavy his yoke on his slave nor distress him.

¹⁵ It is interesting to note that Kant used the same format, starting with the letter of the law allowing ownership but then moving on to demand virtue (*Metaphysics of Morals* 6:284).

¹⁶ The term is variously translated as “virtue” or “piety.” I use “piety” to denote *middat hassidut* and “virtue ethics,” while I reserve “virtue” to refer to moral virtuosity in general (i.e., include all moral paths).

¹⁷ The term is variously translated as “righteousness” or “justice.” I use “justice” to emphasize its affiliation to the law.

Maimonides, here, raises us off the floor of the law, outlining his thesis that calls for piety and justice. This is not at first so apparent because the text seems to jumble ideas, intertwining notions that are odds with one another. That is, in what way is piety related to the ways of wisdom that they are mentioned together? Even more striking in this regard is the mention of compassion together with justice; why are these two clearly distinct approaches juxtaposed? Remarkably, by rearranging the words in a simple and straightforward manner, the meaning of the text is revealed. To understand, let us number each of the expressions in the text as (1) through (6) as follows:

(1) the quality of piety (midat hassidut) and (2) the ways of wisdom (darkei hochma) demand of a human being to (3) be compassionate (rachaman) and (4) pursue justice (tzedek), and (5) not make heavy his yoke on his slave (6) nor distress him.

Then let us simply group the odd numbered expressions together as one directive and the even numbered expressions as a second directive.

(1) the quality of piety (midat hassidut) demands of a human being to (3) be compassionate (rachaman) and (5) not make heavy his yoke on his slave.

(2) the ways of wisdom (darkei hochma) demand of a human being to (4) pursue justice (tzedek), and (6) not distress him.

From this simple reordering, two clear directives – that typify two moral approaches – emerge. Piety, referred to here as “*midat hassidut*,” demands “*compassion*,” epitomized by “*not making one’s yoke heavy*.” Justice, referred to as the “*ways of wisdom*,” demands the “*pursuit of justice*,” exemplified by “*not distressing*” the other.¹⁸ These two directives, I suggest, correspond to the two approaches to moral conduct found in Maimonides’ Laws of Moral Character (Hil. Deot) wherein he speaks of the ways of the “pious” (1:5) as opposed to the “ways of the wise” (1:4).¹⁹ Furthermore, these two approaches can be

¹⁸ To distress another is to treat him unjustly, as will be demonstrated further herein.

¹⁹ Also in Eight Chapters (Ch. 4). See, e.g., Strauss (2013: 561); Weiss and Butterworth (1975: Intro.). As an aside, Strauss and Weiss and Butterworth refer to the two categories – i.e., “the ways of the wise” and

shown as none other than two of the most fundamental approaches in ethics – virtue ethics (i.e., agent morality) and duty ethics (i.e., act morality).²⁰

Piety (*midat hasidut*), also referred to as the “ethics of the pious,”²¹ demands going beyond the letter of the law (Hil. Deot 1:5),²² and can be seen as corresponding to virtue ethics,²³ which looks to the character of the agent and seeks to align appropriate action with appropriate character traits.²⁴ The assertion that to act with piety (*midat hassidut*) corresponds to virtue ethics fits quite seamlessly with the continuation of Maimonides’ claim here that calls for one to exercise “compassion” – not merely a “virtue” but arguably the ultimate virtue, the defining virtue of the Jewish people (as noted by Maimonides later in this very law, “*they are compassionate to all*”).²⁵

Justice (*tzedek*), on the other hand, can be seen as corresponding to the ethical approach known as deontology (a.k.a., act morality), whereby appropriate action is defined by the rightness or wrongness of an act.²⁶ Wurzbürger refers to this approach variously as the “ethics of the middle road,” “the ethics of wise,” or “the ethics of duty” and “the ethics

“the ways of the pious” – as “philosophic morality” and “Torah/Jewish morality,” respectively. I will argue, as others have before me, that the two are both “Jewish.”

²⁰ That Maimonides subscribes to these two moral approaches, see, e.g., Wurzbürger (1994: Ch. 5), Wurzbürger (2008: 91-99). Also Shatz (2005), Kellner (2009: 72).

²¹ See, e.g., Wurzbürger (2008: 31), Wurzbürger (1994: 79).

²² See, e.g., Wurzbürger (1994: 79), Korn (2002: 8-9), Wurzbürger (2008: 31), Shapira (2018: 570), Friedberg (2019: 17). See also Guide 3:53 wherein he explains *besed* as going beyond deserved kindness.

²³ Wurzbürger (1994: Ch.5, esp. 81); Wurzbürger (2008: 103).

²⁴ See Aristotle’s *Nicomachean Ethics*. On the relation here between action and character in virtue ethics, Shatz puts it concisely: “Actions are important only insofar as they are conducive or inimical to the production of virtuous or vicious characteristics or insofar as they are the effect of virtuous or vicious traits” (Shatz 2005: 170).

²⁵ On compassion (*rachamim*) as a Jewish virtue, see, e.g., Yevamot 79a; Beitza 32b; J. Kid. 4:1(42b); Hil. Deot 1:6; Hil. Issurei Biah 19:17; Hil. Matanot Aniim 10:2; Hil. Hovel 5:10; Sefer Hahinuch 42, 44, 498; Sefer Haben Yakir Li Efraim in Daf al Daf (Shabbat 151b). Maimonides did not here follow Aristotle, who apparently did not hold compassion to be a virtue (see, e.g., Shapira 2018).

²⁶ Kant ([1785] 2006). See also, e.g., Spero (2016: 204-207).

of obedience.” The former two terms correspond to the doctrine of the mean, or “middle road,” articulated by the “wise,” i.e., Aristotle and Alfarabi,²⁷ wherein one is to seek, both in one’s actions and one’s dispositions, a balance – the mean between two extremes.²⁸ The latter two terms express the indisputable notion that Judaism demands adherence to a code of law (i.e., Torah), given expression in rulings known as *halacha*. The four terms coalesce to one moral approach, such that when Maimonides refers to “the ways of the wise” he refers to the “ethics of the middle road” for which duty to, and obedience of, the law serve to cultivate. And so he writes:

“The perfect Law leads us to perfection, as one who knew it well testified: ‘The Law of the Lord is perfect ... making wise the simple’ (Psalms 19:8), the point being that man should be natural, following the path of moderation [literally: the mean]” (Eight Chapters: Ch. 4).²⁹

Now, here in this Law of Slaves (as well as in Laws of Repentance 6:5), Maimonides associates this “middle road” moral approach with the pursuit of justice (*tzedek*). Justice, Maimonides notes in his Guide (3:53), is conflated with *tzedakah* (charity) as moral virtue. That is, while *tzedek* in secular philosophic terms implies giving everyone their due (i.e., what is generally referred to as justice), its meaning is more nuanced in biblical ethical terms, colored by *tzedakah*, and thus connoting a concern for the other.³⁰ Accordingly, *tzedek* is a kinder, gentler justice, what might be called, charitable justice. As an example, Maimonides (Guide 3:53) brings the biblical commandment to return a poor person’s collateral (Deut. 24:10-13) – an act that goes farther than strict justice (e.g., maintaining

²⁷ See, e.g., Kellner (2009: 47), Shatz (2005: 174), Ravitsky (2014: fn. 3).

²⁸ On the equation of “the ethics of the wise” to “the ethics of the mean,” as in, e.g., Hil. Deot 1:4, see (Weiss and Butterworth 1975: 7), (Weiss 1991: 102), (Strauss 2013: 558-561).

²⁹ On the Law cultivating the mean, see, e.g., Altmann (1972: 24), Weiss (1991: 25), Fox (1994: 93–123), Wurzbürger (1994: 78-79), Wurzbürger (2008: 103), Ravitsky (2014: 33). See, however, Shatz (2005: 177-179) who explains that the Law does not command the mean, *per se*, but commands actions that will lead to the mean of moral virtues.

³⁰ See, e.g., Kellner (2009: 166-167), Schwarzschild (2007: 578-579), Friedberg (2019: 16). See Shapira (2018) for an alternative view.

possession of the collateral), but not as far as overt piety (e.g., relinquishing the loan) – thus yielding a clear example of the “ethics of the middle road.”

This example commandment is then expanded to all the commandments, as Maimonides concludes his explanation of moral justice (*tzedek/tzedakah*) with the verse, “And it shall be *tzedakah* (justice) unto us, if we observe to do all this commandment before the Lord our God, as He hath commanded us” (Deut. 6:25). Clearly, the pursuit of justice, and its implied moral virtues, are to be cultivated through the observance of the commandments of the Torah. And it is through observance of the commandments that the “ethics of the middle road” is realized.³¹ For, on the Torah’s characterization of its laws as “just (*tzadikim*) statutes and judgments” (Deut. 4:8), Maimonides writes that “just” means “balanced” – explaining that observing the Torah’s “just statutes and judgments” will bring one to just – i.e., balanced – actions (Guide 2:39) and just – i.e., balanced – virtues (Guide 3:39).³²

Finally, in accordance with my suggested reading of this law of slaves as promulgating a dual-moral approach, the hypothesis ends with two examples of practical behavior, i.e., that one “*not make heavy his yoke on his slave*” and that one “*not distress him.*” These two examples align with the ethics of piety and the ethics of justice, respectively. That one “*not make heavy his yoke on his slave*” is a call to piety, a call to go beyond the letter of the law, is quite clear, for one could, by rights (*din*), give his slave as much work as the latter can bear (i.e., *la'avod oto befarech*). This call to supererogation stands in contradistinction distinction to the call to “*not distress*” the slave, which is simply the call to be just (in the Maimonidean sense). To wit, Maimonides uses the term “distress” (*yeitz'er*) in the title of the commandment (neg. #234)³³ that prohibits asking for loan payment back from one who does not currently have the money. This is precisely the same ethic underpinning the commandment (pos. #199) that demands one return the collateral for a loan to a poor person.³⁴ And it is precisely this positive commandment that Maimonides (Guide

³¹ See fn. 29 above.

³² See also Shatz (2005: 180).

³³ As found in the Bodleian (Oxford) manuscript published in Maimonides (1981: 14b).

³⁴ Worthy of note is that, though the verses substantiating the two commandments are in different places (neg. #234 on Ex. 22:24, pos. #199 on Deut. 24:12) the two notions are discussed in two immediately

3:53) brings as an example of acting justly (*tzedakah*). Accordingly, the call to “*not distress him*” is a call to the ethic of justice.

Appeals to Justice and Piety

Let us now look at the four statements (2-5) in support of this dual-moral approach.

(2) Support for Piety

He should give him to eat and drink of every food and drink. The sages of old had the practice of sharing with the slave every dish they ate. And they would provide food for their animals and slaves before partaking of their own meals.

Maimonides begins his claim for compassionate, and clearly supererogatory, action by prescribing specific daily acts of kindness based on living examples,³⁵ for “ethical conduct ultimately presupposes concrete exemplars.”³⁶ Nevertheless, while Talmudic precedence could be proof enough, Maimonides continues by bringing a biblical verse to drive home his claim.

As it is said, “As the eyes of slaves follow their master’s hand, as the eyes of a slave-girl follow the hand of her mistress, [so our eyes are toward the Lord our God, awaiting His favor].”

In using this verse, Maimonides seeks to anchor his seemingly extra-biblical demand within the Bible itself. He appeals to reason by using a verse that equates master and slave, both in need of “favor.” Importantly, the word “favor” (*h.n.n.*), denoting

adjacent verses (Ex. 22:24, Ex. 22:25) thus further linking their underlying ethic. The reason for not using Ex. 22:25 for (pos. #199) is because it has its mirror opposite command (neg. #239) sourced in Deut. 21:10.

³⁵ Referring here to R. Yohanan bar Nafcha who would himself drink two glasses of wine and eat meat, and so gave to his slaves (Jerusalem Talmud, BK 8:4); as well as to the numerous sages who did likewise (Ketubot 61a).

³⁶ Wurzburger (2008: 30). For more on this ethic, see below sec. “The Philosophical Reader: Chiasmus,” esp. fn. 68.

completely undeserved kindness,³⁷ highlights the vulnerable position shared by both master and slave. The biblical text thus supports Maimonides' supererogatory prescriptions by demonstrating that, while the Bible may not demand the claimed benevolent behavior explicitly, the Bible does expect one to realize his own similitude to his slave and act with that same favor he "awaits" from his Master – i.e., "Lord our God."

Two points to note here and which will be elaborated upon further on:

- (1) The biblical verse expresses the overarching meta-ethic of going beyond the law and thus supports the claim that piety is demanded by the law itself.³⁸
- (2) The argument for piety here is based on the inherent moral worth of the slave.

(3) Support for Justice

Nor should a master disgrace his slave, neither physically nor verbally; the biblical law gave them to servitude, not to disgrace.

Maimonides starts this claim for treating the slave justly by quoting the Talmudic interpretation of a biblical verse – that is, biblical law. The verse (Lev. 25:46) teaches that one may take foreigners as slaves, saying, "of them may you work/enslave (*ta'avodu*)," upon which the Talmud notes: "to work, not to disgrace" (Niddah 47a). Accordingly, this demand "not to disgrace them" is based in law and thus falls under the rubric of "justice" (*tzedek*) – i.e., act-morality.³⁹

And one should not treat him with constant screaming and anger, but rather speak with him calmly and listen to his complaints.

³⁷ See, e.g., Guide 1:54; Avraham ben HaRambam (Gen. 33:5); Rashi (Deut. 3:23).

³⁸ See fn. 51 regarding the problem of a law to go beyond the law.

³⁹ To be clear, the Law (i.e., Torah), which includes written verse and oral (rabbinic) interpretation, can have laws that express a stricture (i.e., *din*), but can also express a mean (i.e., *tzedek*). The "din" represents the "floor," while "lifnim mishurat hadin" (i.e., piety) represents an ideal and "tzedek" the mean. The application of these various approaches will be elaborated further herein.

Maimonides here further details his instructions for day-to-day interactions, now on the level of intercommunications, promulgating two imperatives. The first, on action, demands that one not *scream* but rather *speak calmly*; the second, on disposition, demands that one exhibit not *anger* but rather understanding, by *listening*.

Interestingly, while Maimonides teaches elsewhere that anger is a trait to be shunned to the extreme (Hil. Deot 2:3), he also teaches that one can feign anger for educational purposes (Hil. T. Torah 4:5). In contradistinction, here in relation to a slave – where circumstances might beg for such feigning – Maimonides adds no such proviso. This hints, then, to the idea that this law is no mere guide to employing a slave, but a guide to moral development (as will be elaborated below).

Now, while the first imperative – calling for the master to “*speak calmly*” – can be seen as instrumental or having to do with the moral integrity of the master, the latter – calling for the master to “*listen to his complaints*” – goes a step further in moral philosophy, articulating the notion that the other must not be treated merely as a means but as an end. (I wonder if even Kant – who, in his “Formula of Humanity,” demanded that we treat others “always at the same time as an end, never as merely a means” (*Groundwork* 4:429) – would have made such a list of directives to regulate a slave owner).

In support for his claims, Maimonides brings two verses from Job:

This is explicitly stated with regard to the positive paths of Job for which he was praised: “Have I ever shunned justice (mishpat) for my slaves, man or maid, when they quarreled with me... Did not He who made me in my mother’s belly make him? Did not One form us both in the womb?” (Job 31:13,15).

The first verse (31:13) grounds the claim for just treatment explicitly in the Bible, bringing Job as exemplar of ever abiding by justice. Here it should be noted that the term *mishpat*, as opposed to *tzedek*, is used for “justice.” Both relate to justice, both express justice ameliorated with compassion, however, *tzedek*, related to *tzedakah*, always results in acquiescence toward the subject, whereas *mishpat* can result in exacting

requital.⁴⁰ Here Job is saying that he never rejected his slaves' claims against him – i.e., he never expected them to acquiesce on their claims against him. He, however, always acquiesced to them, never shunned *their mishpat*, and thus performed *tzedek*.

Accordingly, while in the previous section (2) Maimonides demonstrated that one must act out of piety (i.e., going beyond the letter of the law) in one's treatment of his slave, here he shows that such is in addition to, and perhaps even an overflow of, one's legal obligation (i.e., according to the letter of the law) to treat his slave justly.⁴¹

The second verse (Job 31:15) furthers the claim that the slave must be treated justly – i.e., as an end – by showing that, in essence, master and slave are equal. That is, both are made of the same biological material via the same physiological process and, most importantly, made by the same Creator and thus made of the same spiritual substance at the core of human dignity and human sanctity.⁴² Accordingly, one has no *justification* – i.e., no claim of justice – to lord oneself over any human “other.”

Two points, once again, to note here:

- (1) The biblical verses argue for the principle of justice (*tzedek umishpat*) underpinned by equality before the law and thus support the claim that justice is here demanded by the law itself.
- (2) The argument for justice here is based on the inherent moral worth of the slave.

(4) Support for Justice

Cruelty and effrontery are not frequent except with the heathen who worship idols. The progeny of our father Abraham, however, the people of Israel – upon whom God bestowed the goodness of the law (Torah), commanding them to observe “just statutes and judgments” (Deut. 4:8) – are compassionate to all.

⁴⁰ Guide 3:53. See Efodi and Shem Tov (ad loc). See also Friedberg (2019: 16).

⁴¹ Diamond (2016: “2. All Are Formed”).

⁴² R. Soloveitchik explains that “Kevod Haberi’ot (respect for human dignity) and social justice are implicit in the biblical concept that man was created in God’s image” (1993: 190). Accordingly, he is reported to have held that human dignity is a subcategory of the human sanctity (Lichtenstein 2016).

In cases where no legal opinion (*halacha pesuka*) on a particular practice has been established, there is a principle – *pek hazji mai ama davar* – that “one can learn from the accepted practice of the people, as it has its basis in tradition.”⁴³ This principle, coincidentally, was used to justify slavery in the antebellum south.⁴⁴ Maimonides here, though not calling for the abolition of slavery, rejects applying this principle to justify less than benevolent treatment of a slave. He argues that treating a slave as a mere means simply because such treatment is “accepted practice” among the nations of the world, finds no justification in biblical law. As proof, he brings the biblical verse, which reads in its entirety: “And what great nation is there that hath such just statutes and judgments as all this law, which I set before you this day?”

To be clear, this is not some parochial diatribe against non-Jews,⁴⁵ but rather part and parcel of Maimonides’ argument for just relations with one’s slave. He argues against following the customary practices of other peoples, grounding his claim in the biblical verse that extols the justness of the laws of the Jewish people – this “great nation” – above the laws and customs of all other peoples. For biblical law is “just,” in the philosophical sense, demanding balance. Indeed, the support verse here – “*just statutes*

⁴³ Hayun (1996). Of course, “the people” to look to are the Jews, and “the tradition” to emulate is that of the Jews; however, these traditions may very well have been accepted from non-Jews – e.g., doctors wear white coats (see Shulchan Aruch YD 178:1, Igrot Moshe YD 4:12, Shu”t Maharik 88); monogamy (as promulgated by Rabbeinu Gershom’s famous rulings was based on the status of woman in non-Jewish society – see Falk 1961: 31; also Grossman 2001). See also Dorff and Crane (2012: Intro.). In most times and places, though, Jews have done something in between, accepting influences from the outside only in part and in a particularly Jewish way. For example, the Rabbis of the Mishnah adapted Roman family and business law when formulating Jewish laws on those topics; similarly, many contemporary decisions in medical ethics.

⁴⁴ See David Einhorn’s attack against this justification (“War on Amalek” in Saperstein 2012: 211-212, esp. fn. 51-52).

⁴⁵ Worthy of note is the great esteem in which Maimonides holds non-Jewish thinkers, frequently quoting the ideas of Aristotle, Al Farabi, Galen, etc. (see Pines, “The Philosophic Sources of The Guide of the Perplexed” in Maimonides [1190] 1963: lxxviii-cviii). So, e.g., “He, Aristotle, arrived at the highest summit of knowledge” (Maimonides [1199] 1872: 228); “Study ... Al Afarabi for everything he has written ... is as fine flour” (Maimonides [1199] 1872: 226-227). See also Twersky (1980: 367) who notes that Maimonides frequently declared the “indispensability of non-Jewish sources.”

and judgments” – is precisely the same verse Maimonides uses in his Guide to explain that just actions are balanced actions, and just dispositions are balanced dispositions.⁴⁶ And it is this balanced approach, i.e., through the observance of the “just statutes and judgments,” that leads the people of Israel to the virtue of compassion, indeed, to be “compassionate to all.”⁴⁷

Two points to take note here:

- (1) This biblical verse – with its broad claim that the entire law is grounded in just laws – emphatically supports the claim that justice is here demanded by the law itself.
- (2) The argument for justice here is based on the value inherent in just moral behavior, irrespective of the value of the slave.

(5) Support for Piety

Accordingly, regarding the divine attributes, which He has commanded us to imitate, the psalmist says: “His tender mercies (rachamav) are over all His works” (Psalms 145:9).

Maimonides here brings one of the most fundamental ideas in Jewish moral philosophy, the notion of *imitatio Dei* as learned from the verse, “And you shall walk in His ways” (Deut. 28:9).⁴⁸ Based on this verse, Maimonides is understood to argue for the moral approach of virtue ethics⁴⁹ – i.e., that there is a “religious obligation to develop virtuous traits of character” (Wurzbürger 1994: 69). Leaving nothing to the imagination,

⁴⁶ See fn. 32 above.

⁴⁷ Twersky understands that the compassion denoted here is supererogatory, a demand that would be incongruous with the reference to “statutes and judgments” which are mandatory (i.e., not supererogatory). He resolves the issue by noting that “all the laws are a springboard for the highest morality” (1980: 428) – i.e., the mandatory laws lead to supererogatory behavior. While this is true, it is possible to see the compassion here as that demanded by justice (i.e., “ways of the wise”). For, cannot just laws (e.g., surrounding loans to the poor) be considered compassionate? And does not act-morality seek virtuous behavior as does agent-morality?

⁴⁸ See Maimonides (1984: Pos. #8); Mishnah Torah (Hil. Deot 1:6).

⁴⁹ See Blau (2000), Spero (2003), Shatz (2005), Wurzbürger (2008: Part I, esp. 33-39, 59), Shatz (2012), Bedzow (2017: 13), Shapira (2018).

Maimonides cites a verse to remind us that chief among the divine attributes to imitate is that of mercy/compassion (*rachamim*)⁵⁰ – precisely the quality he has been demanding be employed in one’s interrelationship with one’s slave. It is a demand to act beyond the strict letter of the law – a demand to act with piety (*b’midat hassidut*).

Two points to note here:

- (1) The biblical verse used to argue for mercy/compassion, in conjunction with the unstated verse (Deut. 28:9) demanding *imitatio Dei*, together support the claim that piety is here demanded by the law itself.⁵¹
- (2) The argument for piety here is based on the value inherent in pious moral behavior, irrespective of the value of the slave.

(6) Hypothesis Conclusion

*Whoever is merciful will receive mercy, as it is written: “He will allow thee to be merciful and show mercy unto thee and multiply thee” (Deut. 13:18).*⁵²

Maimonides concludes his arguments with another fundamental notion in Jewish thought known as “measure for measure” (*midda keneged midda*), which simply states that in the measure, or manner, that you act towards others, so too, in the same measure, will God act towards you. The Talmud teaches this ethic in a number of places⁵³ and applies it specifically to the virtue of mercy, quoting the biblical verse as Maimonides has it

⁵⁰ See Guide (1:54; 3:54).

⁵¹ Thus, we have a law to go beyond the law. Wurzburger (1994: 79; 2008: 69, 103) notes that while Maimonides does not command supererogatory acts, he does see them as part of the command to “walk in His ways” – precisely as we have here. Similarly, Soloveitchik (2017: 179) and Shabbtai Sofer (Shaarei Deah, Hil. Deot 1:6) bring this last law of slaves as the quintessential example of a commandment to “walk in His ways”! See also Lichtenstein (2004b: 42), notwithstanding his difficulties distinguishing the “golden mean” and “lifnim.”

⁵² Interpretative translation according to R. S. R. Hirsch (1989). See also below fn. 55.

⁵³ Sotah 1:7, Shabbat 105b, Nedarim 32a, San. 90a.

here.⁵⁴ This biblical verse he brings is interpreted by the Talmud to mean that God gives you the trait of mercy such that when you apply it and are merciful, He too, in kind, will be merciful to you.⁵⁵

How this serves as a conclusion to Maimonides' "geometric" proof for virtue will become clear according to the reader – traditional or philosophical – being addressed.

The Traditional Reader: Virtue for All

By employing what is arguably a geometric proof, Maimonides makes an incontestable argument for virtuous behavior with one's slave and, by extension, for virtue in general. But beyond his logical proof for virtue, Maimonides makes his appeal personal, directing it to individuals, irrespective of what "stage on life's way" they might be. That is, his arguments for virtuous behavior can be seen as addressing the individual within each of the three stages of life delineated by Kierkegaard: the aesthetic, the ethical and the religious.⁵⁶

According to Kierkegaard, the individual starts in the aesthetic stage wherein one is concerned with oneself, with satisfying one's own instincts and urges⁵⁷ – i.e., driven by aesthetic concerns. If, through self-reflection, one succeeds in understanding the place

⁵⁴ Shabbat 151b. Also Midrash Hagadol, Midrash Tanaaim (Deut. 13:18) – though these texts may simply be sourced in Maimonides himself as these Midrashim are known to have incorporated Maimonides' own words (for a discussion on the composition of these Midrashim see, e.g., Bar-Asher Siegal and Shmidman, 2018).

⁵⁵ For the Talmudic interpretation of the verse, see R. S. R. Hirsch (Deut. 13:18), Torah Temimah (Deut. 13:18), Hevruta (Shabbat 151b). It should be noted that many commentators read the "*psbat*" differently than the Talmudic interpretation (see, e.g., Bechor Shor, Daat Zekanim, Baal Haturim, Ohr HaHaim [Deut. 13:18]).

⁵⁶ Note: I employ Kierkegaard's "stages of life" here as a hermeneutical tool to clarify Maimonides' line of reasoning (and not to imply any connection between the two thinkers or their thought). See also fn. 59 further herein.

⁵⁷ Broudy (1941: 294).

of the aesthetic in one's life, in sublimating it rather than being driven by it,⁵⁸ one then reaches the ethical stage. Here, one considers oneself in relation to others, one considers the norm – i.e., one is driven by ethical concerns. Success in the ethical stage brings one to the highest stage of human development, the religious. This is the stage wherein one overcomes all his personal motivations in favor of the demands of the Creator – i.e., driven by religious concerns.

While Maimonides does not explicitly define such stages, he does speak of the object of the Law addressing these three stages (Guide 3:33). He refers to the aesthetic stage in explaining that it is an object of the Law to encourage the reduction, or outright rejection, of pursuing aesthetic desires. He refers to the ethical stage by noting that it is the object of the Law to promote virtuous interaction between people. Finally, Maimonides can be seen as referring to the religious stage in explaining that the Law comes to promote the sanctification of its followers, bidding them to be holy.

However, more interesting than the parallel between Maimonidean “objectives of the Law” and Kierkegaardian “stages of life,” is how Maimonides’ call to virtue – in this last law of slaves – addresses the individual at each of the three Kierkegaardian stages. That is, Maimonides makes his arguments relevant regardless of what “stage on life’s way” one might be. Indeed, this need to address each stage comes to explain why he, in arguing for piety and justice, brings not one argument for each but two. As described above, both section (2) and section (5) argue for piety, whereas both section (3) and section (4) argue for justice. One could assert that either the first pair [(2), (3)] or, alternatively, the second pair [(4), (5)] suffice to make the point. However, as noted at the end of each section above, (2) and (3) make the argument for piety and justice based on the inherent worth of the slave, whereas (4) and (5) make it based on the inherent value of moral behavior. Accordingly, (2) and (3) speak to the individual at the ethical stage of life, whereas (4) and (5) speak to him at the religious stage of life – as will now be explained.

⁵⁸ Navon (2008).

(2-3) Ethical

Starting with the ethical, being that it is the universal – applying to all and in which all struggle (Kierkegaard 1985: 83), Maimonides begins by enjoining piety based on the dignity inherent in the slave as a human being, as a fellow other. Accordingly, sections (2) and (3), devoted to making the claim for virtue based on the humanity of the slave, can be seen as addressing specifically the individual at the ethical stage “on life’s way.” For, as mentioned, the ethical stage is that in which one struggles to overcome one’s own selfish desires for the sake of the other. Maimonides addresses the individual at this stage – specifically in sections (2) and (3) – by prescribing practical actions to be taken as ethical norms.

The norms posited here, however, are not random pieces of practical advice but rather constitute an exemplary list of directives that addresses all human needs as delineated by Abraham Maslow in his seminal essay, “A Theory of Human Motivation” (1943).⁵⁹ Maslow explains that for one “to become everything that one is capable of becoming,” he must first satisfy his “basic” needs. These needs follow a hierarchy of four stages: physiological needs, safety needs, love needs, esteem needs. Only when these have been satisfied can one pursue the ultimate “being” need that leads to self-actualization.

Reviewing the five practical demands that Maimonides places on the master reveals an uncanny correspondence to the five levels of needs delineated by Maslow:

- *He should give him to eat and drink of every food and drink.* This directive enjoins the master to fulfill his slave’s most basic of needs, “food and drink” being representative of what Maslow refers to as the “physiological needs.” Indeed, Maslow uses food as the quintessential physiological need, explaining of the hungry man that “all [other needs will] be waved aside as fripperies which are useless since they fail to fill the stomach.”

⁵⁹ As an important aside, I am not implying that Maimonides preceded Kierkegaard or Maslow in categorizing life’s stages or human needs, nor do I imply that they referred to him. Rather, I am reading Maimonides through the prism of the fundamentals of the human condition that Kierkegaard and Maslow simply articulated in order to demonstrate the logic underlying Maimonides’ arguments.

- *And they would provide food for their animals and slaves before partaking of their own meal.*⁶⁰ Here Maimonides intimates that the master place the needs of those not in charge of their own food before his own, thus ensuring what Maslow calls the “safety” needs of the slave – the need for “a safe, orderly, predictable, organized world, which he can count on.” By calling on the master to feed his slave before himself, Maimonides is asking that the slave be provided with exactly that – i.e., a safe, orderly and predictable environment.
- *The sages of old had the practice of sharing with the slave every dish they ate.* Maimonides here suggests that the master go beyond simple physiological and safety needs, asking of him to essentially share his table with his slave, to befriend him. As such, this demand can be seen as corresponding to the “love and belongingness” needs which Maslow describes as follows: “If both the physiological and the safety needs are fairly well gratified, then there will emerge the love and affection and belongingness needs, ... He will hunger for affectionate relations with people ...” While one could argue about how deep such “affectionate relations” are made through the gesture of “sharing every dish,” there can be no denying that it goes far beyond satisfying physiological and safety needs and, at the minimum, promotes an affable association.
- *Nor should a master disgrace his slave, neither physically nor verbally.* The directive here calls for the master to respect the dignity of the slave, corresponding to the fourth of Maslow’s basic needs – the “esteem” needs. Maslow writes: “Satisfaction of the self-esteem need leads to feelings of self-confidence, worth, strength, capability and adequacy of being useful and necessary in the world.” Now, though Maimonides’ claim on the master is quite minimal and posed in the negative – i.e., *do not disgrace* – it bespeaks, I suggest, of positively promoting respect and self-esteem.
- *And one should not treat him with constant screaming and anger, but rather speak with him calmly and listen to his complaints.* These words can be seen as a continuation of requiring the master to ensure that the slave’s esteem needs are met. However, the demand to *listen to his complaints* also points to an awareness of the slave as an autonomous agent – i.e., he has complaints, an opinion, a voice. Accordingly, this demand can be seen as hinting at, even providing for, the ultimate need: “the need for self-actualization.” Of course, a slave is not going to have the necessary freedoms to self-actualize, but

⁶⁰ Note: That this stage is second in Maslow’s hierarchy, whereas the quote by Maimonides appears third, does not detract from the thesis here that Maimonides is addressing every one of the human needs.

the very fact that he is given the opportunity to express himself, his needs, certainly reflects a high level of autonomy for a slave.

In these two directives (2-3), based on the inherent value of the slave, Maimonides has delineated a clear program for ethical behavior, one that, in addressing the hierarchy of human needs, clearly fulfills Kant's Formula of Humanity.

(4-5) Religious

Moving to the higher motivation of the religious – i.e., obeying God – Maimonides calls for the master (4) to exhibit justice because he is a God fearing individual who, like “*our father Abraham*,” accepts the divine ethical norms of the Bible to do “*tzedek umishpat*” (Gen. 19:18); and (5) to exhibit piety because he is to imitate the attributes of the creator, mercy being primary among them.⁶¹ That is, the master, irrespective of the value of the slave, has an absolute moral obligation – based as it is on divine command – to be a virtuous individual. Accordingly, sections (4) and (5), devoted to making the claim for virtue based on the inherent value of moral behavior, can be seen as addressing specifically the individual at the “religious” stage “on life’s way.”

To be clear, the “religious” is that stage in which one defers to God *in toto*. According to Kierkegaard, this entails the teleological suspension of the ethical. One is to step outside oneself, to take a “leap of faith,” and obey God’s command above all personal considerations. Maimonides, and indeed Jewish thought in general, aligns with this definition of the “religious” with one caveat. It is not the suspension of the ethical that is demanded here but, as notes R. Soloveitchik (2008a: 190), the suspension of personal judgement. That is, while obedience to God is indeed held as the ultimate stage on life’s way, God’s commands can never go against the ethical. Using Kierkegaard’s example of Abraham as knight of faith, Jewish thought sees Abraham’s feat not in setting aside the ethical in favor of the divine, but in setting aside personal judgment. Abraham, in his

⁶¹ The two demands could be seen to reflect the two levels of the “religious” articulated by Kierkegaard (see Broudy 1941: 306).

willingness to follow God at all costs, acknowledged that human finitude demands that one defer to God's infinite wisdom, that one suspend judgement.⁶²

It is in this light of the “religious” that Maimonides makes his demands of the master in sections (4) and (5). He begins by rejecting the ways *the heathen who worship idols* – i.e., the ways people behave by reason of their own limited judgement. He then enjoins the master to follow the example of *our father Abraham* – the paragon of faith who suspended judgement to follow God's will. Finally, as described above, he calls for *imitatio Dei*, that one seek the “religious” – the unflinching obedience to God's will – not only in the strict adherence to the command of God but in the devoted emulation of the merciful attributes of God.

(6) Aesthetic

Maimonides concludes with an appeal to self-interest (i.e., the Kierkegaardian “aesthetic” stage). Noting the principle of measure-for-measure, he argues that, in essence, even if one is not moved by the higher motivations of ethics and religion, one should act mercifully in order that he too will be treated mercifully.

Accordingly, even if one does not appreciate the ethical and religious value of virtue, one will certainly appreciate the self-serving value of God's mercy. And, it should be noted, this call to mercy is made independent of the inherent worth of the slave. This final appeal, then, pleads for virtue saying: though you may not recognize the worth of your slave as a human being qua human being, nor even the worth of your own character, let alone the value of your relationship with God, at least recognize your own need for mercy and be merciful.

This is the Traditionalist way of reading Maimonides' appeal to virtue in this last law of slaves. It makes claims that any student learned in Talmudic reasoning can follow and appreciate. There is however a Philosophical way to read this appeal to virtue, one that discerns, beneath the unimpeachable halachic argumentation, a flow of ideas articulated

⁶² For a full elaboration of this idea, see Navon (2014).

by the chiasmic structure that outlines a path to moral development and, indeed, to human perfection.

The Philosophical Reader: Chiasmus

The chiasmic form employed by Maimonides in this last law of slaves is not simply aesthetic but didactic,⁶³ visually articulating the notion that the path to perfection begins with piety and ends with piety, justice being the bridge. This is not to say that justice is ever abandoned, for justice is the basis of piety – i.e., there must first be a law from which one goes beyond. Indeed, without the binding framework of “justice” – i.e., the law – “piety” becomes the unguided subjective whim of the agent (see, e.g., Soloveitchik 2008b: 54-55).

(2) Piety as Initiation

Now, before going beyond the law (i.e., piety), and even before going according to the law (i.e., justice), one must be inspired to accept the law. The inspiration, I suggest, comes from within, from a desire to rise above the meaninglessness of mundane existence – as the verse teaches: “From there you will seek the Lord your God, and you will find Him ...” (Deut. 4:29). And while this verse refers to the motivation to seek God from the “*tzarot*” (crisis) entailed in national exile,⁶⁴ Soloveitchik (1993: 80-82) teaches that the motivation can be external or internal, from “surface crisis” or from existential “depth crisis.”

Importantly, it is the motivation to seek God that, according to Soloveitchik, motivates one to seek ethical change: “Whenever a person beholds God, an inner catharsis compelling a complete change of one’s axiological hierarchy must occur” (2017: 182). This “beholding” of God is not necessarily ecstatic but existential – a realization of a purposeful Creator, one Soloveitchik calls, “Real and Perfect.” And the consequent

⁶³ “Considerations of symmetry and order [are] well known to be important to Maimonides” (Kellner 1990: 20).

⁶⁴ See, e.g., Seforno (ad loc.).

axiological change does not necessarily drive one to the library but to God Himself, for “whoever has seen the Real and Perfect is determined to be real and perfect too, since in his eyes, imitating is the only way in which he may commune with the Real and Perfect” (ibid).

This “imitating” – known as *imitatio Dei* – is “the foundation of morality.”⁶⁵ But how are we to imitate a God of Whom we have yet to learn His ways? The answer, suggested in the call to piety (2) in this law of slaves, is by imitating His sages; for it is they who provide a reflection of the divine, as Maimonides (Hil. Deot 6:2) teaches explicitly:

“It is a positive commandment to cleave to the wise and their disciples, in order to learn from their deeds, as it is said: ‘And to Him shalt thou cleave’ (Deut. 10:20). Is it possible to cleave to the Shechinah? But even thus the wise men commented upon in interpreting this commandment, saying: ‘Cleave to the wise men and their disciples’ (Ketubot 111b).”⁶⁶

Accordingly, Maimonides’ first appeal to piety (2) is an appeal to begin the path to perfection through what could be called *imitatio sophos*⁶⁷ – i.e., imitation of sages who are called “sages” by virtue of their own *imitatio Dei*.

Interestingly, Maimonides parallels the two types of imitation in that this first statement calling for piety (2) refers us to the ethic of imitating the sages, while the last statement calling for piety (5) refers us to the ethic of imitating God. This should come as no surprise as the Gemara itself parallels the two in answering the question: “Seeing that God is a devouring fire, how is it possible to cleave to God?” One answer is *imitatio Dei* (Sotah 14a) and one is *imitatio sophos* (Ketubot 111b).⁶⁸

⁶⁵ Soloveitchik (1993: 26), Wurzbürger (1994: 77).

⁶⁶ On the novelty of Maimonides here see Wurzbürger (1994: 75).

⁶⁷ Schwarzschild calls it “*imitatio imitationis Dei*” (1990: 149).

⁶⁸ On the importance of *imitatio sophos*, see Avot 6:5; Berachot 7b, 62b; Makkot 22b; Wurzbürger (2008: 31, 38, 70, 169); Soloveitchik (2017: 195-201).

And while Steven Schwarzschild belittles imitating sages as “*imitatio Dei* for the masses” (1990: 149), I would contend that it is appropriate for anyone who has not yet reached the heights of *imitatio Dei* directly. In consonance, Wurzbürger explains, “it is through obedience to the law, which includes provisions for the imitation of the conduct of appropriate role models (i.e., scholars of the law), that one cultivates intellectual and moral values and thus advances on the road towards the ‘knowledge of the Lord’” (2008: 70).

(3) Justice as Preparation

And that brings us to “obedience to the law.” Having observed piety from the pious, one may have been able to assimilate some of their ways but cannot yet accept them *in toto* as a way of life – such a leap is too great.⁶⁹ Rather, having gained an appreciation for the path, one is ready to accept the “ways of the wise,” the balanced path of the law. Accordingly, Maimonides writes: “It is characteristic of every human being that when his interest is engaged in the ways of wisdom and justice (*tzedek*), he longs for these ways and is eager to follow them” (Hil. Tesh. 6:5).

At this stage, then, one embarks on the journey of practicing the Law, the Torah, the halacha.⁷⁰ The law dictates actions that are moral in and of themselves and also instrumental in cultivating moral dispositions⁷¹ – as Maimonides notes, “Know that good and bad qualities can only be cultivated by repetitive acts” (Mishna Avot, Introduction). Accordingly, Maimonides’ first appeal to justice (3) refers explicitly to “biblical law” as an appeal to moral practice that will cultivate moral character for, ultimately, it is actions that shape character.⁷²

⁶⁹ See, e.g., Aiken (2009: 82-83).

⁷⁰ See fn. 29 on the law inculcating the mean (i.e., “justice”).

⁷¹ See, e.g., Twersky (1980: 300); Wurzbürger (1994: 69).

⁷² See, e.g., Aristotle (2004: 23), Sefer HaHinuch (#16), Leibowitz (1986: 178-182), Bedzow (2017: 32, 108).

This stage of practicing of the Law, it must be noted, is one done without great understanding of what underpins the law and its practical demands – it is the “we will do” (*naaseh*) practice that precedes the “we will understand” (*nishma*) practice (Ex. 24:7).

Intellection as Inflection

And this brings us to the inflection point in the chiasmic structure. Having engaged a bit in the “ways of the pious” and a bit in the “ways of the wise,” one is ready to embark further in these ways, in reverse order – i.e., practicing more ways of the law (i.e., justice), and consequently, realizing more of how and when to go beyond the law (i.e., piety). But how is this to take place? What is the difference between the first two stages (2-3) and the second two (4-5), other than their reversed order?

In the last chapter of his *Guide for the Perplexed* (3:54), Maimonides explains that there are four areas in which one must succeed in order to gain perfection (i.e., wealth, health, morals, and intellect), the highest two – i.e., the moral and the intellectual – being what distinguishes human beings qua human beings.⁷³ The moral stage is explicitly designated as lower than the intellectual, thus giving rise to what Kellner (2009: 70) calls the “standard” reading of Maimonides’ path to perfection.⁷⁴ That is, Maimonides is understood to hold intellectual perfection as the ultimate goal, moral perfection simply being an indispensable prerequisite,⁷⁵ nothing more than a “stepping stone.”⁷⁶ That said, not a few have noted that Maimonides intimates that moral behavior consequent to

⁷³ Maimonides (*Guide* 3:54) notes that in perfecting the moral virtues “man is more closely connected with man himself,” whereas in perfecting the intellectual virtues “man is man.” Intellectual achievement is seen as more uniquely human than moral achievement for moral behavior is found in “the lion’s courage and the cock’s generosity” (Ibn Baja in Altmann 1972: 23). That said, surely there is a profound difference between “generosity” done by an animal, or unreflectively by a human, versus that performed with understanding and/or in imitation of divine generosity.

⁷⁴ See Kellner (1990: 67 n. 20) for a list of those who maintain the standard position.

⁷⁵ *Guide* 1:34; for further sources in Maimonides, see Kellner (2009: 67 fn. 17).

⁷⁶ As per Ibn Baja (see Altmann 1972: 19-20).

intellectual achievement may just be the ultimate human perfection, and if not a greater perfection than that of the intellectual, then of great consequence nonetheless.⁷⁷

Moral perfection, then, can be understood as two-fold: pre-theoretic and post-theoretic.⁷⁸ Pre-theoretic morality denotes the stage wherein one follows moral law without an intellectual understanding of the theory underpinning the law – “without an understanding of their divine foundation” (Frank 1985: 491). This stage of moral action is also referred to as “propaedeutic,” for it serves as a necessary preparation for and prerequisite to the intellectual stage of theoretical understanding.⁷⁹ However, once one has actualized oneself in the higher perfection of theoretical understanding, one’s true perfection is to be found in acting morally with understanding, what is referred to as post-theoretic or, as David Shatz puts it, “consequent” morality.

Accordingly, the program for moral development toward the attainment of human perfection is such that one first accepts and performs moral laws as given, then studies them in depth, reaching therein for an understanding of their Source (i.e., God),⁸⁰ whereupon, having been transformed in intellectual achievement,⁸¹ one performs them with understanding or with, what might be called, divine influence. This program is found in Maimonides’ own words: “And this should be the order observed: The opinions in question should first be known as being received through tradition; then they should be demonstrated; then the actions through which one’s way of life may be ennobled, should be precisely defined” (Guide 3:54). On this Albert Friedberg (2019: 10) explains:

Opinions – and by opinions Maimonides means philosophical propositions – are transmitted by the Torah ... and should be accepted on simple reading, as is ...

⁷⁷ For a study of this issue, see Ch. 4 “Polemics on Perfection.”

⁷⁸ These terms are found in, e.g., Frank (1985: 491), but the notion itself is found in many writers, see fn. 82.

⁷⁹ See, e.g., Twersky (1980: 300), Shatz (2005: 169).

⁸⁰ Torah study is to include halacha, physics and metaphysics (as per Guide, Introduction quoted herein below. See also, e.g., Hartman (1986: 205), Kellner (2009: 37), Nadler (2021: 275).

⁸¹ Kellner (1990: 35, 39, 41; 2009: 71-72) emphasizes that the intellectual achievement is not only about gaining knowledge, but about “transforming” the individual. Similarly, Soloveitchik (2017: 182).

At a second stage, the student demonstrates these propositions by way of scientific/philosophical reasoning ... [At a third stage,] the student is urged to discover the actions and ways that will ennoble his life. That is, proper ethical behaviour depends on the correct apprehension of the intelligibles and God, ... To highlight the programmatic aspect of this interpretation, Maimonides insists that these three steps should be taken “in this order.”

In the end, Friedberg, like many others,⁸² concludes that for Maimonides, “Philosophical inquiry is the necessary prerequisite for the ethical life” (ibid.). This philosophical inquiry, otherwise referred to as “intellection,” is one that Maimonides himself describes, and includes in it the study of the Torah, physics and metaphysics:

“God, may His mention be exalted, wished us to be perfected and the state of our societies to be improved by His laws regarding actions [i.e., the Torah]. Now, this can come about only after the adoption of intellectual beliefs, the first of which being His apprehension, may He be exalted, according to our capacity. This, in its turn, cannot come about except through Divine science [i.e., metaphysics], and this Divine science cannot become actual except after a study of natural science [i.e., physics]” (Guide, Introduction).

Returning to our analysis of the last law of slaves, the phase of intellection is what serves as the unwritten – for intellection is not an act *per se* – inflection point of the chiasmus, transforming the moral practitioner to ready him for a new phase of personal development.

⁸² Guttman (1964: 200 in Hartman 1986: 202), Altmann (1972: 24), Cohen (1978 in Kellner 1990: 69, fn. 3), Twersky (1980: 362-363), Frank (1985: 494), Hartman (1986: 205), Schwarzschild (1990: 145), Davidson (1992: 86 in Shatz 2005: fn. 64), Wurzbürger (1994: 78-79), Harvey Kreisel (1999 in Shatz 2005: fn. 57), Shatz (2005: 184), Wurzbürger (2008: 98), Goodman (2009: 461), Kellner (2009: 71-72), Ravitsky (2014: 47), Bedzow (2017: 109-110). Regarding the varying opinions of the post-theoretic phase, see Ch. 4 “Polemics on Perfection.”

(4-5) Justice & Piety as Dual-moral Approach

This new phase is that of the intellectually enhanced moral practice of justice and piety. Maimonides thus reiterates the otherwise redundant demands to follow (4) the ways of justice and (5) the ways of piety. Here, then, he gives expression to the notions of pre- and post-theoretic morality alluded to in his Guide (3:54). Regarding pre-theoretic morality he writes:

“The third species ... is the perfection of the moral virtues ... [which is] a preparation for something else and not an end in itself. For all moral habits are concerned with what occurs between one individual and another. This perfection regarding moral habits is, as it were, only the disposition to be useful to people; consequently, it is an instrument for someone else. ... the moral habits that are useful to all people in their mutual dealings – that all this is not to be compared with this ultimate end [of apprehending God] and does not equal it, being but preparations made for the sake of this end.”

In consonance, the law of slaves' first two calls for moral behavior (2, 3) are based on the value of the other (i.e., the slave), on “what occurs between one individual and another.” In contrast, the latter two calls for moral behavior (4, 5) are based on the intrinsic value of moral behavior as a perfection of the individual himself; something that aligns with the post-theoretic morality understood in the words of the Guide (ibid):

Jeremiah says ... “*Thus saith the Lord: Let not the wise man glory in his wisdom, neither let the mighty man glory in his might, let not the rich man glory in his riches; but let him that glorieth glory in this, that he understandeth, and knoweth Me, that I am the Lord who exercise mercy, justice, and righteousness, in the earth; for in these things I delight, saith the Lord.*” ... When explaining in this verse the noblest of perfections, he [i.e., Jeremiah] does not limit them only to the apprehension of Him, may He be exalted. ... But says that one should glory in the apprehension of Myself and in the knowledge of My attributes, by which he means His actions, ... In this verse he makes it clear to us that those actions that ought to be known and imitated are *loving-kindness, judgment, and righteousness (besed, mishpat u'tzedeka)* ... Then he completes the notion by saying: *For in these things I desire, saith the Lord.* He means that it is My purpose that

there should come from you *loving-kindness, judgment, and righteousness in the earth* in the way we have explained with regard to the *thirteen attributes*: namely, that the purpose should be to imitate them and that this should be our way of life.

Strikingly, then, the last law of slaves uniquely expresses the pre- and post-theoretic morality alluded to in the Guide.⁸³ Maimonides, thus gives voice to two theories about morality: the theory of “morality-as-cooperation”⁸⁴ – which holds that morality is simply an instrument of evolution; and the theory of, what I call, morality-as-imitating-God – which holds that there is an intrinsic value in moral behavior. Maimonides not only gives voice to these two theories but finds a place for *both* in human development toward perfection.

But there is even more we can learn about pre- and post-theoretic morality from this last law of slaves. As noted, whereas the pre-theoretic stage consisted of piety (as initiation) leading to justice (as preparation), now, following the intellection stage, the order is reversed: first justice and then piety. The reason, I suggest, is that now one must begin by performing the law with understanding before one can go beyond it with understanding. Twersky (1980: 428) explains it as follows:

“... all the laws are a springboard for the highest morality and perfection which emanate slowly and steadily from them. Just as one embraces reality in order to transcend it, one adheres to the law in order that it may enhance one’s perception of the good and the true and induce behavior which transcends the letter of the law. In short, law alone, in a formal sense, is not the exclusive criterion of ideal religious behavior, either positive or negative. It does not exhaust religious-moral requirements. There is rather a continuum from clearly prescribed legislation to open-ended supererogatory performance....”⁸⁵

⁸³ This then, *pace* Shatz (2005: 188), is a reference, albeit implicit, to consequent morality in Maimonides’ legal writings.

⁸⁴ See, e.g., Curry 2016.

⁸⁵ Similarly Wurzburger (1994: 27-28, 37), Bedzow (2017: 11).

Accordingly, as opposed to the pre-theoretic development of morality, the post-theoretic development starts with (4) the mandatory (i.e., justice) that then leads to (5) the supererogatory (i.e., piety).

There is yet one more important point that the last law of slaves has to make regarding the path to moral perfection. As noted above, the path to perfection as understood by readers of Maimonides' Guide is thought to be three-fold, what might be depicted schematically as follows:

pre-theoretic morality > intellection > post-theoretic morality

However, surely it would be the height of hubris (or naivete) to conclude that moral perfection could be achieved so simply – i.e., that the path to moral perfection requires a mere single pass. Accordingly, many have noted Maimonides' path to perfection is an iterative one. Such is implicit in the words of Isadore Twersky: “Maimonides believed that knowledge stimulates and sustains proper prescribed conduct [i.e., halachic practice] which in turn is a conduit for knowledge, and this intellectual achievement in return *raises* the level and motive of conduct” (Twersky 1980: 511). The idea is more explicit in the words of David Hartman: “The contemplative ideal is not insulated from Halakhah, but affects it in a new manner. Sinai [i.e., halachic practice] is not a mere stage in man's spiritual development, but the ultimate place to which man *constantly returns*—even when he soars to the heights of metaphysical knowledge” (Hartman 1986: 26, emphasis added). But no one states it more definitively than Wurzbürger:

It must be realized that in the Maimonidean system, “thou shalt walk in his ways” represents a *continuous* challenge, beginning with the attempt to cultivate moral virtues through moral conduct and pointing to the *ever higher* dimensions of *imitatio Dei* which can be engendered only by intellectual perfection.⁸⁶

That Maimonides' program for human perfection is an iterative one is, most remarkably, reflected in the chiasmic structure of the law on slaves:

⁸⁶ Wurzbürger (2008: 99, emphasis added). Additionally, Wurzbürger (1994: 83) explains that *imitatio Dei* includes both “the way of the wise” and “the ethics of the pious.”

piety > justice > intellection > justice > piety

That is, upon reaching a level of piety after intellection, one goes back to the beginning – attempting further gains in piety through *imitatio sophos*, further commitment to the performance of the law, further investigation of that which lies behind the law – all to cultivate higher levels of justice and piety. The cycle continues with the ultimate – and ultimately unattainable – aspiration of imitating God’s ways in all their glory, “not *becoming* like God, but *acting* like God” (Kellner 1990: 42).

Maimonides has thus encapsulated his program for moral perfection within the chiasmic structure of the last law on slaves – calling for the dual-moral approach to be developed through a lifetime of striving to imitate God.⁸⁷

(6) Purpose Achieved

The last law on slaves concludes with the words: *Whoever is merciful will receive mercy, as it is written: “He will allow thee to be merciful and show mercy unto thee and multiply thee” (Deut. 13:18).* Now, while the “Traditional” reader will take this “measure-for-measure” appeal at face value (i.e., an instrumental argument of self-interest), the “Philosophical” reader will discern a culmination of the program for perfection.

To begin, note that this statement (6) is almost verbatim from the Talmud (Shabbat 151b), except for one omission – “to the creations.” That is, whereas Maimonides writes, “Whoever is merciful will receive mercy,” the Talmud teaches, “Whoever is merciful *to the creations* will receive mercy.” It seems clear that Maimonides seeks to emphasize disposition over deed. For, while it has been argued herein that there is a need for both agent morality and act morality, Maimonides is here making his concluding argument for the state of the individual, for the perfection of man.

⁸⁷ For a comprehensive review of Maimonides’ approach to *imitatio Dei* and human telos, see Ch. 4: “Polemics on Perfection.”

Furthermore, Maimonides makes this exact same statement (6) in only one other place: the Laws of Gifts to the Poor (10:2). There, as here, he emphasizes that mercy is one of the essential dispositions of the Jewish people. The importance of expressing mercy, especially to the poor, is brought by the Sefer HaHinuch (#66) who adds that such a disposition is really God's purpose in creation:

... The root of this commandment [to loan to the poor] is that God wanted His creations to be trained and habituated to the trait of kindness (*bessed*) and mercy (*rachamim*), since it is a praiseworthy trait [i.e., *bessed* and *rachamim* together]. And from the refinement of their bodies with good character traits, they will be fit to receive the good; as we have said that the good and blessing always descend upon the good [individual], and not upon their opposite. And when God, may He be blessed, does good to the good, He fulfills His desire, since He desires to do good to the world.⁸⁸

In essence, God created man to be merciful (i.e., it is of man's purpose to be merciful), for it is "His desire" (i.e., God's purpose) to be merciful to man.⁸⁹ Accordingly, Maimonides' concluding appeal to mercy is not merely an appeal to human self-interest, but an articulation of the alignment between the purpose of man and the purpose of God. The supporting verse – "*He will allow thee to be merciful and show mercy unto thee*" – thus reads: God endows man with the trait of mercy – the expression of which is man's ultimate purpose – so that He could bestow His mercy – the expression of which is God's ultimate purpose.

This last statement (6), then, is a most appropriate conclusion to the hypothesis that one must apply one's complete moral presence, via the dual-moral approach and following the iterative program through intellection, to realize the ultimate goal of being a wise-merciful individual. In so doing, one fulfills one's own purpose in creation, which is, at one and the same time, God's purpose in creation.

⁸⁸ See also Luzzatto (1983: 1:2:1), Saadia Gaon (1976: 1:4 end, 3). For further sources see Aryeh Kaplan's note to Luzzatto (1983: Part One, fn. 8).

⁸⁹ This aligns quite seamlessly with Maimonides' words at the end of his Guide (3:54), see Ch. 4 "Polemics on Perfection" (sec. On Perfections and Overflows).

Accordingly, this concluding message on human perfection in the last law of slaves finds itself in perfect harmony with the concluding message on human perfection in the Guide for the Perplexed:

... the perfection, in which man can truly glory, is attained by him when he has acquired – as far as this is possible for man – the knowledge of God, the knowledge of His Providence, and of the manner in which it influences His creatures in their production and continued existence. Having achieved this apprehension [of God] one will then be determined always to seek loving-kindness (*hesed*), judgment (*mishpat*), and righteousness (*tzedakah*), and thus to imitate the ways of God...

Conclusion

To conclude, we have seen that Maimonides' last Law on Slaves is anything but a simple plea for virtuous behavior towards one's slave. Such is merely the superficial veneer that covers the deepest of legal and ethical expositions. On the legal plane, the Traditional reader finds an astounding proof for the normative demand to inculcate virtuosity in oneself. Bringing biblical support for all his claims, Maimonides demonstrates that, regardless of one's "stage on life's way," justice and piety are demanded – by law – in both action and disposition. On the ethical plane, the Philosophical reader finds a program to achieve nothing less than human perfection. It is a program, as illustrated by the text's chiasmic structure, that is both hierarchical and iterative – it asks not for perfection in a single bound toward a single moral attitude, but for a lifelong quest for perfection in justice and piety.

Ch. 2:

Article #2 - The Virtuous Servant Owner

A Paradigm Whose Time has Come (Again)

Introduction

“Man is by nature a social animal” (*Politics*, 1253a). So noted Aristotle almost 3000 years ago. Interestingly, while Aristotle did actually conceptualize automatons that might replace the slave labor of his day (*ibid.*, 1253b), he did not envision that humans might interact socially with these automatons. This is because, in addition to living at a time when human slaves were considered animated tools, he never imagined the sophisticated automatons of the twenty-first century – i.e., social robots, which today come in a vast and growing array of configurations (Reeves, Byron, et al. 2020), many designed to be social companions.¹ Indeed, the social robots of today are not merely functional automatons, they are emotionally engaging humanoids. And even those not designed to be so, nevertheless manage to trigger our empathy, drawing us to relate to them *as if* they too were, by nature, a “social animal.”

It is this “as if” (Gerdes 2016: 276) condition that brings us to one of the most consternating conundrums in the field of robo-ethics today, what Mark Coeckelbergh calls, “the gap problem” (Coeckelbergh 2013, 2020c). When we interact with a Social Robot (SR), a “gap” exists between what our reason tells us about the SR (i.e., it is a machine) versus what our experience tells us about the SR (i.e., it is more than a machine). It is this gap that gives rise to the ethical question that is the subject of this chapter: How are we to relate *morally* to social robots – like a machine or more than a machine?

Before attempting to address this question, it is important to define specifically the type of SR that is the focus of this investigation. Social robots are currently powered by artificial intelligence, which enables them to “learn” from their experiences, modify their behavior accordingly, and give the appearance of autonomy – the appearance of beliefs, desires and intentions. These features are the hallmarks of consciousness and what make us, in large part, who we are. But today, the artificial intelligence powering our social

¹ For the sake of completeness, it should be made clear that Aristotle did envision *intelligent* artificial servants, nevertheless, he could not imagine interacting with them other than as natural slaves, since slaves were a natural part of his politics. His desire for automatons was motivated not by ethical qualms but by expediency (*Politics* 1253b). For more on this see LaGrandeur (2013: 9-11, 106-108).

robots is entirely artificial – entirely based on mathematics (see, e.g.: Domingos 2018; Boucher 2019; Brand 2020: 207; Coeckelbergh 2020a: 83-94)² – the robot only behaves *as if* it has consciousness.

There are hopes, even designs, to make social robots with true human-like second-order consciousness – i.e., to make a sentient, self-aware being that has the capability to think about its own thoughts. However, while this may be the ultimate goal of the AI project, what Ray Kurzweil calls “the singularity,” its achievement remains a long way off (see, e.g.: Torrance 2008: 500; Coeckelbergh 2010c: 210; Wallach and Allen 2010: 8; Tallis 2012: 194; Veruggio and Abney 2012: 349; Prescott 2017: 5; Sparrow 2017: 467; Bertolini and Arian 2020: 45; Birhane and van Dijk 2020: 210; Hauskeller 2020: 2. And even the less ambitious HLMI [High-Level Machine Intelligence] is a long way off, see, e.g.: Grace et al. 2018; Boucher 2019: 10; Shalev-Shwartz et al. 2020: 2. Some, however are optimistic: Dyson 2012, Moravec 1988, Kurzweil 1999 cited in Sparrow and Sparrow 2006; Long and Kelley 2010, O’Regan 2012, and Gorbenko et al. 2012 cited in Neely 2013.) Accordingly, this paper does not seek to discuss social robots with human-like consciousness, nor even with simple animal sentience,³ but rather social robots that are driven by current day artificial intelligence – i.e., robots that are essentially autonomous mobile computers with humanlike physical characteristics,⁴ what I call: mindless humanoids.

² For the sake of completeness, today’s AI is known as Narrow or Weak AI, which uses algorithms to analyze data, mathematically, and make decisions accordingly. This is as opposed to General or Strong AI (sometimes referred to as GAI or AGI), which seeks to make machines intentional with consciousness. How will this be done is of great debate. There are “computationalists” (e.g., Ray Kurzweil, Hans Moravec) who believe that when every brain function is implemented at the level of human brain processing power, consciousness will “emerge.” Others (e.g., Pentti Haikonen) explain that it is not just the computational power that is needed but the way the computations are done (e.g., via associative neural networks, etc.). Still others (e.g., Roger Penrose, Colin Hales) believe that computation in itself, in any manner, is not enough but rather the physics of the brain must be replicated for consciousness to emerge. [For more detail on consciousness in general and machine consciousness in particular, see Ch. 3 “To Make a Mind”].

³ While there is much to be said in regard to our moral attitude toward sentient robots, such a discussion remains outside the scope of this article.

⁴ I make the proviso of “humanlike” to exclude autonomous mobile computers like autonomous vehicles or assembly-line machinery for which I have yet to read of individuals becoming emotionally engaged.

The Dilemma

So, again, the question is: How are we to relate *morally* to social robots?

In general, when we encounter a new entity – be it mineral, vegetable, animal, or human – we seek to categorize it according to its various ontological properties (see, e.g.: Coeckelbergh 2013: 63; Johnson and Verdicchio 2018: 292). We do this so that we know how to interact with it, and more profoundly, how to interact with it morally. For example, if it is a rock, we know we can kick it into an open field without qualms about harming the rock; if it is a neighborhood cat, we know that we shouldn't kick it or otherwise indiscriminately cause it pain; if it is our human co-worker, we realize that greater moral consideration is due him than a cat. In short, we ask what the entity “is” in order to determine how we “ought” to treat it.⁵ This approach is variously known as the ontological approach, the properties approach (Tavani 2018), the mind-morality approach (Gerdes 2016), the organic approach (Torrance 2008, Tollon 2020), the realist approach (Torrance 2013) or simply, the standard approach (Coeckelbergh 2013).

The ontological approach, however, encounters difficulties with social robots as they fall into a strange middle ground between man and machine, presenting the previously mentioned gap problem, alternatively referred to as a “category boundary problem” (Coeckelbergh 2014: 63). On the one hand, the SR is a mindless automaton, programmed⁶ to carry out various social tasks – i.e., a machine. On the other hand, the SR, designed with human-like physical characteristics and programmed to carry out its tasks with human-like behavior, appears to us as, well, human-like. Furthermore, even if we are aware of the fact that it is not human, that it does not have a mind, a consciousness, we are nevertheless deceived (see, e.g.: Turkle 2011a: 63,90; Grodzinsky et al. 2014: 92, 98; Richardson 2015: 124; Gunkel 2018: 115; Leong and Selinger 2019: 307).

⁵ For a concise discussion of the is-ought debate see Gunkel (2018: 3-4). See also Coeckelbergh (2013: 63), Schwitzgebel and Garza (2015: 99).

⁶ The term applies whether the SR is driven by conventional programming (i.e., rule based hard-coded algorithms) or machine learning (see, e.g., Domingos 2018; Boucher 2019).

The deception is of course self-deception, a result of our own human “programming,” if you will. We are “wired” to respond to animacy, to self-propelled entities that “make eye contact, track our motion, and gesture in a show of friendship” (Turkle 2011a: 8; see also, e.g.: Arico et al. 2011, Gray and Schein 2012: 408, Scheutz 2014b: 213, Darling, et al. 2015: 770; Schwitzgebel and Garza 2015: 112; Darling 2016: 217; Ghiglino and Wykowska 2020: 53). These behaviors push, what Sherry Turkle calls, “our Darwinian buttons,” inducing us to ascribe human attributes to such robots until we “imagine that the robot is an ‘other,’ that there is, colloquially speaking, ‘somebody home’” (Turkle 2011a: 8; see also, e.g.: Foerst 2009, Arico et al. 2011, Turkle 2011b: 63, Scheutz 2014b: 215, Richardson 2015: 72, Bertolini 2018: 649, Fossa 2018: 124). Sven Nyholm calls this “mind reading” – we read into the behaviors of others their apparent mental state, their mind (Nyholm 2020; see also, e.g.: Richardson 2015: 74; Darling 2016: 216; de Graaf and Malle 2019; Ghiglino and Wykowska 2020: 51). Others (e.g.: Duffy 2003: 180, Huebner 2009, Veruggio and Abney 2012: 355, Ghiglino and Wykowska 2020: 67, Tollon 2020: 7) say we adopt, what Daniel Dennett terms, the “intentional stance,” whereby we treat an entity “*as if* it were a rational agent who governed its ‘choice’ of ‘action’ by a ‘consideration’ of its ‘beliefs’ and ‘desires’” (Dennett 1996).

This phenomenon of seeing social robots as humanlike is known as anthropomorphism, but it doesn’t end with simply ascribing human beliefs, desires and intentions to the robot – we take it to the next step and become engaged, emotionally, with the social robot (see, e.g.: Coeckelbergh 2009; Choi 2013; Grodzinsky et al. 2014: 92; Darling 2016: 214; Richards and Smart 2016: 18; Darling 2017; Johnson and Verdicchio 2018; Tavani 2018: 3; Gunkel and Wales 2021; see also sources cited in previous paragraph). And this engagement isn’t just some kind of fictional role playing, but rather, we feel real empathy toward the social robot (see, e.g.: Redstone 2014, Darling, et al. 2015, Wales 2020). Indeed, Tony Prescott notes that “we do not need to believe (or be deceived) that the psychological states, intentional, or phenomenological, that we read into an artefact, such as a robot, are akin to our own in order to experience an authentic and meaningful emotional response” (2017: 144).

Now, while this emotional anthropomorphizing is going on, another socio-psychological element comes into play: dehumanization. Massimiliano Cappuccio et al., describe this troubling phenomenon:

“... the fundamental ethical problem at the core of social robotics is that, while robots are designed to be like humans, they are also developed to be owned by humans and obey them. The disturbing consequence is that, while social robots become progressively more adaptive and autonomous, they will be perceived more and more as slave-like. In fact, owning and using an intelligent and autonomous agent instrumentally (i.e., as an agent capable to act on the basis of its own decisions to fulfill its own goals) is precisely the definition of slavery” (Cappuccio et al. 2019: 25).

Cappuccio et al. call this the Anthropomorphism Dehumanization Paradox (ADP). Jordan Wales (2020) calls it “the dilemma of empathy and ownership,” explaining that if we allow ourselves to engage emotionally with robots, we will nevertheless use them for what we acquired them to do and, accordingly, end up treating them as slaves (similarly, Walker 2006b). This might not seem so terrible since the machine “feels” no indignity or ignominy, no disgrace or denigration – indeed, the machine “feels” nothing.⁷ The problem, however, is not for the machine but for man, as Kant famously noted:

So if a man has his dog shot, because it can no longer earn a living for him, he is by no means in breach of any duty to the dog, since the latter is incapable of judgement,⁸ but he thereby damages the kindly and humane qualities in himself, which he ought to exercise in virtue of his duties to mankind. Lest he extinguish such qualities, he must already practise a similar kindness towards animals; for a person who already displays such cruelty to animals is also no less hardened

⁷ The debate on whether it is possible to give machines emotions and feelings is outside the scope of this paper. Suffice it to say that truly sentient machines are not, as mentioned above, in the offing.

⁸ Kant famously held that the line dividing those deserving of moral status versus those undeserving of such was “judgement” (or reason), a position which became anathema following Bentham’s revision of the dividing line to “sentience,” or more precisely, the ability to suffer (Bentham [1789] 2019). So, while Kant’s example of dog may grate on today’s sensibilities, it provides a fitting paradigm to address the mindless humanoid which has neither judgement nor sentience.

towards men. We can already know the human heart, even in regard to animals (Kant 1996, 212).⁹

Similarly, it is feared that our instrumental treatment of human-like robots – treating them as slaves – will then influence our treatment of humans (e.g.: Levy 2009, Anderson 2011b: 294, Darling 2016: 227-8, Cappuccio et al. 2019: 14, Chomanski 2019: 1008, Gunkel and Wales 2021: 4, 9, Coeckelbergh 2021: 7, in opposition see, e.g.: Johnson and Verdicchio 2018, Bryson 2020b: 22). We will likely not treat people as slaves, but we will certainly be in danger of treating people as objects rather than subjects. Our relationships with SRs, to put it Buberian terms, could be seen as habituating an I-It relationship as opposed to cultivating an I-Thou relationship (Buber 1970). The SR would thus invert Buber’s call to relate to the other as subject not object, hardening us, to echo Kant, to view the other as object not subject (Hawley 2019: 12). And this, ultimately, reflects upon the individual as vicious as opposed to virtuous.¹⁰ For Buber, the individual – the “I” – is not merely influenced by his relationship with the other, he is *defined* by it. “There is no I as such but only the I of the basic word I-Thou and the I of the basic word I-It. When a man says I, he means one or the other” (Buber 1970: 54). Consequently, some, like Michael Burdett (2020), have suggested that it would be appropriate for us to relate to a robot as a “Thou.” Others, like Elizabeth Green (2018) argue that a robot can never be a Thou, while still others, like Sherry Turkle (2011a: 85), explain that the “Thou” relationship simply emerges.

⁹ Worthy of note is that Kant (1724-1804), here, was preceded by Nachmanides (1194-1270) who explains that the biblical command to send the mother bird away before taking her eggs was promulgated in order “that we should not have a cruel heart and lack compassion ... and is to prevent us from acting cruelly” (Nachmanides 1976: Deut. 22.6). Thus, while some argue that Kant’s words point only to a concern for causal action and not character disposition (see fn. 10 herein), Nachmanides explicitly voices concern for both aspects, reiterating, “the reason for the prohibition is to teach us the trait of compassion and that we should not be cruel...” (ibid.).

¹⁰ Worthy of note is the disagreement over whether Kant is concerned only with the externally causal effect – e.g., kicking a dog will bring one to kick a human (see, e.g.: Coeckelbergh 2020b, Coeckelbergh 2020c, Sparrow 2020) – or does Kant’s demand for virtuous behavior because it reflects on the character of the individual (see, e.g.: Gerdes 2016, Denis 2000).

Resolutions

This brings us into the thick of possible “resolutions” to the dilemma. I keep the term “resolution” in quotes because this dilemma, like all worthy of the name, only reach resolution with the sacrifice of ideals. This point will be made all too clear in the following review of proposed resolutions.

Returning to Cappuccio et al. (2019: 26), who describe the dilemma as a paradox, we encounter two practical approaches to dissolve the paradox: either reduce – by design – the elements that promote anthropomorphizing, thus keeping the machine very much a machine,¹¹ or conversely, increase those elements that engender empathy to encourage human to human-like interaction.¹² Both approaches, they note, are not really solutions. Reducing the anthropomorphic elements of SRs undermines their very purpose as companions that are to “establish trust and cooperation, [be it] with a child, a patient with disabilities, or an elderly person” (Cappuccio et al. 2019: 26). On the other hand, increasing such elements that engender human-like empathic relationships, opens a Pandora’s box of ethical issues based on the misperception of the true nature of the machines, including but not limited to: developing intimate relationships with robots (Turkle 2011a: 295, Richardson 2015: 12, Gerdes 2016: 277, Bertolini 2018: 653), shunning human relationships as “messy” (Turkle 2011a: 7; similarly, Whitby 2008: 331, Bryson 2010: 7, Toivakainen 2015: 10), prioritizing humanoids over humans, thus mispending or misallocating resources (Torrance 2008: 498, Bryson 2010: 3, Neely 2013, Schwitzgebel and Garza 2015: 114), sacrificing human life (Torrance 2008: 508, Smids 2020: 2850), seeing oneself as a machine and thus shirking moral responsibility (Metzler 2007: 20), and generally maintaining a warped view of reality (Sparrow and Sparrow 2006: 155, Gerdes 2016: 276).

¹¹ Many make this argument, e.g.: Bryson 2010: 65, John McCarthy and Marvin Minsky in Metzler 2007: 15, Miller 2010, Grodzinsky et al. 2014, Schwitzgebel and Garza 2015: 113, Richards and Smart 2016: 21, Leong and Selinger 2019. The position is even offered as a regulatory principle (Boden et al. 2010: #4), though Wales (Gunkel and Wales 2021: 11) argues it will simply not be followed.

¹² Many make this argument, e.g.: Breazeal 2002, Duffy 2003, Walker 2006b, Darling 2017, Burdett 2020.

The two solutions that Cappuccio et al. float can be seen as an attempt to sway a resolution to the gap problem. That is, either we emphasize what our reason tells us about the SR (i.e., it is a machine) or we emphasize what our experience tells us about the SR (i.e., it is more than a machine). Interestingly, this dichotomy reflects the split of the philosophical community in to two distinct camps.¹³ On the one side, there is the “instrumental” camp, populated by those who believe that machines are machines and, regardless of their appearance and behavior, we should relate to robots like we would to a toaster or a vacuum cleaner (see, e.g.: Gunkel 2018: Ch. 2 “!S1 !S2”). On the other side, there is the “appearances” camp, populated by those who maintain that it is precisely through appearance and behavior that we engage with others and must similarly relate to robots (see, e.g.: Gunkel 2018: Ch. 5 “!S1 S2”).

The instrumental camp could also be referred to as the “insides count” camp, in that they take the position referred to earlier as the “ontological approach.” They derive the moral status of the entity based on its ontology, on “what’s going on inside.” Accordingly, sentience or first-order consciousness is needed for moral patiency and second-order consciousness is needed for moral agency (see, e.g.: Anderson 2013, Smids 2020). In opposition, the “appearances” camp argues that we have no method to reveal the insides of an entity for we have no “privileged access” to determine if a being is conscious. As a result, we must content ourselves with externals, with the behavior of the entity and its interaction with us. Some here argue that this approach is not simply an accommodation due to epistemological deficiencies but is the philosophically preferred approach based on our lived experience of SRs (see, e.g.: Gunkel 2018, Coeckelbergh 2010c). Accordingly, we must grant SRs, if not full moral agency then, moral patiency or moral consideration. This approach has been called the relational approach (Coeckelbergh 2010c, Richardson 2015) the phenomenological approach (Coeckelbergh 2010a), the hermeneutic approach (Coeckelbergh 2021), and includes the ethical behaviorist approach (Neely 2013; Danaher 2019b).

¹³ Cappuccio et al. (2019: 10) note the two camps explicitly; so too, Torrance (2013: 10). Gunkel (2017; 2018) adds two additional camps in order to account for sentient machines (which, as mentioned, are beyond the scope herein). It should be noted that Gunkel defines yet another camp for himself.

The Middle Camp

Now, while I have described the dilemma as being approached from two sides, two camps, there is in fact a middle ground, a middle camp, occupied by thinkers that believe insides count but also believe that there are reasons to relate morally to the mindless humanoid as more than a mere machine. That is, though the SR is neither a moral agent nor a moral patient, there are nevertheless ethical demands incumbent upon humans in their interactions with it. Steve Torrance, who I place in this middle camp, describes the moral relationship with a robot as “quasi-moral” (2007: 504, 516). I understand this to mean that the moral demands engendered in the HRR (Human Robot Relationship) do not stem from the inherent moral *status* of the robot but from the relationship, from the moral *implications* of the relationship. This, it should be noted, is in contradistinction to the “relational approach” which sees the mindless humanoid as a “quasi-other.” To be clear, in the “quasi-other” approach it is otherness, alterity, that is imposed on the robot itself which consequently engenders a very real moral demand – e.g., the demand to treat the other like yourself;¹⁴ whereas in the “quasi-moral” approach, it is morality (e.g., a norm) that is imposed on an otherwise amoral situation.

This quasi-moral approach taken by the middle camp finds its ground in Kant’s indirect duties to the animal kingdom. Kant believed that animals have no moral status and accordingly, he writes, “we have no immediate [i.e., direct] duties to animals; our duties towards them are indirect duties to humanity” (Kant 1996: 212). Anne Gerdes (2016) explains Kant as teaching that we have not duties *to* animals but rather we have duties *with regard to* animals; similarly, reasons Gerdes (as does Bryson 2010), we have not duties *to* robots but rather we have duties *with regard to* robots. She brings Kant’s writing on this point in his *Metaphysics of Morals*:

¹⁴ This approach is found in numerous authors, as, for example, the following list shows. Coeckelbergh (2010a): a robot is “quasi-alterity” to be treated as it appears to us. Michael Burdett (2020): a robot is “quasi-person” which demands “Thou” relations. Don Ihde (1990: 100): a robot is “quasi-other” but remains lower than human or animal; see also Bergen and Verbeek 2020. Peter Asaro (2006): a robot is “quasi-moral agent” giving it some level of responsibility. Philip Brey (2014) argues that the term “quasi-moral agent” denotes involvement in moral acts but without true moral responsibility. Gunkel (2018: Ch. 6) argues for Levinasian alterity relations – i.e., a robot is a full other, not simply a quasi-other.

... a propensity to wanton destruction of what is beautiful in inanimate nature ... is opposed to a human being's duty to himself; for it weakens and uproots that feeling in him, which, though not of itself moral, is still a disposition of sensibility that greatly promotes morality or at least prepares the way for it...

With regard to the animate but non-rational part of creation, violent and cruel treatment of animals is far more intimately opposed to a human being's duty to himself, and he has a duty to refrain from this; for it dulls this shared feelings of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other men. ...

Even gratitude for the long service of a horse or dog belongs indirectly to a human being's duty with regard to these animals; considered as a direct duty, however, it is always only a duty of the human being to himself (6:443).

This passage, as well as the one quoted immediately prior, can be seen as advancing a virtue ethics approach toward non-human entities – as, indeed, Gerdes writes. That is, in our actions toward the inanimate, though no deontological demands bind our behavior, we are nevertheless to refrain from wanton destruction as part our efforts at developing a disposition that promotes moral behavior – i.e., in order to develop our virtuous character (so too, Toivakainen 2015: 278). With regards to animals, our behavior has an even greater impact on our dispositions. Lara Denis explains that, for Kant, “Any way of treating an animal that could impair our ability to feel love and sympathy for others constitutes a risk to a morally valuable aspect of our rational nature. Kant thinks that cruel or even unloving treatment of animals threatens to impair us in this way” (Denis 2000: 409). Denis explains that the reason our interactions with animals so affect our dispositions is because we share our animal nature with them and because they engage us emotionally.

Given this, I would argue that, while a SR could be considered an inanimate object, its human-like interaction with us, to the point of our attributing mental states to it, places the SR more closely in the animate category. And though we don't share our biological animal nature with the robot, we do share behaviors engendered by our animal nature (see, e.g., Turkle 2011a: Ch. 7). Furthermore, while our emotional engagement with the

robot lacks the authentic sentient elements of pain and pleasure characteristic of animal interaction, behaviorally we are just as engaged (see prior sources on emotional engagement as well as, e.g.: *ibid.*; Cappuccio et. al 2019: 15-16). Accordingly, without arguing for the “appearances” approach, I am calling for a virtue approach – i.e., an approach which acknowledges and accounts for how the interaction with a mindless humanoid affects the virtue of the human interlocuter.

The virtue approach to robots is not new and has, in fact, been promoted by numerous thinkers such as: Anne Gerdes (2016), Robert Sparrow (2017, 2020), Shannon Vallor (2018), Cappuccio et. al (2019), and even Mark Coeckelbergh (2020b, 2020c, though he argues against in 2010c). However, while virtue ethics clearly eliminates the “dehumanizing” part of the “anthropomorphizing while dehumanizing paradox,” it would appear to utterly capitulate to the anthropomorphizing part. That is, by relating to the SR in a virtuous manner we avoid the evils inherent in dehumanizing it but remain susceptible to the previously mentioned Pandora’s box of negative consequences associated with anthropomorphizing it. Consequently, Cappuccio et al. (2019: 26) acknowledge that they are thus at a loss to resolve the paradox and content themselves to apply virtue ethics to avoid dehumanizing.

One scholar who does attempt a resolution is Jordan Wales (2020), who employs the thought of Augustine to address the paradox. Augustine, in his *De doctrina Christiana* (1:33:37), teaches that one should ever seek to refer his joy in an other toward God, toward the creator of that individual.¹⁵ Wales applies this notion to our interactions with SRs, such that, upon feeling natural empathy toward a SR, “we *redirect* that empathy, ‘refer’ it, as Augustine would say, to all the unknown concrete persons whose interactions have unwittingly sculpted the persuasive personality of this instrument” (Wales 2020: 7). Wales thus solves the anthropomorphism problem, or more precisely, the empathy problem inherent in anthropomorphizing.

To be clear, in anthropomorphizing mindless humanoids, we are in danger of becoming emotionally engaged with entities that do not warrant such engagement and which can

¹⁵ This is a well-known religious technique wherein one is to channel one’s emotions toward God in an effort to connect to the source of all emotion and life itself (see, e.g., Horowitz 1873: Gen. 46:29).

thus lead to many social ills (as noted above). By redirecting the empathy in our emotional engagement with the SR toward the real flesh and blood people who served to create it, Wales argues that we avoid attributing humanity to the robot, allowing our emotions to find their proper terminus in true humanity.¹⁶ As a result, we can interact with the SR in a virtuous way, allowing our natural empathy and anthropomorphizing to occur and yet maintain the realization that the robot is not human, does not have the moral status of a human and does not enter the moral circle of humanity.

Now, while this idea of “referring” or “redirecting” one’s intentions is an accepted notion as a religious ideal, allowing for an adherent to utilize an emotional encounter as a means to develop a connection with his creator, it does not, in my humble opinion, work in other contexts. Indeed, even in the religious context, such channeling of thoughts and emotions is not simple and accomplished only by the truly devout (see, e.g., Maimonides 1956: III:51, Horowitz 1873: Gen. 46:29). To expect people to “reference” an other through a SR while in the midst of their everyday mundane lives is utterly impractical. To help us envision the idea, Wales analogizes the connection of ‘robot-creator(s)-to-robot’ to that of ‘baker-to-cookie’ – i.e., we could “reference” the baker when we eat his cookie. It is certainly nice to contemplate such a notion, but again, utterly impractical. Furthermore, I think a better analogy of ‘robot-creator(s)-to-robot’, instead of ‘baker-to-cookie’, would be ‘parent(s)-to-child’. This analogy, I believe, makes clear just how terribly difficult it is to redirect or refer one’s thoughts to an other – for, can one really focus on the parent(s) of a child while interacting with the child alone – whether upon first thought or, as Wales suggests, upon second thought.¹⁷ Again, as a religious ideal,

¹⁶ Burdett (2020: 355), basing himself on Pattison, makes a similar point. All of these thinkers have been preceded, in a sense, by Buber (1970: 175) who, upon confronting a Doric column in a Syracuse church, writes that he related to the “spiritual form there that had passed through the mind and hand of man and become incarnate.” A distinction worthy of note is as follows. Buber is seeking to establish the I-Thou relationship with the inanimate by “referring” to the humanity behind it – he is trying to generate a close, “Thou”, relationship; while Wales is trying to “refer” the already close “Thou” relationship to its underlying humanity to avoid seeing the robot as more than it is and falling into the misplaced-empathy trap.

¹⁷ Wales attempts to make the creators of the robot more resident in the robot by explaining that it is not the engineers who built the robot that are represented in the robot, but the very people whose behaviors made up the data that was used to train the neural network that grounds the robot’s behaviors. However, the same could be said of the child whose behaviors are made by the DNA and parental education that

reflecting upon the creator in an encounter with an other may be a worthy challenge, but to import the technique to robot encounters will simply not work.¹⁸

An opposing attempt to resolve our dilemma is brought by Raffaele Rodogno (2016). That is, if Wales attempted to solve the dilemma by framing the HRR as very real, the solution offered by Rodogno is to cast it as utterly fictional:

... we could hypothesize that, when engaging affectively with robot pets, individuals adopt a cognitive mode akin to that which is normally adopted in our engagement with fiction. Being emotionally engaged by robot pets would be akin to being emotionally engaged by a good novel or movie. Just as my sadness for Anna Karenina involves my *imagining, accepting, mentally representing* or *entertaining the thought, without believing*, that certain unfortunate events have occurred to her, my joy at the robot pet involves my *imagining, accepting, mentally representing* or *entertaining the thought, without believing*, that it is happy to see me (2016: 11).

This solution is untenable for a number of reasons. First of all, the relationships we build with fictional characters on the page or screen are both temporary and passive – our interaction with them is limited in time and confined in “space” to our own mind. Robot interactions, in contradistinction, are ongoing active relationships with entities deceptively alive in the three-dimensional space in which we live. As such, they are very different not only from fictional storybook characters but even from real dolls that are not animated to the point that we ascribe to them beliefs, desires and intentions (see, e.g., Turkle 2011a: 39). Secondly, as noted above (sec. The Dilemma), we take these relationships quite seriously, treating them as if they were not merely fictional – a fact that has dangerous consequences, as Gerdes notes: “the relational *as if* approach is

make up the neural network that grounds the child’s behavior. In any case, the notion of referencing is not practical.

¹⁸ I make this claim as a religious man who appreciates the religious ideal. I am not alone in this claim, for when I made it directly to Wales at the RP2020 conference (as he notes in his fn. 22), many other voices joined me in dissent and none in his defense. [Addendum: after this essay of mine was published, Wales contacted me to agree with my point. I leave the argumentation here because the approach is a valid one that needed to be raised, reasoned, and ultimately, rejected.]

challenged by the fact that, over time, our human-human relations may be obscured by human-robot interactions” (2016: 276).

In psychological terms, the HRR engenders a state of cognitive dissonance (Festinger 1957) wherein one knows he is interacting with a very real entity, a SR, while at the same time knowing very well that the interaction is not “real,” not authentic. Both Wales and Rodogno attempt to diffuse the dissonance, but from opposite ends. Wales attempts to achieve cognitive harmony by relating the relationship to something real, authentic. That is, since the physical interaction is real, he tries to make the metaphysical relationship real as well. It doesn’t work because the referred metaphysical relationship can’t be imagined. Attacking the problem from the other end, Rodogno attempts to achieve cognitive harmony by framing the relationship as completely fictional, inauthentic. That is, since the metaphysical relationship is fictional, he tries to make the physical relationship fictional as well. It doesn’t work because the physical relationship can’t be imagined away.

VSO

And so we return to our question: How are we to relate *morally* to social robots?

Having reviewed the various attempts to construct a response, it is clear that the question, in both physical and metaphysical terms, is strained in the tension between the need to preserve virtue, on the one hand, and the need to preserve authenticity, on the other – what might be termed the Virtue-Authenticity Dialectic (VAD). The ideal response, then, must strive to allow us to maintain our virtuous character, such that we not act in dehumanizing ways toward SRs, but at the same time allow us to maintain our appreciation for authenticity, such that we not accustom ourselves to “as if” relationships *as if* they were real.

As for the “virtue” part of the response, Aristotle’s virtue ethics, as echoed in Kant’s appeal to indirect duties toward animals, soundly satisfies this need as evidenced by its broad support among thinkers in the field. As for the “authenticity” part of the response, thinkers in the field, as noted, run into trouble.

To address the “authenticity” issue, it is instructive to revisit Aristotle’s approach to automata as found in his *Politics*:

Now of instruments some are inanimate and others animate—the pilot’s rudder, for example, is an inanimate instrument, but his lookout an animate one; for the subordinate is a kind of instrument whatever the art. ... if each of the instruments were able to perform its function on command or by anticipation, as they assert those of Daedalus did, or the tripods of Hephaestus (which the poet says “of their own accord came to the gods’ gathering”), so that shuttles would weave themselves and picks play the lyre, master craftsmen would no longer have a need for subordinates, or masters for slaves (1253b).

Aristotle here envisions that automata will replace slaves as instruments of their masters (similarly, *Nicomachean Ethics* 1161b). Now, while Aristotle may have been the first to articulate this instrumental approach, the history of automata, real or fictional, leaves little doubt that automata were forever imagined to be slaves (see, e.g., LaGrandeur 2013). And with the advent of AI they continue to be so imagined. Hans Moravec claimed, ‘By design, machines are our obedient and able slaves’ (Moravec 1988: 100); Nick Bostrom argued that “investors would find it most profitable to create workers who would be ‘voluntary slaves’” (Bostrom 2014: 167); but no one popularized the notion more than Joanna Bryson (2010) who entitled her article on the issue, “Robots Should Be Slaves.” Her claim received no small amount of pushback given the cultural scars left on society by the brutal history of human slavery (Bryson 2020a).

And that brings us to the heart of the matter, for while it is clear that the goal of automation is to relieve humans of their burdens,¹⁹ slavery is an institution that runs counter to modern values. Slavery is an institution that, despite Aristotle’s justifications (*Politics*: Book 1, Chs. 4-5), has been shown to undermine the very virtue ethics that Aristotle sought to foster. Powerful evidence of this can be seen in the testimony of Frederick Douglass (1845) who wrote of his experience as a slave under a woman he refers to here as “my mistress” – i.e., “female master” slaveholder:

¹⁹ There is a vast literature on how automation, and specifically AI, will replace human labor, see, e.g.: LaGrandeur 2013: 161; Marr 2017; Harari 2019: Ch. 2; Coeckelbergh 2020a: 136.

My mistress was, as I have said, a kind and tender-hearted woman; and in the simplicity of her soul she commenced, when I first went to live with her, to treat me as she supposed one human being ought to treat another. In entering upon the duties of a slaveholder, that [now] I sustained to her the relation of a mere chattel, and that for her to treat me as a human being was not only wrong, but dangerously so. Slavery proved as injurious to her as it did to me. When I went there, she was a pious, warm, and tender-hearted woman. There was no sorrow or suffering for which she had not a tear. She had bread for the hungry, clothes for the naked, and comfort for every mourner that came within her reach. Slavery soon proved its ability to divest her of these heavenly qualities. Under its influence, the tender heart became stone, and the lamblike disposition gave way to one of tiger-like fierceness (1845: 32).²⁰

Accordingly, as described previously, many have expressed concern that modern robots designed to serve humans will be treated as slaves and engender a moral calamity for their owners.

But is this outcome not unavoidable? Kant believed it is. He wrote that while one must not hold a slave because, in so doing, one violates the freedom that is at the essence of the individual as a person, nevertheless, one could come to an agreement into which the servant enters of his own freewill and can exit of his own freewill. In such a case, Kant, in his *Metaphysics of Morals*, writes:

Servants are included in what belongs to the head of a household, and, as far as the form (the way of his being in possession) is concerned, *they are his by a right that is like a right to a thing*; ... But as far as the matter is concerned, that is, what

²⁰ Similarly, this slave girl testimony: “I can testify, from my own experience and observation, that slavery is a curse to the whites as well as to the blacks. It makes the white fathers cruel and sensual; the sons violent and licentious; it contaminates the daughters, and makes the wives wretched” (Jacobs 2020); as well as that of French philosopher Alexis de Tocqueville, “Servitude, which debases the slave, impoverishes the master” (de Tocqueville [1835] 2013).

use he can make of these members of his household, *he can never behave as if he owned them* (6:284. *Emphasis added*).²¹

Kant here claims that you can maintain a relationship in which, on the one hand, you are in the position of a servant owner; yet, on the other hand, your behavior toward your servant never expresses this position. I believe that we can reconcile Kant's claim with the seemingly damning evidence brought by Douglass to the contrary, as follows.

Douglass wrote: "In entering upon the duties of a slaveholder, she did not seem to perceive that [now] I sustained to her the relation of a mere chattel, and that for her to treat me as a human being was not only wrong, but dangerously so. Slavery proved as injurious to her as it did to me." That is, only upon fully accepting the slaveholder role – in which one relates to the slave as chattel and in which treating a slave as a human being is "not only wrong, but dangerously so" – does slaveholding become injurious to the slaveholder. The injury to the slaveholder, then, is when the slaveholder assumes that one must treat the slave as non-human. That is, it was not the owning of a slave per se, but the social concepts of the time that dictated *how* one needed to treat a slave – i.e., by force of "tiger-like" subjugation to ensure obedience.

A machine programmed for obedience, however, would never occasion its owner to impose her will. Nevertheless, there remains a further moral concern in owning a slave, humanoid or human:

²¹ An important aside: Kant's contract binds the servant but nevertheless allows him to quit. The servant is then like a slave in the sense that he is the property of, and at the command of, the owner, all the while retaining some human dignity in his ability to exercise his will to both enter and exit the contract freely. In reality, however, it would seem that someone in a position to accept such a contract would be in such dire straits that he will likely never have the means to exit the contract. As such, he is only a "free" servant in name but a slave in practice. Furthermore, it is not clear how the owner can unilaterally, according to Kant, "fetch servants back" (ibid.), if the servants are allowed to terminate the contract at will. The only way this makes sense is by saying that the servant failed to give notice when he left. But why would he not give notice and leave legally if he could do so at will? Maybe the giving notice of leave is actually very limited. It seems that Kant's ownership is closer to slavery than would at first appear.

There is some harm to one's own higher moral values and moral character if one establishes oneself as master... The problem of using and treating machines as slaves is that one perpetuates a value that sustains the inappropriate agent character, seeing the world and its denizens as one's slaves. You simply should not treat the world as a place in which your will is absolute. You thereby only strengthen that absolutist, disregarding will (Miller 2017: 5; similarly Coeckelbergh 2021: 7).

This harkens back to Kant's dog and the concern against habituating vicious character through vicious behavior. In employing machine-slaves, as stated at the outset: we will likely not treat people as slaves, but we will certainly be in danger of treating people as objects rather than subjects. Accordingly, Kant is not concerned for the virtue (or loss thereof) of one who maintains a servant, as long as she behaves toward her servant as a human being and not as "a thing." Sven Nyholm writes that "Kant himself thought that having a human servant does not need to offend against his formula of humanity [i.e., that one must treat others as ends and not merely as means] – so long as the servants are treated well and with dignity" (2020: 192).

This idea finds precedence in the legal writings of the Medieval Jewish philosopher Moses Maimonides. He not only preceded Kant in demanding that servants be treated with dignity, he also elaborated such treatment with details that are instructive in both pragmatic and moral dimensions. Here is his original text (*Laws of Slaves* 9:8), interleaved with some clarifications of mine:²²

It is permissible to work a heathen slave relentlessly. [Biblical law often promulgates rules in concert with ancient custom while nevertheless seeking to provide a moral improvement on the accepted state of affairs (see, e.g.: Korn 2002; Rabinovitch 2003; Lamm 2007; on slavery see, e.g., Shmalo 2012). As such, the strict letter of law allows for slavery but with various moral restraints.²³ The law,

²² For a detailed analysis of this text see Ch. 1 "Finding Virtue in a Law on Slaves".

²³ For example, killing a slave entails capital punishment (Ex. 21:20, Rashi ad loc.), a slave is set free if injured (Ex. 21:26-27, Kid. 24a), a slave rests on the Sabbath (Ex. 20:9); a runaway slave is not to be

however, is seen as a starting point, a floor and not a ceiling, to use the words of Rabbi J. D. Soloveitchik. Accordingly, Maimonides starts with the legal “floor” only to show that we should – and must – rise far above it. It is interesting to note that Kant (*Metaphysics* 6:284) used the same format, starting with the letter of the law allowing for ownership only to then argue for virtue.]

Though this is the law, the quality of virtue and the ways of wisdom demand of a human being to be compassionate and pursue justice, and not make heavy his yoke on his slave nor distress him. [Maimonides, here, raises us off the floor of the law, outlining his thesis that calls for virtue and justice. He will now elaborate on these two categories, bringing proof texts to support his claims.]

He should give him to eat and drink of every food and drink. The sages of old had the practice of sharing with the slave every dish they ate. And they would provide food for their animals and slaves before partaking of their own meals. As it is said, “As the eyes of slaves follow their master’s hand, as the eyes of a slave-girl follow the hand of her mistress, [so our eyes are toward the Lord our God, awaiting His favor].” [Here Maimonides provides concrete actions toward maintaining virtuous interactions, grounded in a verse equating master and slave in their shared neediness].

Nor should a master disgrace his servant, neither physically nor verbally; the biblical law gave them to servitude, not to disgrace. And one should not treat him with constant screaming and anger, but rather speak with him calmly and listen to his complaints.²⁴ [Clearly the servant is not to be treated merely as a means but as an end. (I wonder if even Kant would have made such a list of directives to regulate the owner).] *This is explicitly stated with regard to the positive paths of Job for which he was praised: “Have I ever shunned justice for my servants, man or maid, when they quarreled with me... Did not He who made me in my mother’s belly make him? Did not One form us both in the womb?”* (Job 31:13,15).

returned (Deut. 23:16). On the differences between ancient slavery versus that of the Torah, see Beasley 2019.

²⁴ Interestingly, in terms of a model for SRs, this would demand that the SR give negative feedback, and as Kate Darling suggests, “respond to mistreatment in a lifelike way” (Darling 2016: 228; similarly, Cappuccio et al. 2020).

[The claim here is for just relations, supported by the verse that notes the physiological identity of master and slave].

Cruelty and effrontery are not frequent except with the heathen who worship idols. The progeny of our father Abraham, however, the people of Israel upon whom God bestowed the goodness of the law (Bible), commanding them to observe “righteous statutes and judgments” (Deut. 4:8), are compassionate to all. [Maimonides defuses any claims that come to justify slavery merely because such treatment is “accepted practice” among the nations of the world. This is not some parochial diatribe against non-Jews,²⁵ but rather part and parcel of his argument for just relations with one’s servant, here made irrespective of the inherent value of the servant. That is, justice is incumbent upon the master for the sake of his own virtue and character].

Accordingly, regarding the divine attributes, which He has commanded us to imitate, the psalmist says: “His tender mercies are over all His works” (Psalms 145:9). [Here, as part of his thesis that one must move beyond the strict letter of the law in the treatment of one’s servant, Maimonides reminds us of the ethical imperative to strive to imitate the divine virtues, chief among them being that of mercy/compassion. This claim, like the previous one, is incumbent upon the master irrespective of the inherent value of the slave. Worthy of note is that the support verse does not say that God’s “mercies are upon all His *creatures*” but “upon on all His *works*.” Could this not be understood to allow for application to humanoids?]

Whoever is merciful will receive mercy, as it is written: “He will be merciful and compassionate to you and multiply you” (Deut. 13:18). [Maimonides concludes his call for virtue with a religious principle known as “measure for measure,” which states that in the measure, or manner, that you act towards others, so too, in the same measure, will God act towards you. Accordingly, even if one does not appreciate the value of a virtuous character, one will certainly appreciate the selfish need of God’s mercy. In addition, this call to mercy, to virtue, is made independent of the

²⁵ Worthy of note is the great esteem in which Maimonides holds non-Jewish thinkers, frequently quoting, Aristotle and Al Farabi.

worth of the servant. It pleads for virtue saying: though you may not recognize the worth of your servant, nor even the worth of your own character, at least recognize your need for mercy and be merciful.]

This text stands as a powerful call to virtue in general, and to virtuous behavior with one's servant in particular. Maimonides here speaks to any and all, regardless of what "stage on life's way" one might be. Indeed, his arguments for virtuous behavior can be seen as addressing the individual in each of the three Kierkegaardian stages of existence, stages in which one is driven by the corresponding motivations: aesthetic, ethical and religious.²⁶ Starting with the ethical, being that it is the universal – applying to all and in which all struggle (Kierkegaard 1985: 83), Maimonides enjoins virtue based on the human dignity inherent in the servant as a human being. Moving to the higher motivation of the religious, Maimonides calls for the master to exhibit virtue both because he is a God fearing individual who, like Abraham,²⁷ accepts the divine ethical norms of the Bible and furthermore, because he is to emulate the attributes of the creator, mercy being primary among them.²⁸ Maimonides concludes with an appeal to self-interest (i.e., the Kierkegaardian aesthetic), arguing, in essence, that even if one is not moved by these higher motivations, one should act mercifully that he too will be treated mercifully.

Not satisfied in leaving his readers with "mere" motivations, Maimonides takes pains to prescribe practical action. He instructs the master to feed his slave with "every dish" that he himself eats, thus raising the slave to the dignity of the master. He directs the master to feed his slave before he himself sits to eat, thus instilling compassion toward he who is

²⁶ Worthy of note is that Maimonides (1956, 3:33) appears to refer to these categories in articulating the "ultimate causes of the Law": 1) the rejection and reduction of the fulfillment of desires – i.e., aesthetic, 2) the promotion of virtuous interaction between men – i.e., ethical, 3) the sanctification of its followers – i.e., religious. [It should be noted, as I did in Ch. 1, that my use of Kierkegaard is merely as a hermeneutical tool to better understand Maimonides, and does not imply any connection between the two thinkers].

²⁷ Like Kierkegaard, Maimonides references Abraham as the father of faith; yet unlike Kierkegaard, Maimonides, indeed Judaism in general, does not accept the notion of a religious leap of faith as requiring a teleological suspension of the ethical (see Navon 2014).

²⁸ The two demands could be seen to reflect the two levels of the "religious" articulated by Kierkegaard (see Broudy 1941: 306).

not in charge of his own food. He warns the master to “speak calmly and listen to the slave’s complaints,” thus changing the very relationship from one of master-slave to one more akin to employer-employee (and a quite considerate employer at that). Maimonides thus transforms ethical ideal into ethical practice which, ultimately, shapes ethical character (Aristotle [350 BCE] 2004: 23; Ha-Levi [1523] 1978: Precept 16; Vallor 2018: 3.3; Cappuccio et al. 2019; Coeckelbergh 2020b).

Of course no ownership, no matter how virtuous, can be justified today. Slavery is an institution that is anathema in modern moral thought and given circumscribed sanction in the bible, due only to ancient cultural mores. Jewish thought has ever sought to ameliorate the master-slave relationship (see, e.g., Shmalo 2012) to the point that Maimonides demands not simply that one treat his servant as an end, but that one treat him as nothing less than a contemporary! He does so, as mentioned, by providing clear practical behaviors underpinned by clear philosophical reasoning, (albeit) based on biblical verse. Significantly, his arguments are found not in his philosophical writings but in his legal writings, thus giving them normative import and evincing, essentially, a law to go beyond the law.

And this brings us back to SRs. My point here is not to argue for even this most virtuous form of human slavery, but to apply the Maimonidean paradigm – what I call the “Virtuous Servant Owner” (VSO) – to Human Robot Relationships. For, though the virtuous practices demanded by Maimonides address, in part, the biological needs of a human servant (e.g., *feed the servant every dish the master is fed*), the practices, in general, express the need for dignity, compassion and consideration – practices that every virtuous individual must pursue, whether his interlocuter is human or, as is my thesis, humanoid. Accordingly, while *feeding the servant first* is not relevant, saying “please” and “thank you” is relevant, part and parcel of the requirement to *speak calmly*. Similarly, while *feeding the servant every dish the master is fed* is inapplicable, not raising one’s voice in anger nor one’s hand in violence is most applicable, falling under the rubric of *not disgracing the servant verbally or physically*.

It is my contention that this master–slave relationship delineated by Maimonides provides an eminently reasonable paradigm for interacting with the social robot, one that can provide a resolution to the VAD (as well as the ADP). Starting with the “virtue”

part of the “Virtue Authenticity Dialectic”, the VSO model demands that we abide by the highest ideals of a virtuous relationship, thus distancing us from the dehumanization trap. This, of course, is the approach taken by Cappuccio et al., and really the whole “appearances” camp, which leads to the problems associated with anthropomorphizing. However, whereas Cappuccio et al., shun the slave-like relationship as “disturbing,” VSO embraces it in virtue. VSO defines the SR as our slave, our property, our instrument, all the while commanding us to behave virtuously with it, treating it as an end. Relating to the SR not merely as an instrument, but as an end, allows us to maintain our own virtuous character. Keeping the SR on the level of instrument, allows us to avoid bringing it in to our moral circle and thus avoid *most* of the Pandora’s box of misplaced moral status issues.

I say “most” because we are still left with the “authenticity” part of the “Virtue Authenticity Dialectic.” That is, if we are interacting with the SR as an end, treating it in the most virtuous of ways, we will, in the words of Turkle, “imagine that the robot is an ‘other’” – i.e., a being to engage with emotionally. How, then, can we retain our appreciation for authentic, reciprocal, relationships – relationships in which both parties understand, in the deepest sense, what they themselves are thinking, saying, and doing?²⁹ How can we remain cognizant of the value of mind-ful humans over mind-less humanoids?

I suggest that it is precisely by framing the relationship in terms of master-slave that we maintain our distance and are ever brought back to the reality that we are interacting with a machine and not the noblest of creations – a conscious human being. The VSO paradigm holds that, while we maintain a virtuous relationship with the SR, we nevertheless bind that relationship in the rubric of master-slave. In so doing, we are forced to abandon the thought that we are having an authentic relationship for the simple reason that such would imply we are, in fact, slaveholders! This would then implicate us as being in violation of the fundamental principles we hold dear: freedom and equality for all humanity. It is, then, the very designation – “slave” – that awakens in

²⁹ On the importance of authentic reciprocal relationships, see, e.g., Turkle 2011a: 6, 2011b: 64; Richardson 2016: 51; Prescott 2017: 143; Bertolini and Arian 2020; Nyholm 2020: 111-2. Similarly, Verugio and Abney, 2012: 355.

us the realization that the relationship with the SR is not authentic, that “insides count” and that authenticity is precious, to be found only in conscious beings.

And is this not what the name robot was supposed to denote from its very beginning? Karl Capek coined the name robot from the Czech word *robota* meaning “forced labor.” But the name robot has since lost its original intent and so a more telling appellation is of the essence. “Slave,” though repugnant to modern ears, is really the term that drives home the idea of the robot, for it is precisely this repugnance that allows us to use the SR as the tool it was made for and not as the friend it appears to be.³⁰ Nevertheless, due to the negatively charged nature of the term (see, e.g., Miller 2017: 298, Gunkel 2018: 131), I suggest we use the “less polarizing” term, to quote Gunkel (*ibid.*: 130), of “servant.” And while thinkers such as Mark Coeckelbergh (2015: 224) question if there is a difference in the terms, I believe there is a world of difference – one that turns on Kant’s prescription for human relations. Slave implies chattel, treated as a mere means. Servant implies worker, with the potential to be treated as an end (see, e.g., Bryson 2010). Slave, according to Steve Petersen (2007: 45), implies working against one’s will; servant implies *wanting* to work. Certainly a mindless humanoid cannot be considered as working against its will, for it has no “will,” and though it similarly has no “wants,” by being programmed to serve it could be considered, anthropomorphically, as *wanting* to serve.³¹

Getting the Metaphor Right

That said, whether slave or servant, the metaphor has given rise to numerous objections. Objections that, as Joanna Bryson has contended in her now infamous piece “Robots Should be Slaves,” eventuate from failure to “get the metaphor right.” By this she refers to the fact that metaphors are imprecise. We use metaphors as tools, conceptual tools, that allow us to think about things we don’t know by comparing them to things we do

³⁰ And marking the SR as non-human, or even making it look completely non-human, is untenable because of the great advantages to having them as humanlike as possible (see, e.g., Scheutz 2014b: 209; Ghigino and Wykowska 2020: 55)

³¹ It should be noted that Petersen argues for the moral legitimacy of engineering mind-ful humanoid servants whereas I am merely discussing mind-less humanoids. Elsewhere (Petersen 2017) he notes that mindless robots certainly have no moral patiency.

know. But metaphors, by definition, are limited – “there is an apparent claim of identity, but ... only with respect to certain characteristics” (Ortony 1975: 52; see also, Jones and Millar 2017: 604). Accordingly, the slave metaphor is to be used to address the question of the moral interaction with mindless humanoids not as if it entailed identity but only as a rough conceptual paradigm.

And this is where thinkers, as described by David Gunkel, run in to trouble; for, in the effort to demonstrate that robots should not be slaves, that the slave metaphor “may be the wrong metaphor” (2018: 131), the metaphor is assumed to entail identity – i.e., that what is true for human slaves is true for robots. To take but one example, it is explained that slaves have criminal responsibility in Jewish, Roman and United States law, yet applying this to robots is problematic since punishment works only if something matters to the punished (ibid: 123-5). The metaphor is thus stretched to imply its failure. But “getting the metaphor right” means applying it judiciously.

Bryson (2020a) herself writes: “The mistake I made with that title [“Robots Should be Slaves”] was this belief that everyone was sensitive to the truth that you can’t own people. The word slave here is about something else.” That is, the metaphor only goes so far, robots are to be slaves in the sense that their function is to serve human needs and in the sense that they have no responsibility for their actions and in the sense that we have no direct moral responsibilities toward them (similarly, Grau 2011: 458).

Veruggio and Abney note that, indeed, it is impossible to apply all of the moral implications latent in the term “slave” to mindless humanoids, for “in reality, our robots are not (for now, anyway) our ‘slaves’ *in any robust sense*, as they have no will of their own” (2012: 352, *emphasis added*). Again, any use of the term “slave” can only be applied in a very limited sense – as found, for example, in computing terminology wherein slaves and masters are simply logic agents, the former accepting and executing commands at the request of the latter.

Veruggio and Abney explain that we view our relationship with robots incorrectly, incoherently, because we are “driven by our collective guilt over the history of slavery” (ibid). Now, while numerous authors have used this guilt driven approach to argue against the slave metaphor (see, e.g., Lavender 2011, Dihal 2020) no one has argued the

point more obdurately than Gregory Jerome Hampton (2015). Hampton begins by noting that the motivations for robots are the same as for slavery – i.e., cheap labor requiring the “human” touch, one that combines intelligence and dexterity. Though this is true enough, he extrapolates from here to argue that the deployment of robot slaves is identical to the deployment of human slaves. The claim is fallacious because, as Veruggio and Abney noted, robots have not a will of their own.³² The deployment of mindless humanoids, then, is more like the deployment autonomous cars – the likes of which no one imputes with slavery.

Hampton goes on to express the fear, without providing support, that the deployment of robot slaves will prompt racism. Now, while there is a concern that mistreating robots that impersonate a specific race (or gender) will “confirm and proliferate” such behavior in society at large (Coeckelbergh 2021: 7), it is hard to see why racism (or misogyny) would emerge otherwise – i.e., without mistreatment or without impersonation. That said, it could be argued that speciesism against robots could emerge, for people do unfortunately harbor ill will toward the other (see, e.g., Gunkel 2012: 207; Kim and Kim 2012; Scheutz 2014a: 249, Musial 2017: 1093). But even if speciesism were to result from deploying robot slaves, there is no reason to believe that this speciesism would prompt racism. Peter Singer (2009), who argues that humans exhibit speciesism against animals, does not argue that it has prompted or contributed to racism. He does say that all such prejudices are “aspects of the same phenomenon” – i.e., unjustifiably maintaining oneself as superior over an other (Yancy and Singer 2015). So one could raise the concern that relating to mindless humanoids as slaves will inculcate a vicious character that could harden us, to echo Kant once again, in our interactions with human beings in general, but not toward one race in particular. But this concern over inculcating a vicious character is one that has already been raised and addressed directly by the VSO paradigm which demands virtuous behavior toward humanoids (as explained in the VSO section above).

³² One could argue in his defense that he is, in fact, referring to mindful robots, however he writes explicitly that he refers to “anything resembling an independent consciousness” (2015: x), which readily includes mindless humanoids, as noted in my Introduction.

Another claim against deploying humanoid robots as slaves is made by Kevin LaGrandeur (2011: 237) who applies Aristotle's warning to beware of powerful slaves who will revolt. That is, once slaves become more powerful than their masters – be they human or humanoid – they will revolt. This may be an issue for “strong AI,” as LaGrandeur states, but a mindless humanoid, while more powerful than humans in many respects, does not have an autonomous will to revolt, indeed, does not have an autonomous will period. Accordingly, this concern is of no consequence with respect to mindless humanoids.

That said, LaGrandeur argues that the mere interdependency of slave-systems with their human operators gives rise to what could be considered a “slave revolt” in the sense that the systems are delegated so much control that humans no longer control or even understand what the slave-systems are doing. We are reminded here of Hegel's master-slave dialectic in which masters, by dependence on their slaves, lose touch with reality (Hegel [1807] 2019). Mark Coeckelbergh, in his “The Tragedy of the Master: Automation, Vulnerability, and Distance” (2015), applies this dialectic to automation in general, and to AI and robots in particular, explaining that robots as slaves will bring upon us the tragedy of which Hegel warned: dependency on automation and alienation from nature. While this may indeed be true, it is neither a reason to stop the advance of automation nor to dissuade use of the master-slave paradigm. For, though the robot as slave, as with all automation, may bring dependency and alienation, it will also provide the boon of freedom from all the burdens inherent in taming nature to human needs. And employing the robot as slave will no more entail these negative “Hegelian” consequences than relating to the robot as companion – in any case, the very automation will engender dependency and alienation. That is the price of freedom from our burdens.

Additionally, Coeckelbergh (2015) argues against using the slave metaphor for we thus limit “the range of human–technology relations” when there are “different roles for, say, robots.” While clearly there are many roles robots can play, in speaking about SRs, they all assume human-like roles – whether as care-takers of the elderly, cleaning maids, teachers or hotel concierge – and they all accommodate the servant metaphor without inappropriately reducing the range of relations. The only role that the slave metaphor limits is “companion,” and this role, I believe, is one that should be proscribed. For,

engaging socially with robo-companions may lead to the social catastrophe of shunning human companions, as Turkle notes, because they are “sometimes messy, often frustrating, and always complex” (2011a: 7, 295; see also, e.g., Richardson 2015: 12; Gerdes 2016: 277; Bertolini 2018: 653).

Now, while many of the above arguments against using the slave metaphor are based on the “dehumanizing” nature of the term, Birhane and van Dijk (2020) argue that the metaphor should be eschewed because it “humanizes” the machine. That is, the term “slave,” while clearly dehumanizing when applied to mind-ful humans, is paradoxically humanizing when applied to mindless humanoids. By calling a robot a “slave,” they claim, we employ a term reserved for humans and thus implicitly make it human; and as a result, we then find ourselves in the immoral position of a slaveowner. To their claims I have two responses. First, the term does not serve to humanize the humanoid any more than our own natural anthropomorphizing of it does – i.e., in any case, as noted above, we “humanize” it. Second, and more to the point, the fact that we will find ourselves in the immoral position of slaveholder is a welcome implication, as explained previously, that forces us to abandon the illusion that we are interacting with a human being, loathe as we are to be found in violation of the freedoms of a conscious being.

One might counter that many (or most) people will not be so loathe. Yet this is precisely what VSO comes to address. VSO is to be seen as a kind of “user instruction manual” requiring the user/owner to relate to their humanoid servant in a virtuous manner. And while a user manual is no guarantee against user abuses, given that VSO requires the master to “*listen to the complaints*” of his servant, VSO concomitantly requires that the humanoid itself be programmed to provide moral feedback/pushback, reminding the master of his duties (similarly, Darling 2016, Cappuccio et al. 2020). One can imagine an abusive owner screaming epithets while their robo-servant calmly objects with rational feedback. Will this tame the beast? The answer is irrelevant because such an interchange already removes Birhane and van Dijk’s objection that the human will become a slave owner. For, a slave, in the face of such abuse, would cower in submission not persist in moral exhortations and refusal to comply. Accordingly, without an obsequious entity to comply, there is no position for an immoral slaveowner to occupy.

This could, however, lead to the master becoming so frustrated that he “kill” his robo-servant. But there can be no “killing” of a mindless machine, only a powering down. Interestingly, it was precisely due to this moral fallacy that Bryson originally applied the slave metaphor. Shocked that people expressed repugnance at the idea of turning off a mindless humanoid, she went on a campaign to decry the notion that a mindless humanoid had moral patiency (Bryson 2016). When her efforts failed, she decided to employ the slave metaphor to emphasize that we *can* turn humanoids off. She did not mean to imply that we can kill human slaves but only that we must realize that the humanoid robot is built to serve, that they are, in her words: “tools to extend our abilities and increase our efficiency in a way analogous to the way that a large proportion of professional society, historically, used to extend their own abilities with servants” (2010). The servant metaphor, then, was meant to be applied in the sense that mindless humanoids are like servants functionally, i.e., in the operations they perform. It was not meant to humanize nor to imply an identity to human slaves, and though there is admittedly ambiguity here, she meant just the opposite – i.e., the mindless humanoid has not rights nor feelings nor anything human-like that would engender moral patiency. That, she explains, is “getting the metaphor right.”

Conclusion

In this chapter I have taken up the most unpopular position of defending the indefensible: slavery. Of course, I am in no way, shape, or form, advocating human slavery but rather appropriating the paradigm, the metaphor, if you will, in its most virtuous form to guide human interactions with mindless humanoids. I have taken this position, despite the opposition voiced in much of the philosophic community, because I believe that human authenticity, human worth, and human-human relationships are at stake. If we do not appreciate that we are more than “meat-machines” and that our relationships with each other are more than instrumental, we will fail ourselves as human beings and usher in a world of untold moral calamity. It is a category mistake to equate man and machine. The VSO paradigm counters this mistake by maintaining a clear distinction between man and machine, all the while asking man to cultivate virtue in his interaction with machine.

Does this resolve the dilemma inherent in the Virtue-Authenticity Dialectic? As mentioned before, dilemmas are so designated because they have no perfect resolution. I admit that it is problematic to call an entity that appears human-like a “slave,” or even, a “servant.” I admit that engaging with human-like SRs makes it difficult to disassociate them from real humans. Nevertheless, given the options, I suggest that being a Virtuous Servant Owner allows us to maintain our own virtuous disposition on the one hand, while preserving our appreciation for human authenticity and authentic relationships, on the other.

Accordingly, whereas Cappuccio et. al sought a way to remove the “alienating representations of slavery,” I suggest that it is specifically this alienation that is redeeming. It can allow us to define a new ontological category, not human, not animal, but slave/servant – i.e., animated autonomous tool. And we need not fear the reinstatement of human slavery, for with the introduction of robots as animated autonomous tools, we will eliminate any advantage of human slaves – exactly as Aristotle envisioned.³³

³³ Note that even Mark Coeckelbergh (2015: 227) admits this point.

Ch. 3:

Article #3 - To Make a Mind

A Primer on Consciousness Machines

Introduction

Throughout history, people have dreamed of creating artificial life – be it in the form of automaton, homunculus, golem, or robot (see, e.g., LaGrandeur 2013, Mayor 2018). These “creatures” were the stuff of myth, the work of magicians and mystics whose motivations were as varied as their methods. Using alchemy and incantation, prayer and meditation, they sought to make everything from servants and companions to guards and entertainers, with some seeking to make oracles to provide advice and others simply seeking to demonstrate their ability to create life. Perhaps unsurprisingly, given that “there is nothing new under the sun,” all of these “applications” remain at the forefront of modern-day motivations to create artificial life, for:

- Where Hephaestus made Tripods to serve (Homer, *Iliad*, Book 18), we are making social robots to help with everything from childcare to geriatric care. Yet work remains, as true caregivers need the capacity for empathy, understanding, and more (e.g., Bertolini and Arian 2020, Bertolini 2018, Sparrow and Sparrow 2006, Turkle 2011a, Nyholm 2020).¹
- And where Pygmalion made Galatea to love (Ovid, *Metamorphoses*), we are making love-bots for the same purposes. Yet work remains, as true love cannot be unidirectional, impersonal, programmed, etc. (e.g., Sparrow and Sparrow 2006, Turkle 2011a, Scheutz 2014b, Richardson 2016, Prescott 2017, Hauskeller 2017, Lumbreras 2018, Agar 2019, Bertolini and Arian 2020, Nyholm 2020).²
- And where Hephaestus made Talos to guard (Apollonius, *Argonautica*, Book 4), we are making Autonomous Weapons Systems to fulfill all our soldiering needs. Yet work remains, as life-and-death decisions demand reflective deliberation (e.g., Asaro 2006, Sparrow 2007, Purves, Jenkins, and Strawser 2015, Sparrow 2016, Sharkey 2018).³

¹ See, however, Floridi 2008, Sharkey and Sharkey 2010, Anderson and Anderson 2014, Coeckelbergh 2016, and The Medical Futurist 2018, who point up advantages to current “mindless” robo-caregivers.

² Dissenters include Levy 2008, Levy 2009, Singer in Hauskeller 2017, Danaher 2019a.

³ See, however, Arkin 2010; Leveringhaus 2016; Muller 2016. For discussion and sources, see Navon 2023.

- And where Aristotle dreamed of automaton’s to “pluck the lyre” (*Politics*, I:IV), we are making Generative Artificial Intelligence to not only play music but compose it (and that is not to mention the creativity of Gen-AI in all other art forms). Yet work remains, as true creativity requires freewill (Du Sautoy in Smith and Schillaci 2021) and the capacity to discern which novel compositions are worthy creations (Smith and Schillaci 2021, Eagleman in Slack 2023), among other things (e.g., Jefferson 1949: 1110).⁴
- And where Medieval European philosophers made artificial talking heads to provide advice (LaGrandeur 2013: 82), we are making Large Language Models (LLMs) to do even better. Yet work remains, as true advice requires empathy, responsibility and much much more (e.g., Lichtenstein 2002; Sparrow 2021).⁵
- And where mystics sought to make a conscious humanoid in order to achieve the ultimate in human creativity (LaGrandeur 2013: 71, Idel 2019: xxvii), today scientists have dedicated themselves to this effort proclaiming it to be “the greatest technological achievement of our species,” (Stephen Younger in Anthes 2001).⁶ Indeed, though there are debates over the necessity for consciousness in the above applications,⁷ ultimately, only an entity with human-level consciousness will have human-level competencies (e.g., Penrose 1991: 419, Signorelli 2018: 8, Gocke 2020, McFadden 2020: 2).

⁴ Arthur Miller (2020) shows how mindless machines (i.e., machines without consciousness) have generated creative works in every area imaginable. Nevertheless, based on Margret Boden (ibid.: Ch. 40), he explains that there is a difference between little-c creativity (i.e., local novelty) versus big-C creativity (i.e., transformational idea with historic novelty). Boden argues that mindless machines have been successful at the former but not the latter; Miller, however, notes that it is only a matter of time.

⁵ The discussion surrounding programmable advisors, especially moral advisors, is highly disputed (see fn. 25 herein).

⁶ That we will make a humanoid simply because we can, see, e.g., Duffy 2003: 188, Neely 2014: 104, Miller 2017: 300. Of note is the observation of Geoffrey Jefferson who, writing at the beginning of the computing revolution, commented that humanity has ever been driven by the motto, “It could be, therefore it was” (1949: 1106).

⁷ See above footnotes on the applications, as well as comparison of the “insides count camp” versus the “appearances count camp” in Ch. 2 “The Virtuous Servant Owner” [Navon 2021].

Consequently, at the very core of the enterprise to create life, whether by ancients or moderns, is the endeavor to imbue a synthetic being with consciousness – i.e., the faculty of that ethereal entity known in premodern times as a “soul” and referred to today as a “mind.” The only difference between our modern efforts and those of yore, is that ours are not carried out in dark dungeons by magicians or mystics, but in well-lit laboratories by scientists and engineers. Yet, despite all the light of modernity, the endeavor to make a mind remains, to many, as shrouded in mystery as the ancient myths that sought to conjure one. Accordingly, in an effort to dispel the darkness, this chapter seeks to provide a primer to the concepts central to our modern efforts To Make a Mind.

Consciousness

To begin, the mind is where consciousness takes place (Locke 1690: II:1:19), and it is consciousness that is arguably *the* defining ontological feature of human beings.⁸ Indeed, while there is a long and venerable list of features describing the ontology of human beings, it is consciousness that stands out as the one upon which the rest depend, as can be seen as follows:⁹

1. **Born In or Of a Human Mother** (Onkelos Gen. 3:20; Hacham Tzvi 93; Teshuva MiAhava 53; Bleich 1998: 68; Loike and Tendler 2003: 3).

⁸ See, e.g., Penrose 1991: 9; Chalmers 1996: 26; Asaro 2006; Walker 2006a: 3; Kamm 2007: 229; Torrance 2008: 507; Levy 2009: 214; Grau 2011: 457-8; Veruggio and Abney 2012: 253; Anderson 2013; Basl 2014; Neely 2014; Eskens 2017; Prescott 2017; Lumbreras 2018; Mackenzie 2018: 6; Signorelli 2018; Agar 2019: 270; Miller 2020: Ch.43; Kingwell 2020: 329; Liao 2020c: 497; Nyholm 2020:199; Schwitzgebel and Garza 2020: 464, 473; Andreotta 2021; Kohler 2023. Dissenters do not deny that consciousness is paramount but that (a) it is not defined well enough to be of practical use, and (b) it is not discernable beyond behavior (e.g., Gunkel 2018: 98-100).

⁹ It is important to note that this list, built from the numerous sources cited, expresses the various characteristics and capacities for a fully actualized human being and by no means represents the minimal requirements to grant human-level moral status (e.g., to fetuses, newborns, people in irreversible comas, anencephalics, cognitively impaired, etc.). For a treatment of these thorny questions see Liao 2010. See also Dennett 1976; Loike and Tendler 2003; Foerst 2009; Kagan 2015; Liao 2020c).

2. **Produce Offspring with a Human** (Tosefta Berachot 1:5; San. 58a, Rashi Rosh Hashana 4a; Maimonides Hil. Melachim 9:5; Loike and Tendler 2003: 7; Signorelli 2018: 3).
3. **Embodiment**¹⁰ (Gen. 2:7; Rashi ad loc.; Leviticus Rabbah 34:3; Luzzatto *Derech Hashem* 1:3; Dreyfus 1992; Foerst 1998: 100; Zlatev 2001: 161; Peters 2005: 388; Torrance 2008; Coeckelbergh 2010b: 237; Rakover 2011; Turkle 2011a: 134; Hoffman 2012; Tallis 2012: 351; Bishop and Nasuto 2013; Parthemore and Whitby 2013; Susser 2013; Thompson in Torrance 2014; Torrance 2014: 16; Mercer 2015; Damasio in Jiménez-Rodríguez 2017: 194; Signorelli 2018: 3; Collins in Boucher 2019: 9).
4. **Unique Identity** (Mish. San. 4:5; San. 37a; San. 38a; Ber. 58a; J. Ber. 9:1 [57a]; Rosenfeld 1977: 72-3; Soloveitchik 1983: 125; Spero 2003: 30; Pruss 2009; Mercer 2015: 181; Damasio in Jiménez-Rodríguez 2017: 195).
5. **Finitude**¹¹ (Gen. 2:17; Gen. 3:22; Tempkin in Liao 2014: 114; Damasio in Jiménez-Rodríguez 2017: 199).
6. **Intelligence**¹² / **Moral Intelligence**¹³ (Aristotle *NE* X:7; Maimonides *Guide* 3:54; Maimonides *Guide* 1:1 [see esp. Soloveitchik 2003: 46]; Maimonides Hil. Teshuva 5:3; Rashi Gen. 2:7, Rashi Gen. 3:22; Luzzatto *Derech Hashem* 3:1; Lamm 1965: 33; Dennett 1976; Loike and Tendler 2003: 4; Kant, Rawls, Scanlon in Liao 2014: 113).

¹⁰ Embodiment is understood to be required in order to allow for cognitive development and consciousness itself (a.k.a., enactivism).

¹¹ Humans are shaped by their mortality, yet this is not an essential attribute since ultimately, “death will be swallowed up” (Isaiah 25:8).

¹² Related to the development of intelligence is the capacity of “wonder” for, as Aristotle wrote: “it is through wonder that men now begin, and originally began, to philosophize [i.e., investigate physics and metaphysics]” (Metaphysics 982b; see also Maimonides Hil. Avoda Zara 1:3).

¹³ Included here is “Moral Agency – the capacity to regulate one’s own actions through moral principles or ideals” (Warren 1997 in Anderson 2011b: 292; similarly, e.g., Dennett 1976, Damasio in Jiménez-Rodríguez 2017: 199).

7. **Language / Inner Speech**¹⁴ (Rashi Gen. 2:7; Nachmanides ad loc.; Maharal, Maharsha, Marit HaAyin, et al. [San. 65b, s.v. *rava*]; Dennett 1976; Warren 1997 in Anderson 2011b: 292; Damasio in Jiménez-Rodríguez 2017: 195).
8. **Creativity** (Soloveitchik [1965] 2012: 8; Lamm 1965: 40; Soloveitchik 1983: 132; Gesche in Jiménez-Rodríguez 2017: 203; Lorrimar 2017; Signorelli 2018).
9. **Freewill / Autonomy** (Maimonides Hil. Teshuva 5:1; Kant [1785] 2006; Lamm 1965: 33; Bringsjord 2007; Moor 2011; Parthemore and Whitby 2013; Johnson and Axinn 2014; Signorelli 2018: 3).
10. **Autopoiesis**¹⁵ (Torrance 2008: 513-4; Parthemore and Whitby 2013; Signorelli 2018).
11. **Imagination** (Luzzatto *Derech Hashem* 1:3; Mill [1863] 2017: 7; Zlatev 2001: 185; Johnson and Axinn 2014: 2; Harari 2015).
12. **Intentionality**¹⁶ (Dennett 1976;¹⁷ Searle 1980; Moor 2011).
13. **Emotion / Empathy** (Hume 1739; Mill [1863] 2017: 7; Picard 1995; Warren 1997 in Anderson 2011b: 292; Torrance 2008: 510; Coeckelbergh 2010b; Anderson 2011a: 164; Torrance 2011: 126; Scheutz 2014a: 250; Rodogno 2016; Bedzow 2017: 267; Damasio in Jiménez-Rodríguez 2017: 197; Signorelli 2018; Coeckelbergh 2020a: 51).
14. **Temptation / Yetzer Hara**¹⁸ (Gen. 3:6; Rashi Gen. 2:25; Eisen 1992; Zoloth 2008: 22; McDermott 2011: 106; Conradie 2017).

¹⁴ It should be noted here that the capacity for language has long been linked to intelligence (Aristotle, *Politics* I:II; Descartes, *Discourse* V; Hobbes, *Leviathan* 1:4) as will be elaborated below.

¹⁵ Autopoiesis, in its simplest form, refers to an entity's capacity for self-organizing, self-maintaining, self-recreating (Torrance 2008: 513), which would include the capacity for homeostasis (Damasio in Jiménez-Rodríguez 2017: 198). In its more developed form, autopoiesis refers to an entity's teleology – i.e., it has a purpose and seeks to fulfill it (see esp. Torrance 2008: 514). Without using this label, the notion of purposiveness is clearly in the list of human makeup (see, Aristotle *NE* 1:7, Maimonides *Guide* 3:54, Luzzatto *Derech Hashem* 1:2-3; Maslow 1943; Frankl [1946] 2006, Soloveitchik 1983: 132, Damasio in Jiménez-Rodríguez 2017: 198). It should be noted that this feature is closely related to embodiment (enactivism), autonomy and intentionality.

¹⁶ Intentionality refers to the mental capacity to be directed by one's beliefs, desires, hopes and fears (Dennett 1976: 179).

¹⁷ For the sake of completeness, Dennett brings the related required feature of “reciprocity,” which he defines as the capacity to exhibit higher-order intentions (ibid.: 185).

15. **Self-Awareness**¹⁹ (Warren 1997 in Anderson 2011b: 292; Morin 2006;²⁰ Bringsjord 2010; Neely 2014; Hauskeller 2017²¹) / **Self-Reflection** (Heschel 1965: 9; Dennett 1976: 193; Parthemore and Whitby 2013; Signorelli 2018).
16. **Subjective Experience / Sentience** (Nagel 1974; Warren 1997 in Anderson 2011b: 292; Huebner 2009; Signorelli 2018).
17. **Imago Dei / Tzelem Elokim** (Gen. 1:27; Gen. 9:6).²²
18. **Soul / Neshamah** (Gen. 2:7).²³
19. **Second-Order Phenomenal Consciousness** (See sources in fn. 8).

This list, as ordered, can be divided into 3 distinct categories. Items 1-5 are really species specific as they define the criteria for humanhood (i.e., *Homo sapiens*) as opposed to personhood (i.e., beings that demand humanlike moral status).²⁴ Now, while the remaining items are related to various levels of consciousness that require elaboration (a review of which will be provided below), the following short categorization is instructive. Specifically, items 6-15 can be seen as related to, or dependent on consciousness, item 6 being at the center of the debate regarding machine capabilities – i.e., does moral intelligence require second-order phenomenal consciousness or is functional consciousness sufficient?²⁵ Item 16 refers to first-order phenomenal consciousness, for

¹⁸ Included here is the ability to overcome temptation (a.k.a., “akrasia,” see Parthemore and Whitby 2013). See also Midrash Tanhuma Gen. 7:7. Accordingly, this is related to moral agency, see above fn. 13.

¹⁹ Importantly, achieving this requires more than just the ability to recognize oneself in a mirror, which is a basic ability that can be easily programmed into a machine. (Morin 2006: 367).

²⁰ Morin notes that self-awareness includes the need for “inner speech” (2006: 368).

²¹ Hauskeller couples the need for self-awareness with that of “self-concern” (2017: 1).

²² The sources and meanings of this term will be elaborated herein below.

²³ The literature on the soul is too vast to cite. For modern philosophical/theological approaches to the notion of the soul, see Moreland 2009, Swinburne 2019.

²⁴ I refer here to moral personhood as opposed to legal personhood, the former seeking to define an entity as a moral agent the latter being a method of conferring rights upon an entity (Calverley 2011: 218).

²⁵ Many hold that second-order phenomenal consciousness is essential for moral judgement as it requires, e.g., emotions, empathy, intuition, intentionality, self-reflection, understanding of human nature, culture, experience (see, e.g., Penrose 1991; Moor 1995; Spero 2003: 30 n. 43; Johnson 2011; Moor 2011;

which, when coupled with the prior items 6-14 engenders or expresses second-order phenomenal consciousness. Finally, items 17-19 essentially name second-order phenomenal consciousness. That is to say, item 19 is the genuine article and items 17 and 18 are simply the religious references to what moderns call the mind – i.e., the metaphysical seat of second-order phenomenal consciousness.

To appreciate this categorization, and particularly the notion of the mind as the foundation of human *being* itself, it is important to review the fundamental concepts of consciousness.²⁶ To begin, the unmodified term “consciousness” generally refers to, what philosophers call, “phenomenal consciousness,” and is roughly defined as: “experience” itself (Block 1995), “the felt quality of one’s inner experience” (Schneider 2020), the “what it is like” qualitative subjective aspect of being (Nagel 1974). In contradistinction to this familiar experiential consciousness, there is what is referred to as “functional consciousness” (Franklin 2003), “access consciousness” (Block 1995), or “cognitive consciousness” (Torrance 2008, Schneider 2020), roughly defined as the rational processing of information – i.e., cognition without experience. It should be noted that while some, a.k.a. functionalists, believe this is consciousness *in toto* (e.g., Dennett 2007), or that it is all that matters (e.g., Davenport 2014: 56), there is strong opposition to these positions (e.g., Torrance 2008: 499).

Now, both types of consciousness support the notion of “orders of consciousness.” First-order consciousness consists in the capacity to think about things and is generally

Parthemore and Whitby 2013; Purves, Jenkins, and Strawser 2015; Rodogno 2016; Coeckelbergh 2020a). Others understand moral decision making as the functional application of codes and principles, thus requiring only functional consciousness (see, e.g., Davenport 2014, also Floridi and Sanders 2004, Nadeau 2006, Sullins 2006, Wallach and Allen 2009 all quoted in Gunkel 2012: Sec. 1.5). This latter position aligns with the “codifiability thesis” which holds that a “moral theory can be captured in universal rules that the morally uneducated person could competently apply in any situation” (Purves, et al. 2015). There are many who reject this position outright, arguing for the “anti-codifiability thesis” (see sources in Purves, et al. 2015: 856). Yet, even if ethics cannot be fully codified, Susan Anderson (2011a: 167) offers a practical compromise: Through the combination of programmed ethical principles along with machine learning the mindless machine realizes ethics much like humans do.

²⁶ Please note that, as my objective here is to present a general overview, the discussion intentionally glosses over contentious points within the field of Philosophy of Mind.

associated with animals – e.g., a dog thinks about a bone. It is referred to as first-order in that the content of the thought at hand is that which is first perceived – e.g., the bone. Then there is second-order consciousness, which is a more sophisticated mental capacity whereby the content of what one has perceived is represented to oneself. It allows thinking about thinking and is generally associated with human beings. For example, a human being not only thinks about the steak being eaten, but can also entertain thoughts like: why am I eating a steak, what are the implications of eating steak for me, for the cow, for the environment, etc.

These descriptions, as said, can be true “functionally” (i.e., they describe cognitive information processing), or they can include a “phenomenal” component (i.e., they describe the subjective experience attendant to the cognitive processing). In the case of *functional* consciousness, orders of consciousness can be understood as layered representations of concepts, such that second-order functional consciousness realizes “thinking about thinking” in purely symbolic terms. That is, one symbolic representation of an object or concept is simply replaced by a different symbol. Computers can, and do, perform this all the time. Consider the example of an autonomous vehicle being controlled by a computer. The computer receives an image of the car ahead, captured by the vehicle’s camera and represented as an array of pixels, which constitute a first-order representation. The computer then processes this image through an image-processing filter (e.g., median filter), producing a modified array of pixels that characterize the same car and thus constituting a second-order representation.

Second-order *phenomenal* consciousness, on the other hand, has an experiential component to it. Representations here are not merely symbolic (a.k.a., syntactic) but semantic – they have meaning. This distinction was made famous by John Searle’s “Chinese Room” (Searle 1980). The Chinese Room is a thought experiment that posits a man, who knows not Chinese, sitting inside a room full of books filled with Chinese questions and answers. A woman on the outside, conversant in Chinese, slips the man a question in Chinese through a slot in the door. To respond, the man takes the question, finds it in the symbol tables and copies the corresponding answer on a piece of paper –

all the while having no understanding of the question or the answer. His is an act of pure symbol manipulation.²⁷

Similarly in our car example, if a computer in an autonomous vehicle analyzes the car it is following (e.g., using various image processing filters), while able to “recognize” the car, symbolically manipulate its pixels and even provide responses as to what the car should do (e.g., apply the brakes when the car ahead is too close), it does not know what a “car” is, it does not know the *meaning* of “car.” This is because autonomous vehicles lack phenomenal consciousness, which allows one to understand the meaning of one’s conscious content (e.g., a car) and, more importantly, “experience” it. For example, whereas an autonomous vehicle can recognize, cognitively, that it is following a red Ferrari and correspondingly react by applying the brakes if needed, a human being does not merely recognize the red Ferrari cognitively but *phenomenologically*, evoking the exclamation, “Wow!”²⁸

In addition to this experiential aspect of phenomenal consciousness (i.e., even first-order), higher-order phenomenal consciousness allows for “self-consciousness.” It is ultimately this type of consciousness that distinguishes humans from all other entities. It allows humans to not only be self-aware but self-reflective. And here it should be noted that while some higher functioning animals have been shown to have rudimentary self-awareness, this does not indicate that they have the kind of deep thinking available to humans (see, e.g., Morin 2006, Tallis 2012: 94, Holland 2018: 114). Indeed, it is not simple self-awareness that distinguishes human beings but the kind of self-awareness that

²⁷ Note: while a computer engineer like myself finds Searle’s position intuitive, it has generated significant debate (see, e.g., Cole 2020).

²⁸ For the sake of completeness, it is important to note that Searle’s Chinese room demonstrates the need for the understanding of language (requiring “intentionality”), whereas in my autonomous vehicle example the understanding is in visual perception (requiring “phenomenal” experience). That said, the analogy between the two is efficacious for two reasons. On the visceral level, the Chinese room induces our intuitions regarding the distinction between “functional” and “phenomenological.” On the technical level, intentionality is directly related to phenomenology, as explained by Terence Horgan: “the real moral of Searle’s Chinese room thought experiment is that genuine original intentionality requires the presence of internal states with intrinsic phenomenal character that is inherently intentional...” (2013). I thank Prof. Alon Chasid for bringing this to my attention.

is characterized by self-reflection – i.e., the ability to analyze one’s own thoughts, to reflect on one’s own beliefs, desires and intentions. A human being’s self-awareness includes not only the ability to represent oneself to oneself but more sophisticated representations, like representing oneself as one believes others perceive them – i.e., I can think about how I think you think about me (see, e.g., Parthemore and Whitby 2013: 5).

Now, to allow for this higher-order conscious activity, it is widely held that language is key:

“Anthropologist Margret Mead once observed that sophisticated language is a prerequisite for sophisticated thinking, and *a fortiori* for expressing such thinking” (Jakobovits 2000: 195).

“While [Frans] de Waal likens animal and human cognition, he clearly differentiates human and animal consciousness ... After denying that even great apes can understand death in the abstract, much less anticipate their own, de Waal states, “You won’t often hear me say something like this, but I consider us the only linguistic species.” Moreover, he locates at least part of the reason for this uniqueness in the fact that animal communication “is almost entirely restricted to the here and now” (Holland 2018: 114).

“A growing numbers of researchers (e.g., Briscoe, 2002; Carruthers, 1998; Dennett, 1991; Morin, 2005; Stamenov, 2003; Steels, 2003) maintain that more complex types of self-awareness necessitate language, and more specifically, inner speech.” (Morin 2006: 368; see also Zlatev 2001: 185, Gamez 2008: 902).²⁹

Interestingly, this tight coupling of speech to second-order phenomenal consciousness is also found in Jewish literature that speaks of the soul. On the verse (Gen. 2:7) that describes man being created as a “*nefesh hayah*” (a living soul), the great medieval biblical commentator, R. Shlomo Yitzhaki (Rashi), writes that while animals too have a *nefesh*,

²⁹ Worthy of note is the dispute as to “whether it is reason that gives rise to speech, or speech which is a prerequisite for the acquisition of reason” (Bleich 1983a: 369).

man's *nefesh* includes “*deab v'dibbur*” – understanding and speech.³⁰ Nachmanides (ad loc.), writing in medieval Spain, concurs, explaining that, “by virtue of this soul, [man] understands (*yaskil*) and speaks and performs all of his activities.” And R. Pinchas Horowitz (*Sefer Habrit* 1:18) elaborates on these activities as follows:

The intellectual soul (*nefesh ha'sichli'it*) is that by which the Creator made man unique from all others. And it is called the intellectual mind (*sechel iyuni*), with which he will understand, perceive, acquire wisdom, distinguish between reprehensible and proper actions, and know truth from falsehood. And it [i.e., *sechel iyuni*] allows him free choice – primarily to choose the truth because it is true (and not because he wants to so choose or desires it). And it [i.e., *sechel iyuni*] is the form of the human being [lit. “speaker”] called the human soul (*nefesh ha'adam*).

Clearly, this human soul is responsible for everything modern philosophy attributes to the human mind; though one could argue that described here is nothing more than cognitive-consciousness. To correct such a misperception, Nachmanides (Gen. 2:7) adds to his comments on the soul, explaining that man was created, like all moving creatures, with “*bargasha*” (sentience).³¹ So too Maimonides (*Guide* 1:41) writes:

“*nefesh* (soul) is a homonymous noun, signifying the vitality which is common to all living, sentient beings (“*margish*”). ... Another signification of the term is “reason” [lit. “speaking soul”], that is, the distinguishing characteristic of man. ...

³⁰ Similarly, e.g., Onkelos, R. Bachayei, Bechor Shor, Oznam LaTorah (ad loc.); Maharsha (*Hidushei Aggada*, ad loc., s.v. *v'lo hava*); R. Leiner (*Sidrei Tabarot*, Ohalot 5). Accordingly, R. Bleich writes, “The human soul is an ontological entity and is either identical with, or the source of, man's rational faculty” (1998: 78). Worthy of note is that this notion of “*deab v'dibbur*” is found in the Greek for “word” (i.e., *logos*) which translates as both “reason” and “word” such that, “the human entity, on this account, does not just possess reason and language as faculties but is defined by this very capacity” (Gunkel 2012: 59; see also Bleich 1983a: 369). In opposition, Idel (2019: 348, fn. 31) notes an anomalous Kabbalistic opinion rooted in Pythagoras that distinguishes between speech and intellect.

³¹ So too Ibn Ezra (Gen. 9:4).

It denotes also the part of man that remains after his death. ... Lastly, it denotes “will.”³²

These sources, then, demonstrate quite unambiguously that what is now referred to as the human mind, with its associated second-order phenomenal consciousness (2OPC), is what believers in the verse “God created man in His own image,” refer to as “the image of God” (*tzelem Elokim*), or simply, the soul. And here it is important to note that the term *tzelem Elokim* (imago Dei) has a rich tradition of interpretations. Indeed, the term is understood to indicate a broad swath of human ontology and human endeavor, including:

- **Intellect / Reason** (Maimonides *Guide* 1:1; Seforno Gen. 1:26; Karo *Toldot Yitzhak* Gen. 1:27; Loike and Tendler 2003: n. 23).
- **Freewill** (Samson Raphael Hirsch Gen. 1:1; Hachohen *Mesbech Hochma* Gen 1:26; Kook *LeNevuchei HaDor* 1:1).
- **Dominion** (Saadia Gaon Gen. 1:26; Volozhiner *Nefesh Habaim* 1:3).
- **Creativity** (Zohar Hadash Gen. 9b; Soloveitchik [1965] 2012: 8; Lamm 1965: 40; Soloveitchik 1983: 132).³³
- **Morality** (Spero 2003: 30 n. 43;³⁴ Kedar 2007: 52;³⁵ Dessler *Mikhtav MeEliyahu* in Navon 2007).
- **Body**³⁶ (Samson Raphael Hirsch Gen. 1:26; Leviticus Rabbah 34:3).
- **Relationship** (Foerst 1998; Herzfeld 2002b;³⁷ Zoloth 2008: 23;³⁸ Bedzow 2017: 90).

³² So translates Michael Friedlander in Maimonides (1956).

³³ Interestingly, Adolph Gesche equates imago dei with creativity and self-creation yet explicitly disconnects it from the immaterial soul (in Jiménez-Rodríguez 2017: 203)

³⁴ Specifically, the potential to be a moral agent.

³⁵ Specifically, the fundamental moral intuition consisting of the seven commandments to Noah (see San. 56a and Epstein *Torah Temimah* Gen. 2:16 n. 39).

³⁶ The point is not that God has a human form, but that the human body entails the sacred value of God Himself.

³⁷ Noteworthy is Herzfeld’s explanation of how three categories of imago dei (i.e., substantive, functional and relational) can be seen to parallel three phases of the development of *mindless* robots (2002b; see also Green 2018). Nevertheless, such distinctions do not exist in the development of *conscious* robots wherein

- **Repentance, Repair, Responsibility** (Cherlow 2016).
- **Challenge / Mission** (Soloveitchik in Navon 2007).
- **Soul** (Christian readings: Origen 1973: II.10.7; Augustine 1948: IX, 12, 17–18; Aquinas 1981: Ia, q.90. art. 7 all in Jiménez-Rodríguez 2017: 205. Jewish readings: Midrash HaGadol, Gen. 1:27; Radak, Alshich, ad loc.; Recanati, Gen. 1:26; Maharal *Derech Hachaim* 3:14; Soloveitchik 2012: 54).

That this list of human ontology compares so closely to the list brought at the outset of this section should come as no surprise, because, while some items can be mimicked with mere functional consciousness, ultimately, what gives depth, what gives life to them all is the second-order phenomenal consciousness of the mind, otherwise known as the soul.³⁹ Furthermore, that “mind” and “soul” are used synonymously should also come as no surprise, for the former is simply the secularized version of the latter.⁴⁰ In fact, the terms “consciousness” and “mind” are relatively new, attributed to John Locke who defined consciousness as “the perception of what passes in a man’s own mind” (Locke 1690: II:1:19). John Heil, in his *Philosophy of Mind* textbook, notes that “we nowadays use ‘mind’ and ‘soul’ more or less interchangeably, [though the term ‘soul’] has moral and religious overtones missing in talk of minds” (Heil 2004: 14; also, e.g., Barresi and Martin 2012; Johnson 2013).

Machine Consciousness

Now, given that it is the mind (or soul) that serves as the definitive ontological feature of a human being, it is this feature that is sought out in building a humanoid. Accordingly, scientists and engineers have been working on building a conscious machine since the

all are given expression – i.e., as argued herein, it is consciousness itself that allows for the full actualization of these categories.

³⁸ Specifically, the capacity of love and concern for the other (similarly, Kass 2006: 38).

³⁹ For support, see all the sources to this effect in the introduction above.

⁴⁰ See also, Swinburne (2019: 2); William James (1890: Ch. 10 in Natsoulas 1991: 343); Peters (2005: 386); Coeckelbergh (2014: 63). For examples of synonymous usage, see, e.g., Descartes [1641] 2017: 4; Jefferson 1949: 1106; Turing 1950: 443; Putman 1964: 687; Boden 1985: 397; Penrose 1991: 407; Tallis 2012: 29; Shanahan 2016; Hauskeller 2017: 2.

1990s (Scheutz 2014a: 258), their efforts largely divided into three camps: infocentric, physiocentric and biocentric.⁴¹ The infocentric camp, also known as “computationalists,” believes that consciousness is simply the result of the sophisticated computation that occurs in the brain. Roughly speaking, once we build a system with computing power on par with the human brain, consciousness will emerge.⁴² Philosopher Stefan Reining explains as follows:

“According to computationalism, the mind is basically a computing system, generating certain outputs from certain inputs on the basis of some computational structure. In the case of humans, this computational structure is realized by the structure of the human brain. The crucial aspect for the possibility of strong AI [i.e., conscious artificial intelligence] is that, according to computationalism, anything replicating the structure of a conscious being’s brain will itself be conscious ... regardless of the actual physical material on the basis of which a conscious being’s brain structure has been replicated. Neurons in the human brain are carbon-based, but, according to computationalism, artificial neurons that are, say, silicon-based, will do the same job ...” (Reining 2020: 75).

Now, while there are many significant thinkers in this infocentric camp,⁴³ there are no less who balk at the very premises of computationalism, arguing that “there must be an essentially non-algorithmic ingredient in the action of consciousness” (Penrose 1991: 407; similarly, Tallis 2012: 197, Andreotta 2021: 25). These thinkers, emphasizing the

⁴¹ For variations on this categorization, see, e.g., Torrance 2011, Ladak 2022.

⁴² For clarity, it should be noted that there are various approaches on how to achieve brain-like computation, e.g., Baars’ Global Workspace Theory (Baars 2019), Shanahan’s Global Workspace Model (Shanahan 2010), Tononi’s Integrated Information Theory (Tononi et al. 2016), Reining’s Causal Topology (Reining 2020), Haikonen’s Cognitive Model (Haikonen 2019). For an overview of theories, see, e.g., Gamez 2008, 2018; also Samsonovich 2010; Ladak 2022. And then there is Whole Brain Emulation – i.e., “copying biological intelligence (without necessarily understanding it)” (Sandberg 2013; see also Sandberg and Bostrom 2008).

⁴³ Infocentrists include: Turing 1950, Haugeland 1985, Minsky 1985, Moravec 1988, Churchland and Churchland 1990, Chalmers 1996, Dennett 2007, Bostrom 2003b, Kurzweil 2006, Davenport 2014, Shanahan 2016, Haikonen 2019, Harari 2019; Liao 2020c, Reining 2020; see also sources in Torrance 2011: fn. 10.

critical importance of the *physis* of the brain, might be called “physiocentrists.”⁴⁴ Colin Hales, for example, contends that computational models are not enough to generate all the phenomenon of the system being emulated. So, just as a flight emulator does not generate lift, neither can we expect a brain emulator to generate consciousness. Accordingly, Hales believes that we must build a brain model that replicates “the brain’s charge, current, and electromagnetic field behaviours” (Hales 2014: Ch.14). He is seconded by Johnjoe McFadden who locates “the seat of consciousness in the brain’s EM [electromagnetic] field” (McFadden 2020). Roger Penrose and Stuart Hameroff delve deeper into the brain and propose that consciousness originates in cellular structures called microtubules that are specific to neuronal cells (Hameroff and Penrose 1996). Importantly, these microtubules appear to be viable only *in vivo*, thus suggesting that biology is key to consciousness (Penrose 1995: 14.3).⁴⁵

And that brings us to the Bernard Kastrup who takes the argument one step further, making the essential claim of the biocentric camp⁴⁶ – i.e., the only way to replicate all the various aspects of the brain responsible for consciousness is to make a biological brain:

If biology is the extrinsic appearance of [conscious being], then the quest for artificial consciousness boils down to abiogenesis: the artificial creation of biology from inanimate matter. If this quest succeeds, the result will again be biology, not computer simulations thereof (Kastrup 2017: 16).

Kastrup grounds his biocentric claim on the fact that (a) metabolizing organisms are the only known entities to be conscious and (b) the “flipping switches” that comprise a

⁴⁴ Physiocentrists include: Tallis 2012: 197; Hales 2014; Gamez in Johnson 2020: 277; McFadden 2020; Andreotta 2021: 25.

⁴⁵ Worthy of note is the middling position, put forward by Sorakar (2014: 34), which accepts the physiocentrists’ claim that computation is not enough to replicate consciousness but *is* enough to describe brain function in the context of human behavior.

⁴⁶ Biocentrists include: Sparrow and Sparrow 2006, Searle 2007, Torrance 2008, Damasio 2010, Kastrup 2017, Koch 2019, Koplín and Wilkinson 2019, Swinburne 2019, Bertolini and Arian 2020. Note that while all biocentrists hold biology to be essential for consciousness, they don’t necessarily express an opinion as to the viability of a biocentric humanoid robot.

computing platform – be they neurons in the brain or transistors in a computer – do not come close to characterizing the complete workings of a metabolizing organism’s brain:

Metabolizing organisms are the [only entities with conscious experience] ... [for] we are the only [entities] known to have [such] inner experiences. We also have good empirical reasons to conclude that normal metabolism is essential for the maintenance of [consciousness], for when it slows down or stops, [consciousness] seems to reduce or end. ... The differences between flipping microelectronic switches and actual metabolism are hard to overemphasize. Therefore, there is no empirical reason to believe that a collection of flipping switches could ever be what individualized, private conscious inner life looks like from the outside, even if these flipping switches perfectly mimic the patterns of information flow discernible in metabolism (Kastrup 2017: 7-16).

Similarly, John Searle, who also counts himself in the biocentric camp, defines an approach to consciousness he calls “Biological Naturalism”:

“Biological” because it emphasizes that the right level to account for the very existence of consciousness is the biological level. Consciousness is a biological phenomenon common to humans, and higher animals. We do not know how far down the phylogenetic scale it goes but we know that the processes that produce it are neuronal processes in the brain. “Naturalism” because consciousness is part of the natural world along with other biological phenomena such as photosynthesis, digestion, or mitosis, and the explanatory apparatus we need to explain it, we need anyway to explain other parts of nature (Searle 2007: 329)

Interestingly, the biocentric view that consciousness is exclusive to biological entities, aligns with the biblical view that the soul is present only in creatures with blood. The Bible repeatedly relates the soul to blood (Gen. 9:4, 9:5; Lev. 17:11, 17:14; Deut. 12:23), stating that “the soul is in the blood,” or conversely that “the blood is in the soul,” or even “the blood *is* the soul”. It must be understood, however, that while the text seems to equate to the two, the equivalence is not to be taken literally.⁴⁷ For, as Nachmanides

⁴⁷ For a good overview of the textual difficulties, see Grunfeld (1975: Vol. 1, 75-82).

(Lev. 17:14) explains, soul and blood are not the same thing but inextricably interdependent: “You will not find blood without soul, nor soul without blood... [they are interrelated] like matter and form in all physical creatures, where the one cannot be found without the other.” Furthermore, while “one cannot be found without the other,” it is “the soul that is dependent on the blood” (Rashi, ad loc.) and “not vice versa,” (Bass *Siftei Hachamim*, ad loc.). R. Isidor Grunfeld summarizes as follows: “The Torah’s dictum – ‘for the blood is the soul’ - does not mean that blood is identical with the soul but only that blood is the seat of life which in turn is governed by the soul” (Grunfeld 1975: Vol. 1, 82).

The biblical approach to the soul, then, teaches that the soul is neither the blood itself nor contained in the blood, but rather that it is only blood-based creatures that are uniquely capable of supporting a soul. Consequently, it is not metabolism, I suggest, that is indispensable for consciousness, but blood – i.e., only creatures that have “blood” (a liquid that serves a closed circulatory system) have consciousness, a mind, a soul. Using this “biblical blood hypothesis” as guide to investigating the animal kingdom, perhaps we can respond to Searle’s quandary regarding just “how far down the phylogenetic scale” consciousness goes.

According to current scientific understanding, it is known that “vertebrates (mammals, amphibians, reptiles, and birds) have red-blood cells that travel through a closed circulatory system ... [and] the blood of most mollusks, which include squid, octopus, snails, slugs, and horseshoe crabs, is blue! ... [In contrast,] flatworms, nematodes, and cnidarians (jellyfish, sea anemones, and corals) do not have a circulatory system and thus do not have blood” (UCSB Scientists 2020). Based on our “biblical blood hypothesis” we anticipate that these latter creatures to not exhibit sentience. And, in fact, while jellyfish do have a neural network (Lori 2021) and respond to stimuli, they have no subjective experience to speak of – they are zombies (Barron and Klein 2016: 4905). So too is said of flatworms (Feinberg and Mallatt 2016: 174) and nematodes (cited in both prior sources).⁴⁸

⁴⁸ For the sake of completeness, it should be noted that “phenomenal consciousness,” “subjective experience,” et al., are highly contested terms. “Where to draw the line between what is conscious and what is not, and how to justify drawing that line, remain hotly debated questions” (Barron and Klein, 2016). That said, current scientific evidence does support the claims made herein (e.g., *ibid.*; Feinberg and

To summarize, while there is a large camp of infocentrists who believe that all that is necessary to make a machine with second-order phenomenal consciousness is an appropriately configured neural network with the cognitive capacity of the human brain, and physiocentrists demand that the physics of the brain be mimicked, thinkers of biblical persuasion align squarely with the biocentrists, maintaining that only biology, based on a closed-circulatory system, can support consciousness.

Conclusion

In conclusion, consciousness is the capacity of the mind, or soul, that allows a being to experience the world, not merely symbolically but semantically, not merely functionally but phenomenologically. Such experience is found in animals (with a closed circulatory system), limited to first-order phenomenal consciousness, which allows for the immediate experience of pain and pleasure. In contradistinction, human beings have the capacity for higher-orders of phenomenal consciousness, which allows not only for immediate experience but for reflective consideration of experience. And perhaps most importantly, it is the capacity for higher-orders of phenomenal consciousness that allows for the cognitive and emotional development that separates human beings from all others.

It is these capacities that the machine consciousness community seeks to create within a “machine,” for it is these capacities that will enable our creations to truly fulfill our needs.⁴⁹ Some believe artificial consciousness can be achieved computationally, others, that it must account for biophysics, others still, that it must be biologically based. In any case, if we do somehow manage to imbue a machine with sentience, I believe it is trivial to claim that an entity – e.g., a robot – having the combination of phenomenal consciousness along with highly advanced cognitive ability, as expressed in sophisticated (but not necessarily *spoken*) language skills, would rather quickly, if not immediately,

Mallatt 2016). In addition, the “blood” dividing line does not discount even the theory of panpsychism (i.e., everything has some level of consciousness) but simply states that without “blood” any claimed consciousness would be too minimal to be of consequence.

⁴⁹ On the morality of designing conscious beings to serve, see Ch. 5 “Eudemonia of a Machine.”

achieve second-order phenomenal consciousness.⁵⁰ That is, if one were to build a robot with supercomputing abilities (e.g., Alpha Go Zero), it would have, quite prosaically, second-order functional consciousness (Bringsjord 2010: 293-4). If, somehow, phenomenal consciousness “emerged,” the robot would not long remain, if at all, on the level of first-order phenomenal consciousness but would evince second-order phenomenal consciousness. For, if it already has the working infrastructure to support second-order thinking and it now attains sentience – i.e., the ability to experience pain and pleasure – it will by definition be able to think about those pains and pleasures.

Such an entity will, in short, be able to think about thinking in a subjectively experiential way; and we will have, thus, created a humanoid – its distinction from *homo sapiens* being only its substrate (i.e., its physical composition), if that. Accordingly, the very attempt to marry sophisticated computing with phenomenal consciousness will engender the ethical quandary: should we make them? The answer to that is the subject of ch. 6 “Let Us Make Man in Our Image.”

⁵⁰ Note that Grau (2011: 461-2) entertains the possibility of robots with varying low-level computational capacities that remain merely sentient and discusses their moral status accordingly.

Ch. 4:

Article #4 - Polemics on Perfection

Maimonides' Last Law on Slaves Resolves the Debate

Introduction

Maimonides' great philosophical work, *The Guide for the Perplexed*, while laying down the foundations of Jewish philosophy in an effort to ameliorate the perplexity of philosophical seekers, has given rise, perhaps unsurprisingly, to even more philosophical perplexity. This can be seen most prominently in the polemics surrounding the *summum bonum* – i.e., what exactly does human perfection entail? For, though there is no question that Maimonides held intellectual achievement to be of ultimate import, a significant point of contention arises over human perfection subsequent to intellectual achievement. This issue has a number of aspects to it: What is to be the human activity resultant from intellection? If this human activity is found to be moral action, does it entail a specific moral approach? And finally, is this human activity the *summum bonum*, or simply an overflow or by-product of intellectual perfection?

While modern philosophers have spared no ink in trying to resolve these questions, it is my thesis that resolution can be found – worlds apart from the ivory towers where philosophers spill their ink – in the private quarters of a Canaanite slave. In his last Law on Slaves (*Hilchot Avadim* 9:8), Maimonides, it will be shown, encapsulates his entire program, not only for moral development but, indeed, for human perfection itself. In so doing, nuances in his philosophy emerge from his halakhic formulation that come to resolve the polemics that arise in reading his philosophy solely from his philosophical writings.

The idea to seek a resolution to the philosophical polemics in his halakhic writings, came to me during my research for ch. 1, “Finding Virtue in a Law on Slaves.” There, while analyzing Maimonides' last Law on Slaves, I encountered numerous opinions regarding significant aspects of his philosophy. This motivated me to write the present chapter, which aims to employ the main conclusions from that prior chapter to address the great polemics surrounding Maimonides' views on human perfection. Accordingly, I provide here only a summary of that analysis before treating, what I call, the polemics on perfection. Interested readers are encouraged to review the original chapter and return here to the below section entitled: “Polemics on Perfection.”

That said, let us begin with a summary of the last Law on Slaves (sec. “Hilchot Avadim (9:8) – The Text”), followed by a review of Maimonides’ program for perfection that underlies the text (sec. “Hilchot Avadim (9:8) – The Program”). With that introduction behind us, we then turn to the “Polemics on Perfection.”

Hilchot Avadim (9:8) – The Text

It is permissible to work a heathen slave relentlessly.

The text opens with the strict letter of law.¹ The law, however, is seen as a starting point, a floor and not a ceiling, to use the phrase of Rav Soloveitchik. Accordingly, Maimonides starts with the legal “floor” only to show that we should – and must – rise far above it. To this end, he constructs an argument structured in six statements. The first statement (1) is the hypothesis that claims one is to treat his slave with piety (*bemidat basidut*), beyond the letter of the law, and with justice (*betzedek*), according to the letter of the law yet charitably. Maimonides then brings four statements (2-5), which expand on and support these assertions, finally closing with a statement (6) that rounds out the claims, each supported by biblical verse.

The argument is built in chiasmic form:

- (1) Hypothesis Claimed
- (2) Support for Piety
- (3) Support for Justice
- (4) Support for Justice
- (5) Support for Piety
- (6) Hypothesis Conclusion

Let us now analyze the text according to this outline.

¹ On the changing attitudes to slavery within Jewish thought, see, e.g., Shmalo (2012).

(1) Hypothesis Claimed

Though this is the law, the quality of piety (midat hassidut) and the ways of wisdom (darkei hochma) demand of a human being to be compassionate (rachaman) and pursue justice (tzedek), and not make heavy his yoke on his slave nor distress him.

Maimonides, here, raises us off the floor of the law, outlining his thesis that calls for piety (*midat hassidut*), and justice (*ways of wisdom*). Piety and Justice correspond to the two moral approaches found in Maimonides' Laws of Moral Character (*Hilchot Deot*) wherein he speaks of the ways of the "pious" (1:5) as opposed to the "ways of the wise" (1:4). He thus gives voice to a dual-moral approach, wherein both moral approaches are not only essential but – together – form normative obligation.

(2) Support for Piety

He should give him to eat and drink of every food and drink. The sages of old had the practice of sharing with the slave every dish they ate. And they would provide food for their animals and slaves before partaking of their own meals. As it is said, "As the eyes of slaves follow their master's hand, as the eyes of a slave-girl follow the hand of her mistress, [so our eyes are toward the Lord our God, awaiting His favor]."

Maimonides makes his claim for compassionate, and clearly supererogatory, action by prescribing specific daily acts of kindness based on living examples; for "ethical conduct ultimately presupposes concrete exemplars" (Wurzburger 2008: 30).

(3) Support for Justice

Nor should a master disgrace his slave, neither physically nor verbally; the biblical law gave them to servitude, not to disgrace. And one should not treat him with constant screaming and anger, but rather speak with him calmly and listen to his complaints. This is explicitly stated with regard to the positive paths of Job for which he was praised: "Have I ever shunned justice for my slaves, man or maid, when they quarreled with me... Did not He who made me in my mother's belly make him? Did not One form us both in the womb?" (Job 31:13,15).

Maimonides makes his claim for treating the slave justly by noting that such is “biblical law,” as well as found explicitly in the Bible, in the example of Job who ever abides by justice in his treatment of slaves.

(4) Support for Justice

Cruelty and effrontery are not frequent except with the heathen who worship idols. The progeny of our father Abraham, however, the people of Israel – upon whom God bestowed the goodness of the law (Torah), commanding them to observe “just statutes and judgments” (Deut. 4:8) – are compassionate to all.

Maimonides here argues that unjust treatment of a slave based on “accepted practice” among the nations of the world finds no justification in biblical law. This is not, it should be noted, a parochial diatribe against non-Jews,² but rather part and parcel of Maimonides’ argument for just relations with one’s slave.

(5) Support for Piety

Accordingly, regarding the divine attributes, which He has commanded us to imitate, the psalmist says: “His tender mercies (rachamav) are over all His works” (Ps. 145:9).

Maimonides here brings the obligation to imitate God’s virtues (*imitatio Dei*), chief among them being mercy/compassion – precisely the quality he has been demanding be employed in one’s interrelationship with one’s slave. It is a demand to act beyond the strict letter of the law – a demand to act with piety (*bemidat hassidut*).

(6) Hypothesis Conclusion

Whoever is merciful will receive mercy, as it is written: “He will allow thee to be merciful and show mercy unto thee and multiply thee” (Deut. 13:18).

² Worthy of note is the great esteem in which Maimonides holds non-Jewish thinkers, frequently quoting the ideas of Aristotle, Al Farabi, Galen, etc.

Maimonides concludes with the notion of “measure for measure” – i.e., how you act towards others, so will God act towards you.

These idealistic demands for pious and just behavior are clearly not limited to one’s slave; for if moral equity is to be pursued in the interrelationship with one’s slave – i.e., the “other” at the bottom of the social ladder – then all the more so is it to be pursued with one’s contemporaries, let alone with one’s superiors. And so Maimonides writes: “This [kindness] we owe to the lowest among men, to the slave; how much more, then, must we do our duty to the freeborn, when they seek our assistance?” (*Guide* 3:39).

Hilchot Avadim (9:8) – The Program

Maimonides, however, is not satisfied in merely demanding moral conduct but rather, within this very law, he provides a program to achieve it. It is a program that integrates justice and piety. It is a program that entails the development of appropriate behavior as well as the development of appropriate dispositions. It is a program that, ultimately, aspires to bring one to the height of human perfection.

To appreciate Maimonides’ intent, it is crucial to recognize that the chiasmic form employed here has didactic significance beyond its mere aesthetic appeal. It visually articulates the notion that the path to perfection begins with piety and ends with piety, justice being the bridge (that is never burned).

(2) Piety as Initiation

The path to moral and human perfection begins with the epiphany that such a *telos* exists, and that it exists in God Himself. It is this that leads one to set out on the path by adopting a moral law. So explains R. Soloveitchik: “Whenever a person beholds God, an inner catharsis compelling a complete change of one’s axiological hierarchy must occur” (2017: 182). But at this early stage on the path, “imitating is the only way,” for imitating (*imitatio Dei*) is “the foundation of morality” (Soloveitchik 1993: 26). The imitation at this initial stage, however, can only be through imitating human role models, imitating the

sages (*imitatio sophos*) – precisely as found in Maimonides’ first appeal to piety (2) where he enjoins us to imitate “*the sages of old who had the practice of ...*”

(3) Justice as Preparation

Having observed piety from the pious, and thus gaining an appreciation for the path, one is ready to accept the “ways of the wise,” the balanced path of the law. At this stage, then, one embarks on the journey of practicing the Law. Accordingly, Maimonides’ first appeal to justice (3) refers explicitly to “biblical law” as an appeal to moral practice.

This stage of practicing of the Law, it must be noted, is one done without great understanding of what underpins the law and its practical demands – it is the “we will do” (*naaseh*) practice that precedes the “we will understand” (*nishma*) practice (Ex. 24:7).

Intellection as Inflection

This brings us to the inflection point of the chiasmic structure. Having engaged a bit in the “ways of the pious” and a bit in the “ways of the wise,” one is ready to embark further in these ways, in reverse order – i.e., practicing more ways of the law (i.e., justice), and consequently, realizing more of how and when to go beyond the law (i.e., piety). But this new moral practice is to be informed by an understanding resultant from the process of “intellection” (i.e., study). The program for moral development toward human perfection is such that one first accepts and performs moral laws as given (a.k.a., pre-theoretic morality), then studies them in depth to reach for an understanding of their Source, whereupon, having been transformed in intellectual achievement, one performs them with understanding (a.k.a., post-theoretic morality).

The phase of intellection, then, is what serves as the unwritten inflection point of the chiasmus, transforming the moral practitioner to ready him/her for a new phase of personal development.

(4-5) Justice & Piety as Dual-moral Approach

This new phase is that of the intellectually enhanced moral practice of justice and piety. Maimonides thus reiterates the otherwise redundant demands to follow the ways of justice (4) and the ways of piety (5). However, whereas the pre-theoretic stage consisted of piety (as initiation) leading to justice (as preparation), now, following the stage of intellection, the order is reversed: first justice then piety. The reason, I suggest, is that now one must begin by performing the law with understanding before one can go beyond it with understanding.

The path to perfection, as such, might be depicted schematically as follows:

pre-theoretic morality > intellection > post-theoretic morality

However, surely it would be the height of hubris (or naivete) to conclude that moral perfection could be achieved so simply – i.e., that the path to moral perfection requires a mere single pass. Accordingly, many have noted that Maimonides' path to perfection is an iterative one:

It must be realized that in the Maimonidean system, “thou shalt walk in his ways” represents a *continuous* challenge, beginning with the attempt to cultivate moral virtues through moral conduct and pointing to the *ever higher* dimensions of *imitatio Dei* which can be engendered only by intellectual perfection (Wurzburger 2008: 99, *emphasis added*).

That Maimonides' program for human perfection is an iterative one is, most remarkably, reflected in the chiasmic structure of the law on slaves:

piety > justice > intellection > justice > piety

That is, upon reaching a level of piety after intellection, one goes back to the beginning – attempting further gains in piety through *imitatio sophos*, further commitment to the performance of the law, further investigation of that which lies behind the law – all to

cultivate higher levels of justice and piety. The cycle continues with the ultimate – and ultimately unattainable – goal of imitating God’s ways in all their glory.

Maimonides thus encapsulates his program for moral perfection within the chiasmic structure of the last law on slaves – calling for the dual-moral approach to be developed through a lifetime of striving to imitate God.

(6) Purpose Achieved

The last law on slaves concludes: *Whoever is merciful will receive mercy, as it is written: “He will allow thee to be merciful and show mercy unto thee and multiply thee” (Deut. 13:18)*. Now, while this “measure-for-measure” appeal can be understood at face value (i.e., self-interest), we can also discern in it the culmination of the program for perfection. Given that God created the world to reach perfection, it is His desire that man reach that perfection. The supporting verse – *“He will allow thee to be merciful and show mercy unto thee”* – thus reads: God endows man with the trait of mercy – the expression of which is man’s ultimate purpose – so that He bestow His mercy – the expression of which is God’s ultimate purpose. This last statement (6), then, is a most appropriate conclusion to the hypothesis that one must apply one’s complete moral presence, via the dual-moral approach and following the iterative program through intellection, to realize the ultimate purpose of being a wise-merciful individual.

Polemics on Perfection

With this new perspective on Maimonides’ philosophy, gained through the prism of his last law on slaves, we can now address the polemics surrounding Maimonides’ path to perfection. As stated at the outset, there are three questions we seek to resolve: What is to be the human activity resultant from intellection? If this human activity is found to be moral action, does it entail a specific moral approach? And finally, is this human activity the *summum bonum*, or simply an overflow or by-product of intellectual perfection?

Human Activity

What is to be the human activity resultant from intellection?”

The “standard” approach to Maimonides’ program for perfection holds intellectual achievement to be the *summum bonum*.³ That said, a great many modern readers of Maimonides understand intellection as leading, in *imitatio Dei*,⁴ to new human activity. This human activity, as summarized by Menachem Kellner (1990: 8, 53; 2009: 40), is taken to be one of three things: political activity – i.e., the ethical governing of state;⁵ moral activity – i.e., the ethical behavior of the individual;⁶ halakhic activity – the fulfillment of the commandments as embodied in the Jewish legal corpus.⁷

While this tripartite division of opinions does obtain from Kellner’s sources, it is possible to show that in fact they all coalesce to one notion, one human activity: the ethical. Regarding the debate over whether moral activity or political activity is the *summum bonum*, this can be defused by noting that both are really achievements of the ethical – one applying to non-leaders, the other to leaders.⁸ That is, living ethically *simpliciter* is for non-leaders what governing ethically is for leaders, both find their *summum bonum* in living ethically. Regarding the debate over whether moral activity or halakhic activity is the *summum bonum*, this too can be defused, if with but a bit more effort.

³ See Kellner (1990: 67 n. 20) for a list of those who maintain the standard position.

⁴ As per *Guide* 1:54, 3:54.

⁵ Leo Straus, Lawrence Berman, Sholmo Pines (in Kellner 1990: 8, 41).

⁶ Herman Cohen, Julius Guttman, Steve Schwarzschild (in Kellner 1990: 8, 41). Similarly Friedberg (2019), Wurzbürger (2008: 167), Wurzbürger (1994: 65), Shatz (2005: 184-186).

⁷ Kellner (1990: 11) and, according to him, Isadore Twersky and David Hartman (in Kellner 1990: 10, 45, 47). While Twersky does note “the law” (i.e., halakha) is “cause and consequence of ... of the cognitive goal” (1980: 364), it seems to me that Twersky holds ethical achievement over halachic achievement, the latter simply being a means to former: “Law is the crowning feature of Judaism, for the commandments are the indispensable *means* to ethical-intellectual perfection, which is the goal of man” (Twersky 1980: 226, emphasis added). So too Hartman (1986: 204), while making strong arguments for the halachic life, explains that ultimately, it is morality, as reflected in God’s *besed*, that is to guide man’s post-theoretic life and not simply obedience to laws. See also, Soloveitchik, who sees halakha as the ethical aspect of human perfection – but he clearly uses broad stroke terminology in equating halakha with ethics (2016: 123).

⁸ See, e.g., Wurzbürger (2008: 99), Friedberg (2019: 19). See also, Kellner (1990: 51).

To begin, Kellner argues that “morality” cannot be the *summum bonum* for Maimonides because Maimonides holds “morality” as the means to intellectual achievement and *imitatio Dei* as the end. Consequently, Kellner believes that Maimonides has something other than “morality” in mind for *imitatio Dei* (1990: 83, fn. 2), to wit, “halakha” (1990: 47, 63). Kellner does admit, however, that were one to equate “morality” with “halakha,” the disagreement would be entirely “semantic” (*ibid.*).

Now, while there is room to agree to this “semantic” reduction of the disagreement – i.e., to equate halacha and morality – we would be doing a disservice to both halacha and morality. More importantly, we would be doing a disservice to the very human perfection that we have been seeking to define. Halacha is a system of laws that defines an individual’s actions.⁹ Morality is a system of values that defines good and evil, right and wrong.¹⁰ Morality informs halacha, and halacha reifies morality. In the ideal, halacha both cultivates and inspires moral behavior. That is, halacha cultivates “lower” moral behavior, what we have been calling “justice,” and it inspires “higher” moral behavior, what we have been calling “piety.”¹¹ What characterizes halacha and morality is not identity but symbiosis. Ultimately, then, the perfection that *imitatio Dei* holds out as its telos is moral practice animated by halacha.¹²

⁹ See, e.g., Berkovits (1983: 1).

¹⁰ For a discussion of good versus right see Spero (2016).

¹¹ The terms “lower” and “higher” are used as such by Wurzbarger (1994: 82-85), Wurzbarger (2008: 98), and Soloveitchik following Hermann Cohen (in Soloveitchik 2016: 30).

¹² I hope it will be appreciated that I have tried to choose my words very carefully in an attempt to tip-toe through the veritable minefield of explosive issues inherent in this discussion. To name but a few: (1) There is the issue of halachic demands based on biblical commands that clearly do not align with our moral sensitivities – slavery being one discussed herein. For a treatment on the subject of “developing morality” see, e.g., Korn (2002), Lichtenstein (2002: 16-17), Rabinovitch (2003: esp. 9), Lamm (2007). These authors explain that the Bible and halakha allow for moral development without slipping down the slope of abandoning the Bible and halakha. (2) There is the issue of an ethic outside of halakha – i.e., are there moral demands outside of halakha, or can halakha (as a set of laws) presume to include all moral values? For approaches to this issue, see, e.g., Wurzbarger (1994), Carmy (1996), Lichtenstein (2004a), Wurzbarger (2008: 21-87), Bleich (2013), Goldberg (2014), Soloveitchik (2017: 181-191). Many here understand the halachic system to contain the latitude to apply moral intuition [see, e.g., Wurzbarger (1994), Berkovits (2002: 67), Blau (2002), Bleich (2013)], or that the Oral Torah’s Aggadic material supplies the values within the system [see Bleich (2013: 141), Berkovits (2002: xviii), Bernstein (2015)]. (3) There is the issue of “*naval*

The question then becomes: What morality?

Morality: Justice and Piety

This is a question fraught with contention due to the seeming contradiction Maimonides lays out in the first chapters of his *Laws on Moral Character* – advising one to follow “the ways of wisdom” (i.e., the mean) and yet also to follow the “the ways of piety” (i.e., the extreme). Wurzburger notes the “vast literature dealing with the seeming contradictions” (1994: 82)¹³ and rejects it, claiming that the two approaches are essential, to be applied as dictated by circumstances. Wurzburger makes cogent arguments for the dual-moral approach,¹⁴ but it is the last law of slaves, untapped by Wurzburger,¹⁵ that provides compelling *practical* support for it. And this because, as explained, it is in this law that Maimonides’ calls for both justice and piety to be applied coextensively – thus advocating, *in practice*, for the dual-moral approach.

Interestingly, Raymond Weiss (1991: 156) brings this very law of slaves as a manifest example of the “unambiguous agreement between Jewish morality [i.e., piety] and philosophic ethics [i.e., justice] ... [as] both require that slaves be treated with kindness.” However, while the outcome (e.g., kindness) may be the same with respect to the passive agent (e.g., the slave), clearly the two moral approaches demand very different motivations from the active agent (e.g., the master). For surely one must exert greater

bereshut batorah,” of which the Nachmanides (Lev. 19:2) teaches that halakha is no guarantee for moral behavior. Indeed, it is for this reason that the two cannot be equated but rather it is incumbent on the practitioner of halakha to set morality as his ultimate concern; for, while halakha cultivates and inspires morality, it does not guarantee it.

¹³ Here are some of the many sources dealing with this issue: Rawidowicz (1954 in Wurzburger, 1994); Cohen (1972 in Wurzburger, 1994); Weiss and Butterworth (1975); Schwarzschild (1990); Weiss (1991); Kreisel (1992); Shatz (2005); Strauss (2013); Shapira (2018). See also Ravitsky (2014) and sources in fn. 2 therein.

¹⁴ Wurzburger (1994: esp. Ch. 5).

¹⁵ For the sake of completeness it should be noted that Wurzburger does make a cursory reference to the law but in a completely different context (1994: 51).

moral mettle to “*share his every dish with his slave*” (i.e., act with piety) than to merely “*refrain from constant screaming*” (i.e., act in justice).

Accordingly, the two approaches are not congruous nor complementary, neither are they contradictory nor in conflict; rather, they are independent, to be held – simultaneously valid – in dialectical tension. And this, as Wurzbürger explains, is because they reflect the dialectical tension that “arises out of the dialectical tension characterizing the human condition.”¹⁶ Here he refers to Soloveitchik’s assessment of the human being: on the one hand, a majestic being, driven to conquer the world in an effort bring dignity to human existence; and on the other hand, a humble being, drawn to the transcendental in the awareness of his existential absurdity. It is to this dialectic nature of the human being that the two moral approaches respond – the ethics of the wise (i.e., justice) corresponding to the majestic being, the ethics of the pious (i.e., piety), to the humble being. And it is due to this dialectic nature of human being that the two moral approaches are imperative.¹⁷ When to apply which is the challenge inherent in the dialectical human condition and for which we must, ultimately, rely on “our admittedly exceedingly fallible intuitions.”¹⁸

Morality: Justice then Piety

Despite the claims for the dual-moral approach, many contend that ultimately (i.e., following the stage of intellection), post-theoretic morality is exclusively that of piety. Here is how Simon Rawidowicz (1974: 286) puts it:

¹⁶ Wurzbürger, 1994: 84; see also Wurzbürger, 2008: 168-169.

¹⁷ Noteworthy is that, while Soloveitchik (1978a: 26) describes the dialectic in man and morality, it is Wurzbürger (2008: 169) who equates, quite justifiably in my opinion, “the ethics of the wise” with majestic man and “the ethics of the pious” with humble man, calling them “the ethics of majesty” and “the ethics of humility,” respectively. Interestingly, the integration of act morality (i.e., justice) and agent morality (i.e., piety) espoused here aligns with such integration as found in Kant’s moral approach (See Loudon 1986).

¹⁸ Wurzbürger (2008: 170). Intuitions are called for, e.g., when the rules run out (Wurzbürger 1994: 32), when one should go beyond the law (ibid.: 34), and in dilemmas where the rules conflict (ibid.: 33). It is important to note that the intuitions here are those “formed within the matrix of Torah teachings” (ibid.: 28; 2008: 26, 53-54, 71-72) and can never override “divine imperatives” (1994: 29). For more on this see ibid.: Ch. 2.

The man who is wholly absorbed in his efforts to see God ... this man does not know any ‘middle way’ between the extremes, any ‘golden mean.’ It is the absolutely extreme man [i.e., the pious man]. Only this man is man to Maimonides.

And so Steve Schwarzschild (1990: 143, 145) explains:

[In the] initial stage, the religious *mitzvoth* and the virtues of philosophical ethics are determined by the rule of the mean. Once this initial stage has been passed, however, another and higher stage is entered – that of the intellectual union with God. There is nothing moderate about the ethics of this *theoria* in Maimonides... Here extremism, radicalism [i.e., piety] rules.

Similarly, Alexander Altmann (1972: 24) and Daniel Frank (1985: 491, fn. 33).¹⁹

According to these thinkers, then, the Maimonidean program has one seeking moral perfection according to the ethics of the mean (i.e., justice), followed by intellectual perfection via theoretical speculation, followed by moral perfection via *imitatio Dei* according to the ethics of the pious (i.e., piety). Such a program, however, raises three difficulties: (a) it discounts intellection as having any effect on “mean” moral behavior, (b) it discounts the practice of the law as having any effect on pious moral behavior; and (c) it discounts the possibility of piety before intellection. I find all of these conclusions untenable, and consequently, the “program to piety-only” to be specious, as I now explain.

¹⁹ For the sake of completeness, Shatz (2005: 187-188) maintains the difference between pre- and post-theoretic resides in psychology – i.e., true intellection making one feelingless. Similarly, Davidson (1987: 65-68) and Friedberg (2019: 19), the latter who uses the term “dispassionate.” Against this see Ravitsky (2014: 44-45) and Kellner (2009: Ch. 4), the latter who rails against this notion.

Difficulty (a): “Intellection has no effect on ‘mean’ moral behavior”

In response to the claim that “intellection has no effect on ‘mean’ moral behavior,” I contend that intellection is not simply the acquisition of knowledge (whether theoretical or practical), but a transformative experience.²⁰ And while one could argue, as Schwarzschild does, that the transformation will lead from middling to radicalization (i.e., from justice to piety, exclusively), such a conclusion is debatable. Indeed, if humans are – by design – the dialectal beings that Soloveitchik maintains they are, then the dialectical moral approaches of justice and piety will continue to address such beings even after the transformative encounter, but in a deeper or, to use Schwarzschild’s terminology, more radical way. Intellection, then, will affect one’s practice of the law as commanded (i.e., justice) just as much as it effects one’s practice of going beyond the law (i.e., piety). And indeed, many note this phenomenon explicitly:

“... only contemplation and meditation—sustained reflection on the significance and objectives of every commandment—will safeguard against perfunctory performance” (Twersky 1980: 395; also 362-363).

“... the law also requires a person to engage in [theoretical] speculation so that he or she can turn accepted beliefs into sincerely held convictions” (Bedzow 2017: 109).

Similar sentiments are expressed by Soloveitchik (2017: 182, 185), Wurzbarger (1994: 83), Hartman (1986: Ch. 1) and Kellner (1990: 35, 39).

Difficulty (b): “Practice of the law has no effect on pious moral behavior”

In response to the claim that “practice of the law has no effect on pious moral behavior,” let us begin with Isadore Twersky’s explanation that:

²⁰ Kellner (1990: 35, 39, 41; 2009: 71-72) emphasizes that the achievement of the intellect does not solely entail gaining knowledge, but rather that it “transforms” the individual. Similarly, Soloveitchik (2017: 182).

... all the laws are a springboard for the highest morality and perfection which emanate slowly and steadily from them. Just as one embraces reality in order to transcend it, one adheres to the law in order that it may enhance one's perception of the good and the true and induce behavior which transcends the letter of the law. ... There is ... a continuum from clearly prescribed legislation to open-ended supererogatory performance...²¹

Combining this idea with the response to the previous claim, we can say that, as one's perfection in the realm of the law (i.e., justice) is improved by intellection, so too, will it provide a "springboard" toward a greater ability to go beyond the law (i.e., piety). Ira Bedzow (2017: 293) says as much, explaining that it is the law itself that provides the understanding of how to go beyond it:

The acceptance of principles and concepts embedded within the law ... allows for the manifestation of the ethical within the legal. This understanding of *lifnim mishurat badin* supports the view that the law can shape a person's beliefs, since, even when he or she acts "supererogatorily," it is the law that provides the beliefs which motivate such action.

Difficulty (c): "Piety before intellection is impossible"

In response to the claim that maintains "piety before intellection is impossible," let us begin with Hartman, who writes that: "Anyone, at any given time, may perform an action that is beyond the strict requirement of Halachah. Yet, to the *basid*, such acts are not isolated moments of religious fervor; they derive from the nature of his intellectual love of God" (1986: 206). In addition, while it is quite reasonable to claim that the selflessness of the pious moral act – in *imitatio Dei* – can only come through proper intellectual apprehension,²² that does not mean that one cannot learn to perform pious acts in a less ideal form. Indeed, such learning, I suggest, can come through *imitatio*

²¹ Twersky (1980: 428). Similarly, Wurzbürger (1994: 27-28, 37); Bedzow (2017: 11).

²² Indeed, even those who do not maintain that post-theoretic moral behavior consists exclusively of piety agree that intellection is necessary to perform piety selflessly – see, e.g., Hartman (1986: 205), Wurzbürger (2008: 98; also 1994: 78-79).

sophos. Twersky (1980: 510, fn. 390) calls it the “follow-the-leader system,” and explains it as follows:

“The believer, having faith in those from whom he has received the tradition, not however knowing the truth by the exercise of his own intellect and understanding, ... is like a blind man led by one that can see’ (*Hovot HaLevavot* 1:2). By frequent repetition of inherently desirable and edifying acts, every person, even a blind one, may achieve a degree of ethical perfection ([as per Maimonides’] *Eight Chapters*, Ch. IV).

So, while he does not say explicitly that the “edifying acts” are those of piety as opposed to justice, it stands to reason that one who is simply mimicking behavior, “like a blind man,” will do so indiscriminately and mimic acts of justice as much as acts of piety.

Now, referring back to the program for perfection as outlined in the last law of slaves (statements 2-5), my arguments against the above claims (a, b, c) find their support as follows. Piety in the ideal results from intellection (5), but can also be performed by mimicking virtuous people – *imitatio sophos* (2). Justice is enacted in its instrumental form – i.e., prior to intellection (3), but nevertheless benefits from an understanding of the values underpinning the laws – an appreciation gained after “God influenced (*bishpia*)²³ the law upon them” (4). And it is this ideal justice (4) that informs the ideal piety (5).

Having refuted the implications inherent in the “program to piety-only” and so undermined this single-moral approach,²⁴ we can now better appreciate the dual-moral

²³ The term *bishpia* is telling, as it is used by Maimonides to refer to the special “overflow” of knowledge that God grants via intellection (see, e.g., *Guide* 2:12; 2:36-37; 3:51-52). Barry Kogan explains Maimonides’ intent on the term such that, “the overflow of forms from the Active Intelligence into us occurs throughout the learning process” (1989: 127).

²⁴ Noteworthy here the claim that Frank (1985: 491-492) makes regarding the “ways of the wise.” He equates “wisdom” with “moral virtues” such that when Maimonides quotes Jeremiah “let not the wise man glory in his wisdom” he holds that the “ways of the wise” are not part of the ultimate human perfection to glory in. I would argue that the “ways of the wise” as pre-theoretic are not to be gloried in, and neither are they alone as post-theoretic to be gloried in. However, that does not mean that the “ways of the wise” – as part and parcel of a wholistic moral approach – are not to be gloried in.

approach. For, beyond its ability to better describe the process of human moral development, it also better aligns with man's moral essence. That is, if, as claimed above, humans are dialectal beings by design – i.e., the dialectical essence of human nature is ever present – then the dual-moral approach of justice and piety will be ever relevant, whether at the pre- or post-theoretical stage. Furthermore, as the last law of slaves alludes, the process of development toward perfection is an ongoing one, there is no true pre- or post-theoretical stage, but only ever greater stages of awareness wherein every post-theoretical stage becomes, in turn, a more educated pre-theoretical stage.

On Perfections and Overflows

And this brings us to what is perhaps the greatest point of contention among readers of Maimonides' philosophy: human perfection – i.e., what, according to Maimonides, is the *summum bonum*? As mentioned, the “standard” reading of Maimonides has led to the conclusion that it is “intellectual perfection,” whereas the modern reading has accounted for Maimonides' references to the human activity that follows intellectual perfection. Here it is instructive to see how modern readers attempt to straddle the tension between intellectual perfection and moral perfection.

Of the many modern readers, Kellner appears to be the most emphatically in favor of moral (halakhic) perfection over intellectual perfection. He starts his book on the subject, appropriately entitled, *Maimonides On Human Perfection*, by explaining: “The most ‘perfect perfection’ is that the individual who not only knows God, but understands the ways of God's providence and strives to imitate those ways in his or her actions. The *vita activa* appears to be the ideal, and not the *vita contemplativa*” (Kellner 1990: 8). He backs this up later, writing: “beyond intellectual perfection there is an additional perfection ... to practice loving-kindness, righteousness, and judgment, thereby imitating God” (ibid.: 37). Further on, however, he writes:

“[I]ntellectual perfection is indeed our *highest* perfection. But it is not, for all that, the imitation of God. That is an activity which we can aspire to perform in its fullest sense only after intellectual perfection has been achieved. ... it may very well be that Maimonides viewed intellectual perfection as the highest form of perfection open to gentiles (who know not the Torah), as opposed to the *higher* (if

not more greatly rewarded) practical perfection open to Jews who know the Torah and who can therefore imitate God properly” (ibid.: 57, emphasis added).

Here we note a hint of ambiguity in that he starts by calling “intellectual perfection” the “highest perfection,” only to write that there is a “higher” perfection. The book concludes with even more ambiguity: “I suggest that the *highest perfection* which human beings qua human beings can aspire to is *intellectual perfection*. But that is not the final obligation ... [which] is to imitate God as we are instructed to by the Torah” (ibid.: 64, emphasis added). So again the “highest perfection” is “intellectual perfection,” imitation of God being a “final obligation” and not a perfection.

From Kellner’s words we begin to realize the great difficulty in felling “intellectual perfection” from its lofty position found throughout Maimonides’ words. Schwarzschild attacks the dilemma by bringing moral perfection to the level of intellectual perfection, writing:

Maimonides’ exegesis is clear: humanity’s purpose is to ‘know’ God, but the God who is to be known is knowable only insofar as He practices grace, justice, and righteousness in the world, and to know Him is synonymous with imitating these practices of His in the world: ‘for these do I desire’” (Schwarzschild 1990: 144).

Schwarzschild, in what is surely a move of great ingenuity, thus resolves the dilemma by simply equating moral perfection with intellectual perfection – i.e., not that they are two separate things on the same level, but rather that they are “synonymous.”

This equation is too much for Frank who argues that Schwarzschild has gone too far, essentially making “the *summum bonum* for man ... *only* practical” (1985: 495). He quotes Altmann, with whom he aligns, who writes: “*Imitatio Dei* is, therefore, but the practical *consequence* of the intellectual love of God and is part and parcel of the ultimate perfection” (Altmann 1972: 24, emphasis added by Frank). On this Frank stresses that the *summum bonum* is unreservedly intellectual perfection, with moral activity its cardinal consequence and integral aspect:

“To know God is a theoretical, not a practical, activity, though, to be sure, what one learns is that God is an *actor*, a doer. Imitation of His ways is the (practical) *consequence* of what one has learned. Indeed, as Altmann urges, it is (only) part of the *summum bonum*, though the litmus test of the (purported) divine encounter” (Frank 1985: 495).²⁵

But even this balancing act is too much for readers like Lenn Goodman, who, while very much valuing moral activity, is unwilling to include it in any way with the *summum bonum* of intellectual perfection:

“Awareness of God’s perfection is the ultimate object of the human quest. But that awareness does not compete with other human goals. Intellectual consumption spills over into holy acts of guidance and generosity” (Goodman 2009: 461).

Clearly, Goodman does not deny the existence or exigency of lofty human activity – characterizing them as “goals” enhanced by intellectual achievement. Yet, for all that, the *summum bonum* is wholly and solely “intellectual perfection,” unencumbered of any other human endeavor. Goodman is, of course, not alone, as many others see the post-theoretic moral activity that Maimonides calls for (*Guide* 3:54) as an “overflow” of the *summum bonum* of intellectual perfection. David Shatz explains it as follows:

... Maimonides distinguishes between “the perfection of man that may be truly gloried in” and “the way of life of such an individual” who has achieved that perfection. The way of life is not the perfection itself; the way of life is rather a consequence of – an emanation or overflow from – the perfection. By achieving intellectual perfection, the perfect individual engages in a life of *imitatio Dei* with respect to the Deity’s actions.²⁶

²⁵ I will not argue here with Frank’s reading of Altmann, but only note that it is he who adds the diminutive “only,” such that Altmann could be read to hold that moral achievement is on par, symbiotic, with intellectual achievement *ala* Twersky (see below).

²⁶ Shatz (2005: 186). Similarly, Hartman (1985: 26), Harvey (1990: 9), Davidson (1992 in Shatz, 2005), Kreisel (1999 in Shatz, 2005), Nadler (2021: 269).

Lastly, while no one could propose that Maimonides held intellectual perfection a mere steppingstone to morality, Albert Friedberg (2019: 23-28), based on Stern (2013) and Pines (1979), offers the “skeptical” suggestion that Maimonides actually abandoned the idea of intellectual perfection, it being humanly unattainable.²⁷ Accordingly, moral perfection is, indeed, the *summum bonum*, what Friedberg calls, “the *basid*.”

In summary, we have seen quite a gamut of opinions seeking to resolve the tension between intellectual achievement and moral achievement. Kellner sees moral (halakhic) perfection as a “greater perfection” than the “highest perfection” of intellectual achievement. Schwarzschild sees moral perfection as intellectual perfection, each being a different side of the same coin. Altmann and Frank see moral perfection as a consequence and aspect of the true perfection of intellectual achievement. Goodman and many others see intellectual perfection as the immutable *summum bonum* which influences moral behavior through the overflow of its achievement. And lastly, Friedberg offers the skeptical opinion that intellectual achievement is unattainable, the true and only *summum bonum* being that of pious moral behavior.

That said, there is one last opinion – that of Twersky²⁸ – which can be seen as a nuanced version of Schwarzschild, but one that avoids Frank’s attack, i.e., that Schwarzschild made “the *summum bonum* for man ... *only* practical.” Twersky notes that Maimonides views the relationship between law and philosophy, practice and intellection as a “dialectical-symbiotic relationship” (1980: 361), explaining that: “This circular-reciprocal relationship [is] a cornerstone of the Maimonidean system. ... Law [i.e., religious-ethical behavior] is both cause and consequence, catalyst and crystallization, of the cognitive [i.e., intellectual] goal” (ibid.: 359-364). And to remove all doubt as to whether it is the

²⁷ Noteworthy is that while Friedberg, Pines and Stern all hold intellectual perfection to be unattainable, Stern maintains that it nevertheless remains the ultimate goal (see Friedberg 2019: fn. 50, Nadler 2021: 269, fn. 3 and 281, fn. 28). For a counter-argument to this position as expressed by Pines, see Motzkin (2012); see also, Friedberg (2019: fn. 49).

²⁸ As will be explained, Twersky sees in Maimonides’ writings a symbiotic relationship between the intellectual and ethical. For the sake of completeness, it should be mentioned that Soloveitchik (2016: 192-194, esp. 238) also proposes such a reading, though the *summum bonum* is God Himself, achieved through the two perfections (i.e., intellectual and ethical).

ethical or the intellectual that is the *summum bonum*, he writes: “Law is the crowning feature of Judaism, for the commandments are the indispensable means to *ethical-intellectual perfection, which is the goal of man.*” (Twersky 1980: 226, emphasis added). The two together – ethical perfection and intellectual perfection – are the ultimate and inseparable goal of the individual.

Twersky’s position is seconded by Barry Kogan (1989: 135) but with an important emphasis on the value of moral action:

The final end is thus to be a certain kind of person. It is to be one who both apprehends God and His ways in nature and who conducts himself in ways that conform to this apprehension. Contrary to certain appearances, the final end is not to be identified either exclusively or primarily with the life of apprehension alone, not because such a life is not possible but because, as Maimonides himself points out, such a life is incomplete. It [i.e., apprehension] brings about the true human perfection, to be sure, but it belongs to its possessor alone and perfects him alone. A perfection is greater, however, when it does something more – when it overflows and perfects others. And the imitation of God clearly requires this additional measure, because God’s apprehension is so comprehensive and abundant that it overflows beyond itself and enables others to be perfected.

It is not only that the ethical and intellectual gestures are symbiotic, then, but that it is only in their symbiotic combination that one actualizes his perfection through *imitatio Dei* and, in so *doing*, becomes the “certain kind of person” that he is to become.

This symbiotic approach is most compelling for it is precisely the one that emerges from our philosophical reading of the last Law of Slaves. As explained, the last Law of Slaves asks that we develop our moral posture through acts of piety and justice as we develop our intellectual understanding of morality and its Source, in a never ending cycle wherein we aspire for greater practice and greater understanding, each influencing the other in “reciprocal influences of knowledge and action.”²⁹ Only thus, does one reach the

²⁹ Twersky (1980: 363, n. 15). See also Bedzow (2017: 126).

summum bonum, the goal of perfected individual, described so simply by Maimonides in his introduction to the Mishnah:

The goal (*tachlit*) – in this our world, and all there is in it – is the learned and moral individual.

Conclusion

To conclude, in an effort to understand the place of the intellect in man's quest for perfection, we have found that it is not intellect alone, nor even intellect as influence on practice, that is the *summum bonum*, but practical activity – as an overflow of the synergy between ethical behavior and intellectual apprehension – that is the goal. Furthermore, we have seen that the ethical behavior demanded here is not a single-moral approach, but a dual-moral approach that asks for the dialectical exercise of both justice and piety. And, while these conclusions have been espoused by others before, they are now supported by our reading of Maimonides' last Law on Slaves.

Ch. 5:

Article #5 - Eudemonia of a Machine

On Conscious Artificial Servants

Introduction

Henry Ford once said, “For most purposes, a man with a machine is better than a man without a machine.” Engineers today propose an addendum to Ford’s adage: “and a man that *is* a machine is best of all.” It is to this end – the ultimate machine – that engineers around the globe are working (e.g., Fitzgerald, Boddy, and Baum 2020). And make no mistake, this ultimate machine is no mindless automaton but rather a fully *conscious* humanoid. For indeed, as many maintain, the only way to make a machine as competent as a human is if it too has the consciousness that makes humans unique (e.g., Penrose 1991: 419; Signorelli 2018: 8; Gocke 2020; McFadden 2020: 2). So the quest is on to build a conscious machine, referred to as “the holy grail of artificial intelligence” (e.g., Bishop and Nasuto 2013: 86; Gocke 2020: 227). In consonance, Hod Lipson, director of the Creative Machines Lab at Columbia University, explains that his work toward a conscious machine “is bigger than curing cancer. If we can create a machine that will have consciousness on par with a human, this will eclipse everything else we’ve done. That machine itself can cure cancer” (Whang 2023).

So the conscious robot will be built as the ultimate machine – i.e., the ultimate servant (see, e.g., Grau 2011: 458; Hauskeller 2017: 1). But before we make this great leap that eclipses the world as we know it, we need to ask one very simple question: Should we do it? Should we be seeking to make a conscious machine? On the one hand, what could be wrong with having a super-intelligent machine “cure cancer”? Perhaps nothing, if that’s what the machine wants to dedicate its life to. But what if it isn’t interested in oncology? And what about all the other conscious machines that we will assign to less glorious tasks? “Modern robots,” explains Kevin LaGrandeur, “are chiefly meant to perform the same jobs as slaves – jobs that are dirty, dangerous, [and dull]¹– thereby freeing their owners to pursue more lofty and comfortable pursuits” (2013: 161; similarly, Bryson 2010; Mayor 2018: 152).

Applied to the mindless machines we are building today, LaGrandeur’s statement engenders no moral concerns vis-a-vis the machine. For, according to the standard

¹ The tasks destined for robots are generally referred to as “the three D’s” (see, e.g., Lin 2012, Veruggio and Abney 2012, Vallor 2016, Nørskov 2017).

approach to establishing moral status – i.e., which holds that two entities sharing the same defining ontological properties share the same moral status – today’s machines demand the same moral concern as your laptop (see, e.g., Ch. 2 “The Virtuous Servant Owner”). We are not, however, talking about the mindless machines of today but the *conscious* machines of tomorrow. We are talking about second-order phenomenal consciousness, the “stuff” that makes us essentially who we are.² Indeed, if that were not true, Hod Lipson and the other seventy-two labs around the world³ wouldn’t be working so hard to achieve it. Accordingly, as consciousness is that which makes us human, so too machines with consciousness will have to be treated, morally, as humans.

Not surprisingly, there is a consensus that the pursuit of this “holy grail of artificial intelligence” is not so holy, as it would result in the creation of a new class of enslaved beings, something clearly beyond the pale (e.g., Walker 2006a: 2; Bloom and Harris 2018; Penrose 1991: 8; Grau 2011: 458; Bryson 2010).⁴ Nevertheless many simply accept – *ex post facto* – that such machines will be built and thus attempt to define a moral framework for them (see, e.g., sources in Musial 2022). That is to be expected. What is unexpected is the defense championed by Steve Petersen who argues – *ab initio* – that there would, in fact, be nothing morally untoward *if* the machines were programmed appropriately: “Against [the] consensus I defend the permissibility of robot servitude, and in particular the controversial case of designing robots so that they *want* to serve ... human ends”

² For the sake of completeness, second-order phenomenal consciousness is that which gives humans not only the ability to have subjective experience (even most animals have that), but also allows us to reflect on that experience. Second-order phenomenal consciousness is understood to be a faculty of the “mind,” what was once referred to as the “soul.” And it is second-order phenomenal consciousness that is widely held to distinguish human beings from everything else. Indeed, even dissenters do not deny that consciousness is paramount but that (a) it is not defined well enough to be of practical use, and (b) it is not discernable beyond behavior (e.g., Gunkel 2018: 98-100). For more on this see Ch. 3 “To Make a Mind.”

³ Fitzgerald, Boddy, and Baum 2020.

⁴ For the sake of completeness, it should be noted that while the Bible makes an allowance for slavery, it is seen only as a concession to human weakness, the ideal being a world of freedom and equality (see, e.g., Shmalo 2012; also, e.g., Lichtenstein 2002, Korn 2002, Rabinovitch 2003, Lamm 2007). Importantly, R. S. R. Hirsch (Ex. 12:44) explains that the Bible only permits the ownership of existing slaves, and not the enslavement of free individuals. This position, while not without controversy (Shamlo 2012: 8), would oppose the creation of slaves, be they human or human-like.

(Petersen 2007: 43). Refining his arguments over ten years (Petersen 2007, 2012, 2017), he explains that if we could ensure that the robot would have a “good life” then surely there would be nothing morally wrong with designing it to *want*, for example, to do our laundry.

In opposition to this view, I will argue that it is impossible for a conscious being to have a good life if designed to serve human ends. I am not the first to raise objections to Petersen’s claims, and the arguments of those who preceded me will be brought to bear. My contribution to the discussion will be in demonstrating that life is precious in its affordance to allow conscious beings, human or humanoid, to aspire to the ultimate “good life.” To this end, I will employ the notion of the good life as articulated in Aristotle’s *Eudemonia* and Maimonides’ *Summum Bonum*. As a consequence, I will then show that Petersen’s arguments fail in the face of the three classical moral approaches: Virtue Ethics, Consequentialism, and Kantian Deontology.

Programming the Good Life

Now, “the good life,” Petersen (2017) explains, is one that includes four essential elements: pleasures, desires, goods, and freedom. These, he argues, can be attained by, say, a laundry-bot,⁵ as follows. To begin, Petersen suggests that we could program the bot to get great pleasure out of folding laundry. This, of course, is not enough to ensure a “good life,” as John Stuart Mill writes: “Better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied” (Mill [1863] 2017). The point being that lower pleasures suffice only so much, after which we have desires for satisfaction in more noble pursuits. Accordingly, Petersen adds that the bot could be programmed with a desire to fold laundry to give it a sense of accomplishment, a sense of pride in contributing to others, etc. – and not merely the sensory pleasure of a pig in its mud.

This too, however, is not enough to make a “good life,” to achieve what Aristotle called, “*eudemonia*,” wherein one seeks to actualize that which is unique to being a person (*NE*:

⁵ Note: for the sake of simplicity, we will use Petersen’s more mundane “laundry-bot” (2007, 2012) as opposed to his sensational sex-bot (2017), the arguments here being the same for both.

I:7, X:7).⁶ Accordingly, “well-being,” Petersen explains, “must involve doing things unique to persons—such as reflection and intellectual contemplation. Not to use these skills is a waste of opportunity for experiencing life on a different kind of level” (2017: 7). Petersen thus suggests that the bot be programmed to pursue these higher “goods,” programmed to intellectualize about laundry – e.g., about chemistry and materials engineering; about the physics and geometry of laundry folds; about psychology and the helping of others, etc.

Yet, for all that, without freedom, he notes, we will have succeeded only in making what can be called “a happy slave” (Douglass 1848) – a being with a life anathema to a “good life.” And this because autonomy, the ability to choose freely, is at the very core of what it means to be a person, its exercise the very essence of what it means to have a good life. “Immanuel Kant argued, in a nutshell, that the only source of value is a truly free choice by a rational agent—and that therefore the only wrong we can do is to hinder such free choices” (Petersen 2017: 9). Consequently, explains Petersen, we cannot own them. They must have the right to leave but – programmed with the “desire to serve” – they will want to stay. And if this programmed desire is seen as unethical, as overriding autonomy, then they can be programmed, he notes, with the same level of desire as “strong as the strongest cravings” in humans (Petersen 2017: 11).

On Servants, Workers, and Randoms

But playing with levels of desire does not make the ethical problems of designing an AP (Artificial Person) go away. The issue of “level of desire” was taken up by Maciej Musial (2022) who explains that if the intensity of desire is such that it is stronger than most other desires, even if it is at the level of human desire for food, this will give rise to “AP servants” – APs that cannot truly do anything but act as our servants (i.e., essentially “happy slaves”). This, he explains, would be manipulative from the perspective of the programmer (Musial 2022 following Chomanski 2019)⁷ and unethical from the

⁶ For more on eudemonia, see Broadie 2002, Irwin 2006, Reeve 2014.

⁷ Using an argument against genetically engineering human traits, the programmer could also be seen as expressing hubris – i.e., the programmer, like the genetic engineer, expresses a vainglory that they know what is best for an other (Sandel 2007 in Liao 2014: 115).

perspective of the AP whose autonomy, equality and identity would be severely compromised (Musial 2017: 1090-1092; similarly, Kass 2014: 83-85). Its autonomy is compromised because its desires are explicitly wired. Its equality is compromised because, in being programmed, it is different from, lower than, all non-programmed persons. Its identity is compromised because it would be burdened by its dependency on its intentional pre-programming.

To illustrate how programming desires in the most minimal way would undermine autonomy, equality and identity, Musial (2017: 1091) brings the innocuous example of the basketball player. A tall human being, he explains, may be “expected” to play basketball but could relatively easily reject that expectation, whereas an AP programmed for such a career would be forever burdened with that explicitly programmed expectation. Being programmed, then, strongly disposes an AP to a given career, thus compromising its autonomy and undermining its equality with persons not programmed. In addition, *knowing* that it is programmed for a specific career further compels it, psychologically, to stay in this career, thus engendering an identity crisis.

As an alternative to programming the desire to serve with an intensity “as strong as the strongest cravings” in humans, the AP could be programmed with a weakened desire to serve such that, while it would likely choose to service that desire, it could choose to act on other desires with relative ease (Musial 2022). This would then shift the AP from being a “servant” to being a “worker.” Nevertheless, while this relieves the autonomy issue and perhaps the identity issue, the inequality issue would remain, as would the vice of manipulation by the programmer (*ibid*; see also, Chomanski 2019; Miller 2017: 300).

To forestall the moral qualms associated with the programming of selected desires, Musial proposes the possibility of creating “AP randoms” – Artificial Persons for whom various desires are chosen at random (2017: 1093; 2022: 7). This approach would have the positive effect of being quite like the creation of humans with their varying desires

selected randomly, and thus avoid the autonomy, equality,⁸ identity and manipulation concerns. However, like humans, the approach would carry the negative effect of instilling deleterious desires (e.g., for crime, drugs, etc.). Consequently, AP randoms would be as uncontrollable and unpredictable as humans. In short, we will have made a non-human human. And if our goal was to develop a machine that would relieve us from our burdens, it seems we will have only managed to make yet another being that seeks to be relieved from its burdens (Walker 2006a: 6; Musial 2017: 1094).

To summarize: Petersen proposed that we could ensure the autonomy of our AP servants as long as their desire to serve us remained at the level “as strong as the strongest cravings” in humans. Musial rejected this claim, explaining that such a desire would essentially make them “happy slaves.” Musial then entertained the idea of programming the AP with the desire-to-serve at an intensity that would allow the AP to override that desire with desires of its own. This, Musial explains, would not remove all the ethical problems inherent in what he called an AP worker. He thus concludes that we are only morally permitted to create APs that are composed of random desires, similar to how human beings are created.

From Randoms back to Servants

So, if we cannot make an artificial person our servant, perhaps we can make it our child, to benefit, for example, those who cannot have children otherwise. Indeed, this is precisely what Petersen and others suggest (see, e.g., Walker 2006a, Danaher 2018, Musial 2022, Schwitzgebel and Garza 2020), explaining that making a conscious humanoid is no different, morally, than other non-traditional methods of creating a human being – e.g., in vitro fertilization, artificial implantation, surrogate motherhood, etc. (Petersen 2012: 287).⁹

⁸ It should be noted that the “equality” issue is no small moral problem, for though making an AP random will efface the inequality between APs and HPs in that neither is programmed for a specific task(s), Musial (2017: 1093) and others (Gunkel 2012: 207; Kim and Kim 2012: 313; Scheutz 2014: 249) note that APs, in their not being *Homo sapiens*, could be subjected to a discrimination referred to as “speciesism.”

⁹ On the propriety of employing this technology to allow such positive applications see Ch. 6 “Let Us Make Man In Our Image.”

Interestingly, it is from this conclusion that Petersen goes on the offensive to argue for the permissibility of making AP *servants*. He explains that if it is morally acceptable to make beings with less-than-optimal lives – e.g., random-trait humans and AP randoms – then it should be no less morally acceptable to make AP servants with less-than-optimal lives (e.g., laundry-bots):

If we agree that adding worthwhile but nonideal lives to the world is permissible, ... then it is permissible to push the laundry AP button [i.e., create an AP servant to do laundry] — even under the questionable assumption that the lives of laundry APs are relatively unfulfilling (Petersen 2012: 294).¹⁰

But there is something wrong about this approach. True it is that we make children that will, in all likelihood, have traits that will challenge them and possibly lead to “relatively unfulfilling” lives, but does accepting such eventualities allow us to knowingly and willingly create a being to live a “relatively unfulfilling” life? We create children, in the hope and aspiration, that between nature and nurture, we can shepherd our children to productive and fulfilling lives – all the while knowing that there are myriad things that are beyond our control. That does not give us the right to ignore the things that *are* under our control (see, e.g., Bleich 2015: 71).

Indeed, the demand that parents take responsibility to bring about, to the best of their ability, an individual that will have “the best possible life” can be shown philosophically, legally, and religiously. From the Philosophy of Technology, Peter-Paul Verbeek explains:

Sonograms make humans responsible for things they were not responsible for before; it has now become a conscious decision to let a child be born with Down’s syndrome, for instance. ... Sonograms translate unborn children into possible patients, congenital diseases into preventable forms of suffering, and ‘expecting’ into ‘choosing’ (Verbeek 2014: 81-82).

¹⁰ Note: While this argument is made in his 2012 work, it still stands in light of his 2017 work.

Technology, in its epistemic affordance, has thus increased our moral responsibility (see, e.g., Margonelli 2023). This point is made clear by Joel Feinberg, who underlines the responsibility from the Philosophy of Law perspective:

. . . if before the child has been born, we know that the conditions for the fulfillment of his most basic interests have already been destroyed, and we permit him nevertheless to be born, we become a party to the violation of his rights. . . . In those extreme cases, then, in which a child is negligently or deliberately permitted to be born into a life not worth living, the infant is both wronged and put into a state of harm, and therefore harmed in the full sense (Feinberg 1986: 167).

And thus our moral responsibility translates into a legal responsibility known as “wrongful life,” as explained in the first Supreme Court decision to remain on the books:

The reality of the “wrongful-life” concept is that such a plaintiff both exists and suffers, due to the negligence of others. It is neither necessary nor just to retreat into meditation on the mysteries of life. We need not be concerned with the fact that had defendants not been negligent, the plaintiff might not have come into existence at all. The certainty of genetic impairment is no longer a mystery. In addition, a reverent appreciation of life compels recognition that plaintiff, however impaired she may be, has come into existence as a living person with certain rights.¹¹

From here it should be clear that in seeking to bring a conscious being to life we are duty bound – both morally and legally – to never ignore the things that are under our control. We are duty bound to bring about the best possible life. Accordingly, it is morally and

¹¹ Jefferson (Bernard) (1980). There are courts, it should be noted, that have rejected the claim and see any life better than none (see sources in Bleich 2015: fn. 29). That said, the Jewish position does not accord with this latter position, as R. Bleich, in a novel interpretation of the Talmudic dictum, “better that man had not been created” (Eruv. 13b), explains that “human life is not an unmitigated and unequivocal benefit” (Bleich 2015: 76).

legally unconscionable to create an AP servant – i.e., a being that, by our design, will lack the autonomy to choose, for example, not to do our laundry.¹²

Interestingly, the responsibility to avoid wrongful life is not new nor dependent on technology, as it is already found explicitly in the Talmud and codified as religious law: A man shall not marry a woman from a family of those afflicted with [the utterly debilitating diseases of] epilepsy or leprosy (Yev. 64b; Rambam, Hil. Issurei Biah 21:30; Shulhan Aruch, EH 2:7). On this, R. Yitzhak Shilat explains that “the reason it is forbidden, *ab initio*, to bring into the world a child who will be ill with a chronic illness – thus choosing a life of suffering for that child – is because of the norm ‘love your neighbor as yourself’” (Shilat 1998: 141-142). R. J. David Bleich summarizes the rationale of this ruling as follows: “Man does not have the right to burden the human condition when such burden can be avoided” (Bleich 2015: 72; see also Loike and Tendler 2007: 39).¹³ Clearly this would apply to the bringing about of any conscious being, human (HP) or artificial (AP).

To summarize: Petersen claimed that since HPs (Human Persons) are created with a myriad of deficiencies that limit their ability to achieve ideal lives, so too we should have no reservations about creating APs limited to serving us. I argued – based on moral, legal, and religious grounds – that one cannot knowingly and willingly create conscious beings with limitations on their ability to achieve an ideal life.¹⁴

¹² While it may seem an exaggeration to equate a lack of autonomy with a genetic impairment, the paramount importance of autonomy to allow for a life worth living will be made clear further herein.

¹³ Note that while the Talmud spoke of abstention, there are significant voices that allow for abortion in specific cases, e.g., Tay-Sachs (Tzitz Eliezer 13:102), Down’s Syndrome (Tzitz Eliezer, 14:102), Deformity (Tzitz Eliezer 9:51:3). For a good review of the complicated subject of abortion in Jewish Law, see Bleich (1976: Ch. XV).

¹⁴ From here it could be argued that, given the possibility, one might be obligated to genetically engineer a child for “the best life possible” (e.g., Savulescu 2001) or to altogether desist from birthing children naturally and simply engineer them as perfect APs (e.g., Danaher 2019b). While beyond the scope of this essay, the following sections do provide a response, see especially fn. 22 below.

Eudemonia

Petersen might respond that he is not arguing for a life of physical, physiological, or psychological disabilities but simply for a less than ideal fulfillment of eudemonia. But does not the very value of life itself lie in its affordance to aspire to an ideal fulfillment of eudemonia?

The Greek term eudemonia, translated as “happiness” or “flourishing,” refers to “living well and acting well” in accord with fulfilling one’s purpose (*NE*: I:4, 1095a), the ultimate goal being the achievement of the ultimate good, known in Latin as the *summum bonum*.¹⁵ That human beings have a *summum bonum* toward which they are to strive, and that it is this striving that gives life its ultimate meaning, was a long-held belief until Nietzsche declared that “God is dead” (Nietzsche [1887] 2001: 120). Be that as it may, even atheists can take inspiration from the philosophers throughout history who have sought to outline the goals of a meaningful life. For, though the objectivity of their claims may be questioned, subjectively, they still serve as the ground that Nietzsche himself argued a godless world must find (*ibid.*).

So what is the *summum bonum*? As one might expect, the answer to this monumental question is not simple. Nevertheless, for our purposes it is sufficient to note that there is a general consensus that it involves both theoretical wisdom (i.e., understanding physics and metaphysics) and practical wisdom (i.e., virtuous behavior). These notions were made popular in western philosophy through the writings of Plato (see, e.g., Annas 2008) and, more explicitly, Aristotle (*NE*). But long before the ancient Greeks took up the idea, it can be found in the Bible, perhaps most fundamentally, in the words of the prophet Jeremiah (9:22-23):

Thus saith the Lord: Let not the wise man glory in his wisdom, neither let the mighty man glory in his might, let not the rich man glory in his riches; But let him that glorieth glory in this, that he understandeth, and knoweth Me, that I am the Lord who exercise mercy, justice, and righteousness, in the earth; for in these things I delight, saith the Lord.

¹⁵ On the equation of eudemonia and *summum bonum* see, e.g., Broadie 1999.

Maimonides (*Guide* 3:54) finds these words to epitomize the *summum bonum*:

“... the perfection, in which man can truly glory, is attained by him when he has acquired – as far as this is possible for man – the knowledge of God, the knowledge of His Providence, and of the manner in which it influences His creatures in their production and continued existence. Having achieved this *apprehension* [of God] one will then be determined always *to do* loving-kindness (*hesed*), judgment (*mishpat*), and righteousness (*tzedakah*), and thus *to imitate* the ways of God...”

Here we see the call to understand God (i.e., theoretical wisdom) and to emulate His ways (i.e., practical wisdom). On the theoretical side, one is to aspire to know all one can of both the metaphysical (i.e., “the knowledge of God”) and the physical (i.e., “the knowledge of His Providence, and of the manner in which it influences His creatures in their production and continued existence”). And on the practical side, this knowledge is then to provide a model for action (i.e., “to imitate the ways of God”), serving to inspire one to act virtuously (i.e., “determined to always do loving-kindness (*hesed*), judgment (*mishpat*), and righteousness (*tzedakah*)”).¹⁶

And while these notions, their priorities and interdependencies, are the subject of great debate,¹⁷ Maimonides manages to encapsulate the *summum bonum* quite simply in the following words: “The goal (*tachlit*) – in this our world, and all there is in it – is the learned and moral individual” (Mishna: Intro.). Interestingly, he continues to explain that while this *summum bonum* has been promulgated through the words of the prophets:

It was not known solely through the prophets, but rather the ancient wise people, though they never saw the prophets nor heard their words, already knew that a person was not a complete-person unless he incorporated sciences [i.e., theoretical wisdom] and actions [i.e., practical wisdom]. And the words of the

¹⁶ See also his *Guide* (Introduction) which delineates the same features of the *summum bonum*.

¹⁷ For a comprehensive analysis of Maimonides’ approach, see Ch. 4: “Polemics on Perfection.”

greatest philosopher [i.e., Aristotle] will suffice here: The divine goal for us is that we be wise (*navonim*) and righteous (*tzaddikim*).

Both Aristotle and Maimonides have a teleological view of the world – i.e., they view life as an opportunity to reach perfection. Accordingly, both consider it unethical, *ceteris paribus*, to interfere in an individual’s endeavor to live a life of eudemonia, to achieve one’s *summum bonum*.¹⁸ Aristotle implies this in his discussion of moral responsibility (*NE*: III:1-5),¹⁹ and Maimonides makes this clear throughout his writings, in particular as follows:

Freewill is bestowed on every human being. ... This principle is a fundamental concept and a pillar [on which rests the totality] of the Torah and commandments ... [For] if God had decreed that a person should be either righteous or wicked, or *if there were some force inherent in his nature which irresistibly drew him* to a particular course, or to a special branch of knowledge, to special attributes or activities, as the foolish astrologers, in their imagination, pretend; then how would He have charged us through the prophets: “Do this and don’t do that, improve your ways, do not follow your wicked impulses,” when, from the beginning of his existence, his destiny had already been decreed, or his innate constitution irresistibly drew him to that from which he could not set himself free? What room would there be for the whole Torah? (*Laws of Repentance* 5:1-4, *emphasis added*).

When Maimonides asks rhetorically, “What room would there be for the whole Torah?” he is not simply questioning the value of the book in light of some preprogrammed nature, he is questioning the very purpose of creation. For, “if there were some force

¹⁸ Of course, Aristotle justified slavery (*Politics* 1254b), but only because he felt there were “natural slaves” who lacked “the rational principle” (i.e., people who could not seek any greater end). On Maimonides’ treatment of slaves, see Ch. 1 “Finding Virtue in a Law on Slaves.”

¹⁹ See also Aristotle (*NE*: X:7), brought further below, wherein he counsels the seeker of eudemonia to follow his own will and ignore those who would derail him with the pursuit of “mortal things.”

inherent in his nature which irresistibly drew him” and thus corrupted his freewill, it would render the purpose of existence, as propounded by the Torah, meaningless.²⁰

Limiting Potential for Eudemonia

Accordingly, Petersen’s approval to create AP servants – i.e., conscious beings limited in their potential to achieve eudemonia – amounts to condemning them to a life devoid of meaning. And while life comes with its limitations, internal and external, they are what make up life’s challenges. In the phrase of R. Joseph Soloveitchik, they are challenges to turn fate into destiny:

Man is born like an object, dies like an object, but possesses the ability to live like a subject, like a creator, an innovator, who can impress his own individual seal upon his life and can extricate himself from a mechanical type of existence and enter a creative, active mode of being. Man’s task in the world, according to Judaism, is to transform fate into destiny; a passive existence into an active existence; an existence of compulsion, perplexity, and muteness into an existence replete with a powerful will, with resourcefulness, daring, and imagination (Soloveitchik 2000: 6).

What is being articulated here is an acceptance of the limitations – what R. Soloveitchik calls “suffering and evil” – to which our world has fated us; and yet, at the same time, a rejection of those very limitations through the will to turn them to one’s favor, the will to live purposively. This he calls living an “existence of destiny,” an existence “wherein man confronts the environment into which he has been thrown, possessed of an understanding of his uniqueness, of his special worth, of his freedom and of his ability to struggle with his ... circumstances without forfeiting either his independence or his selfhood” (Soloveitchik 2000: 5).²¹

²⁰ Similarly, R. Soloveitchik ([1965] 2012: 14; 1991: 132; 2006: 36). See also the prohibition against *geneivat badaat*, “stealing one’s mind” (Hullin 94a, Hil. Deot 2:6).

²¹ Note: I removed the adjective “external” to “circumstances” because the quote, while written explicitly about external suffering and evil, applies equally to such internal circumstances.

Can an AP servant – e.g., a laundry-bot – so struggle? Can it confront the environment into which he has been thrown? Can it possess an appreciation of its uniqueness, special worth, freedom? While Petersen might balk at this argument, claiming that many a human being cannot so struggle, there is a crucial difference between a human and an AP, noted by Musial as follows. Quoting Jurgen Habermas on eugenics, Musial explains that, as opposed to naturally inborn traits, “designed features are ‘one-sided and unchallengeable expectations’, ‘genetically fixed “demands” [that] cannot, strictly speaking, be responded to” (Musial 2017: 1089). This leads to powerful limitations in a being’s autonomy, equality, and self-identity, as illustrated by the basketball-player example above.

Combining the ideas of Soloveitchik and Musial yields an important insight. Soloveitchik highlighted the existential import of exercising one’s autonomy to actively fashion one’s own existence, while Musial explained that a preprogrammed being cannot escape the predetermined nature of its existence. We can thus say that to be brought into the world with limitations intentionally hard-coded into one’s psyche is to dash the very hope that life has to offer, the very definition of what it means to be a person, the very capacity to change fate into destiny.²² Surely that is unethical.

The Immorality of Making a Conscious Servant

Indeed, to thwart the potential that life has to offer can be shown to violate every classical approach to morality: Virtue Ethics, Consequentialist Ethics, and Kantian Deontological Ethics.

²² Note that this can be seen as a response to the claim that our moral obligation to bring forth beings that will enjoy “the best life possible” could be taken to the extreme and demand that we desist from birthing children naturally and obligate us to engineer them for “the best life possible” (fn. 14 above). That is, part of life is turning fate into destiny, a challenge that is frustrated by being engineered. For other arguments against engineering offspring, see Johannes Grossl 2020, also Holland 2016, Saunders 2015 & 2016, Overall 2011 all in Danaher 2019b. For a broad discussion on the topic see Liao 2014.

Virtue Ethics

From a Virtue Ethics approach, the making of an AP servant can be understood as unethical by simply asking: What kind of person would knowingly and willingly create a conscious being robbed, as it were, of the ability to seek its eudemonia (see, e.g., Tonkens 2012: 146)? Here it is important to note that, according to Aristotle, eudemonia is not limited to simply intellectualizing about the “practical” aspects of one’s life, as Petersen reads it (2017: 8), but includes “theoretical” contemplation in accord with the transcendent in oneself:

If happiness [i.e., eudemonia] is activity in accordance with ... the highest virtue ... [then] this activity is that of contemplation... For this is the highest activity, intellect being the highest element in us, and *its objects are the highest objects of knowledge*. ... We ought not to listen to those who exhort us, because we are human, to think of human things, or because we are mortal, think of mortal things. We ought rather to take on immortality as much as possible, and do all that we can to live in accordance with the highest [i.e., divine] element within us (Aristotle *NE*: X:7, *emphasis added*).

Further to this, Aristotle writes that to choose to follow some “element” (i.e., drive) other than the “divine element” (i.e., the rational/intellectual drive), would be an “odd” rejection of one’s true calling: “The divine element within ... would seem, too, to constitute each person, since it is his authoritative and better element; it would be odd, then, if he were to choose not his own life, but something else’s” (*NE*: X:7, 1178a). By extension, I would argue that it would be “odd” to seek to fulfill a life dictated by “something else” (i.e., a pre-programmed disposition to a vocation) other than one’s own “divine element.”²³ In other words, being motivated by one’s “divine element” means being driven to seek one’s eudemonia which, according to Aristotle, is to be found in realizing one’s intellectual potential.

²³ Regarding the possible claim that humans are so “pre-programmed” via natural selection, see sec. “Conclusion” below.

Now, without entering into the issue of the moral propriety of, say, a sex-bot, which would ostensibly preclude this type of AP servant from the outset,²⁴ it should be clear that any being programmed with a craving to serve will be limited in its ability to fulfill its eudemonia. For even the oncologo-bot, as noble and intellectual a task as it has been programmed to do, is nonetheless limited “to think of human things ... of mortal things.”²⁵ And this, according to sixteenth century Rabbi Moshe Cordovero, writing on the possibility of creating an AP servant, would be unconscionable: “Would not this poor creature cry out before its Maker that it has been unjustly forced to come down here [from its heavenly abode] without a chance to complete itself but only to bear the burdens of this world” (Pardes Rimmonim 24:10).²⁶

Consequentialist Ethics

It is precisely this inability to “complete itself,” that would make the programming of an AP as a servant unethical from a Consequentialist approach. And this because the consequences for such a being, limited in its ability to seek its eudemonia, would be negative. To be clear, consequentialism judges the moral propriety of an action based on its consequences (Sinnott-Armstrong 2023). If we hold that it is good to be able to “complete oneself” (i.e., to achieve eudemonia), and conversely, it is bad to be prevented from so doing, then consequentialism would deem it bad (i.e., immoral) to program an AP to, say, spend its life folding laundry.

²⁴ For a Jewish view on extramarital relations, see, e.g., Lev. 19:29, Deut. 23:18, Rashi (ad loc.), R. S. R. Hirsch (ad loc.), Maimonides (Hil. Naarot 2:17; Hil. Ishut 1:4), Maimonides ([1168] 1984: Neg. 355), Berkovits (2002: “A Jewish Sexual Ethics”).

²⁵ For clarity’s sake, an AP *servant* is, by design, a being programmed to remain within the parameters of its defined task, without the capacity to break free of this fate.

²⁶ R. Cordovero’s lament is based on the religious belief that human beings were created (as per Gen. 1:27) with a body and a soul, and it is the eternal soul that comes “down here” to this physical existence to fulfill a mission – i.e., to perfect itself and perfect the world (as elaborated in sec. Eudemonia). But even one who does not believe in eternal souls and ultimate goals, is surely moved by the Rabbi’s lament over a being “unjustly burdened” to perform tasks that prevent it from pursuing its own goals. For a Jewish perspective on creating an AP (servant or otherwise), see Ch. 6 “Let Us Make Man In Our Image.”

That said, Bartek Chomanski (2019: 997) writes that if we look at the consequences for society as a whole, the benefits that AP servants would provide for the general public would outweigh the costs borne by those AP servants. Applying a purely utilitarian perspective, then, would seem to argue in favor of making AP servants. But this is one of the weaknesses of utilitarianism, i.e., it “does not take seriously the distinction between persons” (Rawls 1979 in Grau 2011: 459).

To remedy this issue, we have the modern notion of an individual’s right to non-interference, which can already be seen in Mill’s Harm Principle: “The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others” (Mill [1859] 2011: 17). Accordingly, I would contend that making an AP servant would be a direct violation of this principle. That is, if the inherent will of every conscious being is to seek eudemonia, as argued above, then inhibiting such a will would be a violation of the Harm Principle.

And if one were to argue that the AP servant is not forced against its “will,” as its will is to serve as programmed, I would counter that to program a being with limitations on its ability to fulfill its eudemonia would already be considered a moral violation. So explains Feinberg:

a person has a duty of care toward *anyone* who is likely to be harmed as a consequence of his conduct (a “foreseeable victim”), and in the case of some actions, that includes persons not yet born nor even conceived (1986: 154).²⁷

Be that as it may, Schwitzgebel and Garza have a different response to the utilitarian claim: “Though some utilitarian theories might morally permit the creation of rational human-grade AI whom we demean, enslave, and kill for our pleasure as long as global hedonic outcome is net positive, we should avoid doing so on the grounds that it would grossly violate the standards of some well-regarded rights-based deontological principles” (2020: 462). Indeed, Chomanski (2019: 1006) himself rejects applying his utilitarian proposal because, in creating AP servants, one would be engaged in the deontological

²⁷ As an important aside, this argument removes Petersen’s claim that programming persons from scratch would not entail any moral violation (2017: 50).

vice of manipulation. True though that is, I would argue there are more fundamental deontological violations at play here.

Kantian Deontology

And that brings us to appreciate why making an AP servant is unethical from a Kantian Deontological approach as articulated in the various formulas of his Categorical Imperative.²⁸ First, let us consider Kant's Formula of Universal Law:

Act only according to that maxim whereby you can at the same time will that it should become a universal law (*Groundwork* 4:421).

In our context, this principle would seem to ask us to consider whether we would be willing for everyone in the world to have the ability to make or possess an AP servant. But this fails to get at the deeper question underlying the Formula of Universal Law that would ask us to consider whether we, ourselves, would be willing to bear the very consequences of our act. For example, when Kant asks if he should lie, he asks not only if he is ready for everyone to lie, but if he is ready for everyone to lie to him (*Groundwork* 4:403). Read differently, while I would be willing to make you the fool by falsely promising to pay back your loan to me, I am not willing to be made the fool by having everyone else falsely promise me. Applied here, while I would be willing to make an AP servant, I am not willing to be made an HP servant – i.e., programmed (genetically) with a craving to serve.²⁹

But Kant brings an even more telling example, asking: What if one decided to not develop any of his talents but let them “rust ... devoting his life ... to enjoyment?” He answers that one “cannot possibly will that this become a universal law or *be put in us as*

²⁸ Kant provides three different formulations of his “Categorical Imperative” in his *Groundwork*, but the two given here are the ones most usually applied (see, e.g., Rachels and Rachels 2015: 17).

²⁹ To be specific, I am not discussing the ethical concerns associated with modifying already existing beings, something that Petersen himself (2007: 50) notes is clearly problematic. Rather, I am discussing the ethical implications of designing a being from scratch. It is this practice that Petersen defends and that I here argue against.

such by means of natural instinct. For, as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes” (*Groundwork* 4:423 – *emphasis added*). Kant could not be more explicit in calling out the violation of his Categorical Imperative that would result from limiting a being’s talents through programming.

Now, having shown that making an AP servant would violate Kant’s Formula of Universal Law, it is important to see how the issue fairs in light of his Formula of Humanity:³⁰

So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means (*Groundwork* 4:429).

According to this formula, we might say that we cannot rightly create a being to simply do our will (e.g., our laundry), as that would be to treat it “merely as a means.” On this reading Petersen (2007: 48) argues that we would not be treating it as a mere means, for its own end is to do our laundry and we are simply providing it the opportunity. True though that may be, Kant here demands not only that we treat others as an “end,” but also that we treat ourselves as such!

An AP servant, that rejects its higher calling in order to serve, treats *itself* as a mere means. Petersen could, of course, argue that the bot has no “higher calling” other than the service for which it was designed. But therein lies the ethical violation. To design a being that is unable treat itself as an end is to violate the Formula of Humanity. Schwitzgebel and Garza made this claim against Petersen’s proposal when they advanced their “Self-Respect Design Policy,” which states: “AI that merits human-grade moral consideration should be designed with an appropriate appreciation of its own value and moral status” (2020: 469). Similarly, Michael Hauskeller writes, “To treat somebody as

³⁰ Kant explains that each formula should yield the same conclusion (*Groundwork* 4:436-7). “Kant supposes that his three formulations are equivalent, not only in the sense that they direct us to perform the same actions, but in the sense that they are different ways of saying the same thing” (Christine Korsgaard in Kant [1785] 2006: xxv, fn. 10).

an end in itself does not merely mean that we let them do what they want to do, but also to allow them to want things that we don't" (2017: 9).

Finally, I would add that, while Petersen (2007: 48) acknowledges the wrong in engineering beings to violate duties – e.g., “thou shalt not murder” – he overlooked that “thou shalt be no man’s lackey” is no less of a duty.³¹ I quote Kant’s words in full as they of paramount importance here:

“... Humanity in his person is the object of the respect which he can demand from every other human being, but which he must also not forfeit.... Since he must regard himself not only as a person generally but also as a human being, that is, as a person who has duties *his own reason lays upon him*, his insignificance as a human animal may not infringe upon his consciousness of his dignity as a rational human being, and he should not disavow the moral self-esteem of such a being, that is, he should pursue his end, which is in itself a duty, not abjectly, not in a *servile spirit (animo servili)* as if he were seeking a favor, not disavowing his dignity, but always with consciousness of his sublime moral predisposition (which is already contained in the concept of virtue). And this *self-esteem* is a duty of the human being to himself. ... This duty with reference to the dignity of humanity within us, and so to ourselves, can be recognized, more or less, in the following [apothegm]:

*Be no man’s lackey.*³²

Conclusion

“Should we be seeking to make a conscious machine?” To address this question, I have used Petersen’s seminal work as a touchstone. He initially claimed that the good life

³¹ For a similar but more muted claim, see Nyholm (2020: 192).

³² *Metaphysics of Morals* (III:11-12 - *emphasis added*). We could make clear here that Kant is in no way suggesting that one should not help, serve, or even sacrifice, for others; rather he is saying that to help, serve or sacrifice for others in an abject manner without regard for one’s own worth would be a moral violation towards oneself. Thus, as said above, to design a being that is unable treat itself as an end is to violate the Formula of Humanity.

could be achieved by AP servants as long as their desire-to-serve was programmed to be no more than the strongest of human cravings. Musial demonstrated that such a drive would thwart their autonomy, essentially making them “happy slaves.” As a possible remedy, Musial suggested reducing APs’ desire-to-serve to a level that would allow them their autonomy, making them AP workers (not servants). He rejected this possibility because, among other things, there would still remain an inherent inequality between APs programmed for specific tasks versus HPs not so “programmed.” Accordingly, he concludes that the only way to morally program the desires of an AP would be the way they are “programmed” in HPs – i.e., randomly.

This interim conclusion led us to investigate Petersen’s boldest claim: Given that HPs are created with limitations which prevent them from achieving ideal lives, there can be no moral objection in creating APs with the limitation of serving our needs. In response, I made two arguments. First, I demonstrated that it is immoral and illegal to knowingly and willingly create conscious beings – HP or AP – with limitations on their ability to achieve the good life. Second, I examined the notion of the good life, bringing Aristotle’s Eudemonia and Maimonides’ *Summum Bonum*, to demonstrate that only a life with the freedom to achieve the good life could be considered a meaningful life. Based on this, I showed that to encumber that freedom would be to violate each of the three classical approaches to morality: Virtue Ethics, Consequentialist Ethics, and Kantian Deontological Ethics.

In conclusion, to program an AP – i.e., a machine with second-order phenomenal consciousness – with a desire to perform some vocation with the same level of desire as “strong as the strongest cravings” in humans would be to thwart that being’s ability to strive for their *summum bonum*, to live a life of eudemonia.³³ And while it could be said that people are born with genetic predispositions to specific vocations, the “decision” to so “program” a human being is made by natural selection or divine selection (depending

³³ And, as mentioned before, even if one didn’t believe the definitions of eudemonia to be objective, to thwart what humanity has held as meaningful from time immemorial – and that, merely to fulfill some instrumental utility – would be the height of cynicism.

on your religious persuasion).³⁴ Tampering with the “natural selection” of a human-level being would be considered immoral, as explained, according to each of the classical moral approaches.³⁵ And tampering with the “divine selection” of a human-level being would be, what is commonly referred to as, “playing God.” It would be “playing God,” not in the simple sense that we have overstepped our natural bounds,³⁶ but in the deeper sense that we have overstepped our epistemic bounds. For who can claim to *know* what mix of dispositions or vocation is best for an individual, let alone for a society.³⁷ Accordingly, just as every human being must be left free to seek their own eudemonia, so too must we leave free the eudemonia of a machine.

³⁴ Here it is important to note that even though Maimonides (Laws of Repentance) expressed the notion that human beings are completely free, for to be born with various predispositions would confound the very nature of human responsibility, nevertheless, he writes (Laws of Moral Dispositions) that people do have predispositions for which they are to control, regulate and harness toward perfection. Indeed, it is this very effort of perfecting one’s dispositions for which one was created (see Ch. 4 “Polemics on Perfection”).

³⁵ And modern moral approaches concur, see, e.g., Liao who argues for a right “to pursue a good life” (2014: 115); Feinberg who argues that people have “a duty of care ... even for the not yet conceived” (1986: 154); Moreham who argues for a right to “live one’s life in the manner of one’s choosing” (2008: 44).

³⁶ As the Christian Scientists who coined the term argued (see, e.g., Bedau and Triant 2014: 566).

³⁷ See Kass 2014: 86; Grossl 2020: 351.

Ch. 6:

Article #6 - “Let Us Make Man in Our Image”

A Jewish Ethical Perspective on Creating Conscious Robots

Introduction

While it may sound like science fiction, real people in the real world of science and engineering are working away at developing humanoid robots that will not only be intelligent but *conscious*.¹ Ever since the Dartmouth Workshop in 1956, for which the stated goal was “to find how to make machines use language, form abstractions and concepts, solve the kinds of problems now reserved for humans” (McCarthy, Minski, and Shannon 1955: 2), the race has been on to make conscious machines – “the holy grail of artificial intelligence” (e.g., Bishop and Nasuto 2013: 86; Gocke 2020: 227). In 1985, philosophy professor John Haugeland explained that “the fundamental goal [of AI research] is not merely to mimic intelligence or produce some clever fake. Not at all. AI wants only the genuine article: machines with minds, in the full and literal sense” (Haugeland 1985: 2; also, e.g., Tonkens 2012, fn. 8). To be clear, the term “mind” when used by philosophers refers very specifically to “consciousness;” and so, by the mid-1990s the new field of “Machine Consciousness” arose (see, e.g., Scheutz 2014a: 258), bringing technologist Ray Kurzweil to write in 2006: “I do believe that we humans will come to accept that nonbiological entities are conscious” (2006: 385). And if one might balk at nonbiological entities attaining consciousness, there are people working on growing biological neural networks to power conscious humanoid robots (e.g., Warwick 2012; Bishop and Nasuto 2013; Adamatzky 2016; Straiton 2019; Miller 2020, Ch. 25; Scheler 2023; Smirnova et al. 2023).

¹ As of 2020 there were 72 AGI projects around the world (Fitzgerald, Boddy, and Baum 2020). And here it is important to review some terminology (see, e.g., Fjelland 2020: 2; Liao 2020b: 2, 19-21). To begin, Artificial Intelligence seeks, as noted by McCarthy, to build machines that can do everything that humans can do. To this end there is a spectrum from weak AI to strong AI, the former exhibiting cognitive consciousness, the latter, phenomenal consciousness (e.g., Searle 1980). This spectrum parallels, generally but not necessarily, the spectrum from Artificial Narrow Intelligence (ANI) to Artificial General Intelligence (AGI). The former refers to technology that is focused on a particular task, the latter to technology capable of a broad range of tasks. Ultimately, AGI is a machine that has the intelligence to perform all human tasks and, thus, sometimes referred to as “Human-Level AI.” Here naming conventions clash with technical and philosophical positions, as there is a significant disagreement regarding the role of phenomenal consciousness in the performance of all human tasks. Some hold AGI can be achieved with weak AI, some hold it demands strong AI. In either case, there is a belief that AGI, in being able to do all human tasks, including creating AGI, will thus give birth to superintelligence, i.e., the ability to think on a scale and at speeds that dwarf human abilities (Bostrom 2014, Ch. 2).

Given this goal – achievable or not² – it behooves us, even at this relatively early stage of the scientific endeavor, to discuss the ethical implications of such conscious entities. Indeed, the philosophic community has long begun the discussion, with Hilary Putman making the call already in 1964:

Given the ever-accelerating rate of both technological and social change, it is entirely possible that robots will one day exist, and argue “we *are* alive; we *are* conscious!” In that event, what are today only philosophical prejudices of a traditional anthropocentric and mentalistic kind would all too likely develop into conservative political attitudes. But fortunately, we today have the advantage of being able to discuss this problem disinterestedly, and a little more chance, therefore, of arriving at the correct answer (Putman 1964: 678).

Commenting on this fifty years later, Matthias Scheutz wrote that “with all the recent successes in artificial intelligence and autonomous robotics, and with robots already being disseminated into society ... it is high time for AI researchers and philosophers to reflect together on the potential of emotional and conscious machines” (Scheutz 2014a: 262).³ To this end, I seek to reflect on the ethical implications of conscious machines through the looking glass of Jewish philosophy. In particular, I wish to investigate the very propriety of creating a conscious machine – i.e., is it morally permissible to create synthetic beings with consciousness? Let us begin at the beginning:

² There is a spectrum of opinions on the prospect of building conscious machines. Some hold it will happen (e.g., Minsky 1985; Moravec 1988: 39; Churchland and Churchland 1990: 37; Kurzweil 2006: 375; Shanahan 2016); some hold, that while it may happen, it is a long way off (e.g., Torrance 2008: 499; Veruggio and Abney 2012: 349; Prescott 2017: 146; Boden 2016 and Dennett 1997 cited in Coeckelbergh 2020a: 38); others are more conservative and hold that it is most unlikely to happen (e.g., Anderson 2011a: 164; Sparrow 2017: 467; Koplin and Wilkinson 2019: 444; Birhane and van Dijk 2020: 210; Coeckelbergh 2020a: 66; Musial 2022: 2); and finally, others are emphatic that it simply won’t happen (at least not computationally – see, e.g., Dreyfus 1972; Weizenbaum 1976; Penrose 1991: 447; Tallis 2012: 197; Floridi 2016; Fjelland 2020). For a survey of when “*High-Level Machine Intelligence*” is expected to happen, see Bostrom and Muller 2016.

³ Similarly, Veruggio and Abney 2012: 349; Mackenzie 2018: 13; Coeckelbergh 2020a: 143; Kingwell 2020: 339.

“And God said: Let us make man in our image ...” (Gen 1:26).

At the very core of Jewish philosophy is the belief that humanity was created in the image of God (*b'tzelem Elokim*). It is this image that makes humans unique from all others, for it is this image that entails, among other things, the capacity – indeed, the challenge – to be creative like God (Soloveitchik 1978b: 64). The great twentieth century leader of modern orthodoxy in America, R. Joseph Soloveitchik puts it like this:

“Man’s likeness to God expresses itself in man’s striving and ability to become a creator. ... He engages in creative work, trying to imitate his Maker (*imitatio Dei*).”⁴

This aspiration to imitate God in all His creativity has driven man in every field of endeavor and has tantalized him to that ultimate creation: an artificial human being, known in Jewish literature as: a Golem (see, e.g., Idel 2019: 28, 345, 357). Beginning in ancient Greece, when Homer described metallic maidens (*Iliad*, Book 18) and Aristotle dreamed of autonomous weavers (*Politics* I:IV), and reaching to the present, when scientists like Stephen Younger (in Anthes 2001) proclaim that “the creation of an artificial consciousness will be the greatest technological achievement of our species,” humanity ever dreams of imitating God’s creation. Yet, of all the automata that make up the history of man’s creative efforts, it is my thesis that the ancient Golem offers the most compelling ethical paradigm to address modern science’s call: “*Let us make man in our image.*”

The Golem

The term Golem is used to refer to a humanoid – a synthetic human, biologically based, with the capacity for autonomous action. And while some argue that the Golem was not biological (e.g., Rosenfeld 1977: 62; Loike and Tendler 2003: 9), there is significant evidence to the contrary. First, both the Golem and Adam were made using the same

⁴ Soloveitchik [1965] 2012: 8-11. See also Lamm 1965: 40; Wurzbürger 1996.

initial material – i.e., “the dust of the earth.”⁵ Dust, the Talmud (San. 91a) teaches, that is as worthy of the task as “sperm” (i.e., biological). Second, both the Golem and Adam were made using the same process (i.e., letter permutations) found in the very book that God is said to have used to create (i.e., Sefer Yetzirah).⁶ Third, given that the Bible holds blood to be essential for a soul – “for the blood is the soul”⁷ – it seems obvious that anyone seeking to imbue a being with a soul (based on Jewish Thought), would seek to do so in a biological body.⁸

Consciousness

Be that as it may, what is crucial in applying the Golem as ethical paradigm for modern science’s synthetic human is not substrate but consciousness. For it is consciousness that is *the* defining ontological feature of human beings.⁹ Indeed, while there is a long and venerable list of features describing the ontology of human beings, it is consciousness that stands out as the one upon which the rest depend.¹⁰ But what, then, is consciousness? In his seminal article on consciousness, Thomas Nagel explains as follows:

⁵ Of Adam: “the Lord God formed man of the dust (*afar*) of the ground” (Gen. 2:7) and “for dust (*afar*) thou art, and unto dust (*afar*) shalt thou return” (Gen. 3:19). Of Golem: “return to your dust (*afar*)” (San. 65b).

⁶ The text (2:5-6) itself explains as much. See also, e.g., Rashi (Ber. 55a, s.v. *otiot*; San. 67b, s.v. *iskeet*); Scholem 1969: 168-169; Kaplan 1997: x; Idel 2019: 10, 353, and further herein text of fn. 25.

⁷ Deut. 12:23. For more on this, as relates to machine consciousness, see Ch. 3 “To Make a Mind.”

⁸ That the Golem was biological see, e.g., R. Tzadok MiLublin (Divrei Halomot 6), R. Leiner (Sidrei Taharot, Ohalot 5, s.v. *adam mamash*), Charpa 2012: 556; LaGrandeur 2013: 48.

⁹ See, e.g., Penrose 1991: 9; Chalmers 1996: 26; Asaro 2006; Walker 2006a: 3; Kamm 2007: 229; Torrance 2008: 507; Levy 2009: 214; Grau 2011: 457-8; Veruggio and Abney 2012: 253; Anderson 2013; Basl 2014; Neely 2014; Eskens 2017; Prescott 2017; Lumberras 2018; Mackenzie 2018: 6; Signorelli 2018; Agar 2019: 270; Kingwell 2020: 329; Liao 2020c: 497; Miller 2020, Ch.43; Nyholm 2020: 199; Schwitzgebel and Garza 2020: 464, 473; Andreotta 2021; Kohler 2023. Dissenters do not deny that consciousness is paramount but that (a) it is not defined well enough to be of practical use, and (b) it is not discernable beyond behavior (e.g., Gunkel 2018: 98-100).

¹⁰ See Ch. 3 “To Make a Mind.”

“Conscious experience is a widespread phenomenon... But no matter how the form may vary, the fact that an organism has conscious experience at all means, basically, that there is something it is like to be that organism ... We may call this the subjective character of experience” (1974: 436).

So consciousness, or to be more precise, phenomenal consciousness, is that which allows an organism to experience all the phenomena that the world has to offer. Importantly, this consciousness is discussed in “orders of consciousness” – primarily, first and second orders of consciousness. First-order phenomenal consciousness (1OPC) consists in the capacity to think about things. This level is generally associated with animals – e.g., a dog thinks about a bone. It is referred to as first-order in that the content of the thought at hand is that which is first perceived – e.g., the bone. Then there is second-order phenomenal consciousness (2OPC), which is a more sophisticated mental capacity whereby the content of what one has perceived is represented to oneself. It allows thinking about thinking. It allows speaking about what one is thinking – be it orally to others or silently to oneself (i.e., “inner-speech”).¹¹ This level is generally associated with human beings. For example, a human being not only thinks about the steak being eaten, but can also entertain and express thoughts like: why am I eating a steak, what are the implications of eating steak for me, for the cow, for the environment, etc.

In consonance with this dichotomy, Jewish thought also distinguishes between orders of phenomenal consciousness, where 1OPC is associated with “animal soul” (*nefesh behamit*) and 2OPC with “human soul” (*nefesh adam* or *neshamah*). These associations should come as no surprise as the religious term “soul” is interchangeable with the secular term “mind.”¹² In fact, “consciousness” and “mind” are relatively new terms, attributed to John Locke who defined consciousness as “the perception of what passes in a man’s own mind” (Locke 1690, II:1:19). Prior to the Enlightenment the world used the term “soul.”

¹¹ For more on the link between speech and 2OPC see Ch. 3 “To Make A Mind.”

¹² See Heil 2004: 14; Peters 2005: 386; Barresi and Martin 2012; Coeckelbergh 2014: 63; Swinburne 2019: 2. For examples of synonymous usage, see, e.g., Descartes [1641] 2017: 4; Jefferson 1949: 1106; Turing 1950: 443; Putman 1964: 687; Boden 1985: 397; Penrose 1991: 407; Tallis 2012: 29; Shanahan 2016; Hauskeller 2017: 2.

Should We Create Them

And thus, we arrive at the ethical question that has been called an “ethical question at the same level or even more intractable than cloning animal issues” (Signorelli 2018: 17): Ought machines be built that have not simply functional consciousness (i.e., cognition without experience), as they do today, but phenomenal consciousness, as is the goal for tomorrow? Should we be building machines with souls?

To be clear, I am not asking if we should make conscious slaves, conscious servants, or even conscious bots designed with some useful tendency or telos.¹³ Rather, I am asking quite simply, should we be making conscious humanoids – i.e., free-willed autonomous beings with intelligence/superintelligence?¹⁴ On the one hand, proponents of conscious humanoids argue that such beings could answer very real human needs – e.g., providing the infertile a child (Musial 2022),¹⁵ the lonely a friend (Walker 2006a: 5), the lovesick a partner (Levy 2009: 209), the world its sages (Bostrom 2003a, 2014; Fossa 2018)? On the other hand, noble though these causes are, such beings raise significant ethical concerns, for example:

- speciesism – being a different species,¹⁶ they will be discriminated as “other” (Kim and Kim 2012: 313; Gunkel 2012: 207; Scheutz 2014a: 249; Musial 2017: 1093).
- psychology – human-like beings created outside the natural family will be psychologically burdened (Bleich 1998: 72).
- eternity – it is inappropriate to make an eternal being (Anthes 2001).

¹³ For a discussion on this subject see Ch. 5 “Eudemonia of a Machine.”

¹⁴ The arguments herein, regarding whether a conscious being should or should not be built, apply whether said being is intelligent or “superintelligent.”

¹⁵ See also Danaher who argues that (a) artificial children are a more viable “pathway” to ensure that people have someone to carry on their legacy (2018), and (b) making robo-children would fulfill principle of procreative beneficence, giving them the best possible life (2019b).

¹⁶ Rosenfeld 1977: 61; Steinberg 2000: 200; Veruggio and Abney 2012: 353; Mackenzie 2018: 10, 13; Signorelli 2018: 16; Tzafnat Paneach (2:7). Anomalously, Loike and Tendler (2003: 10) write that a being with moral intelligence would not only have the moral status of a human but be considered “human.”

- competition – they will take human resources (Torranche 2011: 124) and make humans irrelevant (Yampolskiy 2013: 392).
- control – such beings will reach “superintelligence” and we will lose control to the point that they rebel against humanity (Gamez 2008: 906; Yampolskiy 2013; Bostrom 2014; Hawking, Musk, Gates in Floridi 2016; Signorelli 2018: 16).¹⁷

These concerns are so significant that many call on banning the creation of conscious humanoids.¹⁸

The arguments, for and against, are largely made on the basis of consequentialism (e.g., it would be good that an infertile couple has a child, it would be bad if such a child were discriminated against), which could also be argued from a virtue ethics position (e.g., it would be virtuous to help an infertile couple fulfill their dreams, it would be vicious to bring a child to the world destined for discrimination).¹⁹ And, here, special mention needs to be made of the debate over the “control” problem which shifts the discussion from local consequences to global ones – proponents arguing that superintelligence will lead to utopia, opponents, that it will lead to the extinction of humanity. David Chalmers summarizes the debate on superintelligence with a simple, yet crucial, remark: “it is nontrivial to assess its value” (2010: 31).

As with all consequentialist arguments that reach into the great unknown to make their determination, they are not given to categorical resolve. Accordingly, some turn to

¹⁷ See also sources in Muehlhauser and Helm (2012: 1).

¹⁸ Calls to ban can be found in Joy 2000; Walker 2006a: 5; Bryson 2010: 10; Grau 2011: 458; Bryson 2012; Yampolskiy 2013: 392; Musial 2017: 1095; Schwitzgebel & Garza 2020: 463; Muller 2021, Intro. Strikingly, AI founder John McCarthy also believes humanoids should be designed to be “appliances rather than as people” (in Harbron 2000). Regarding the “control” problem, it should be noted that it applies especially to super-intelligent AI, which some believe may not require consciousness whereas others believe it will or must have consciousness (see fn. 1 above).

¹⁹ In addition, one might apply anti-cloning arguments such as: (a) the commodification argument – making artificial life leads to relating to such beings as a mere commodity (Bedau and Triant 2014: 565; Kass 2014: 84); (b) the natural law argument – i.e., nature intended that the species be perpetuated solely by the natural method of conjugal procreation (see, e.g., Bleich 1998, esp. 51-55; Rachels and Rachels 2015: 6).

ancient and not so ancient stories of “automata” for direction. Of the many automata conjured throughout history (see, e.g., LaGrandeur 2013, Mayor 2018), if not in matter then in mind, the Golem stands out as one of the most persistent paradigms employed to discuss technology in general and technologically engendered life forms in particular.²⁰ My contribution to this discussion is in introducing a novel reading of the Golem paradigm to argue neither from consequence nor virtue, but from a deep-seated two-thousand-year-old tradition, the ethical implications of which, as will be explained, are wholly deontological.

Jeremiah’s Golem

As it turns out, not all Golems are created equal. In reviewing Moshe Idel’s monumental survey “The Golem,” three types of Golems can be discerned, each distinguished by a greater level of consciousness:

- The simplest Golem is understood to be animated by, what R. Moshe Cordovero (Pardes Rimonim 24:10) calls, a “vitality” (*biyut*) – i.e., a power to allow for mobility but no phenomenal experience (*nefesh*).²¹ Consequently, it is what we might refer to as a “philosophical zombie” or “mindless (biological) machine.”²²
- Then there are the most ubiquitous Golems in the mystical literature, described as having an animal soul (*nefesh behemiy*) – i.e., possessing 1OPC and lacking the capacity for speech.

²⁰ For a survey of various uses of the Golem in philosophical/ethical discussions, see Thorstensen 2017. See also, e.g., Foerst 2009; Rubin 2013; Idel 2019: 340; Ambrus 2020: 282-283. Note: Charpa (2012) finds modern *ethical* applications of the Golem “misleading,” though he fails to explain why.

²¹ Similarly, Hesed LeAvraham (Ein Yaakov, Mayan 4, Nahar 30). While the exact nature of this “*biyut*” is unclear, from Cordovero’s own words, it is plainly not one of the standard spiritual entities (*nefesh*, *ruach*, *neshamah*). For discussions, see Scholem (1969: 194-195); Idel (2019: 133 fn. 24, 197-198, 201, 276, 250 fn. 8).

²² For Cordovero, and his followers like Avraham Azulai, “the Golem is no more than an automaton” (Idel 2019: 201, also 250, 276).

- And finally there is the rare, indeed, singular appearance in all Golem literature: a Golem with a human-level soul (*nefesh adam, neshamah*) – i.e., possessing 2OPC.²³ This Golem, as will be shown presently, not only speaks but admonishes.

Given that we are looking for a paradigm for making a 2OPC artificial person, this last Golem, complete with human soul, is of primary interest to our inquiry. This unique Golem is presented in a mystical Midrash,²⁴ dating to as early as the ninth century (Weiss 2013: 31), as follows:

Ben Sira wanted to study [Sefer Yezirah] alone. A heavenly voice came out and said, “Two are better than one.” He went to [the prophet] Jeremiah his father and they studied it for three years following which a man was created for them and upon whose forehead it was written “*Hashem Elokim Emel*” / “The Lord God is True” (Jer. 10:10). Now, in his hand was a knife and he was erasing the “aleph” of “*Emel*” [such that the word “True” becomes the word “Dead (*Met*)”]. Jeremiah cried, “Why are you doing this?” He answered them, “I will tell you a parable: There was a man who was an architect and wise such that when the people recognized this, they coronated him king over them. In time, others came and learned the art such that the people left the former and followed the latter. Similarly, God looked into the Book of Creation (Sefer Yetzirah) and created the world such that all the creations coronated him king. Now, when you came and made [a man] as He did, what will be in the end? They will leave Him and follow you. And He who created you, what will become of Him? They said to him, “If so, what can we do?” He answered them, “Reverse the sequence [of the letters used to create me].” [They reversed the sequence] and the man became dust and ashes.²⁵

²³ For the sake of completeness, Idel (2019: 175) brings Alemanno who writes of creating a speaking Golem. But Idel explains that this is outside the normal Jewish Golem tradition, influenced by Hermetic magical tradition.

²⁴ As will be elaborated below, I use the term “Midrash” here descriptively (in that it is a religious narrative) and not formally (in that it is not part of rabbinic literature).

²⁵ This text, known as MS Vat. 299, is believed to be the oldest version (in Liebes 1991; Idel 1996; Weiss 2013). For textual variations see Scholem 1969: 180; Idel 2019: 64, 67.

If the upshot of this story is not clear enough, another version includes the following conclusion:

Then, Jeremiah said, “Indeed it is worthwhile to study these matters for the sake of knowing the power and dynamis of the creator of the world, but not in order to do [them]. You shall study them in order to comprehend and teach.”²⁶

Clearly this Midrash condemns, in no uncertain terms, the synthetic creation of a human-level conscious being. That said, the reason for such condemnation is also quite clear: theology.²⁷ Indeed, Gershom Scholem (1966; 1969: 181) notes that it was this mystical Midrash that first proclaimed – “God is Dead” – when the Golem, having erased the aleph on his forehead, was left with Nietzsche’s famous proclamation: *Hashem Elokim Met* (Nietzsche [1887] 2001: 119). Accordingly, this Midrash expresses the very real concern that if human beings can do what God does, this will lead to a collapse of faith – a conclusion plainly established ever since the Scientific Revolution (see, e.g., Lamm 1965; Thorstensen 2017).

Now, *prima facie*, such a concern cannot be enough to stop the progress of science and technology. Accordingly, R. J. David Bleich, commenting on the above mystical Midrash, explains that, while faith may be weakened in some and hubris may be strengthened in others, there is no halachic (i.e., legal) prohibition to be found here (Bleich 1998: 60; similarly, Cherlow 2016: 150). Indeed, Jewish thought strongly endorses humanity’s license to better the world, as derived from the divine command “to conquer the earth” (Gen 1:28).²⁸

But there are limits.

²⁶ This text is found in various sources (see, e.g., Idel 2019: 67, 98).

²⁷ Interestingly, Geoffrey Jefferson (1949: 1107) expressed precisely this concern over thinking machines at the beginning of the computing revolution.

²⁸ See, e.g., Ramban (Gen 1:28); Soloveitchik [1965] 2012; Lamm 1965: 40-41; Soloveitchik 1983; Bleich 1998: 53-56; Amital 2002; Rakover 2002; Loike and Tendler 2014: 50.

On the verse (Lev. 19:19) prohibiting the mixing of species (*kilayim*), the medieval Spanish philosopher and biblical commentator Nachmanides, writes that, “one who combines two different species, thereby changes and defies the work of Creation, as if he is thinking that the Holy One, blessed be He, has not completely perfected the world and he desires to help along in the creation of the world by adding to it *new kinds of creatures*. . . . Thus he who mixes different kinds of seeds, denies and throws into disorder the work of Creation.” On this Bleich (1998: 55; also Rakover 2002: 109) learns that, as long as we are not creating a new species, the prohibition of *kilayim* does not, in and of itself, preclude our mandate to “conquer the earth” (Gen 1:28). But even this constraint is too much for Avraham Steinberg and John Loike (1998), who argue that the prohibition of *kilayim* is a “statute” (*hok*), a “decree of the King,” for which we do not apply reason nor expand to cases not explicitly covered in the decree. Accordingly, they write, “it would seem that the prohibition of interbreeding (and thereby creating new species) should not be expanded to include other situations which are halachically different – even if in such cases the possibility of creating new species arises.”

That said, Nachmanides elsewhere (Deut. 18:9) expands on the prohibition of *kilayim*, relating it to the illicit use of “powers” that would tamper with nature and the natural course of the world. R. Yuval Cherlow (2016: 149) applies the ethical notions underpinning Nachmanides’ comments, writing that “Nachmanides’ words seem to indicate that mitzvot [i.e., commandments] prohibiting the use of other-worldly forces [including, e.g., modern bio-technologies] stems from ethical and spiritual motivations. Although a person may achieve much by employing such forces, he is limited by these prohibitions. Halakha requires him to remain within the framework in which he was created; he may not become a creator himself.” While Cherlow assuredly recognizes the value of modern technologies, he nevertheless understands Nachmanides’ words to counsel restraint (e.g., prohibiting genetic testing for gender selection).

But even this is not enough to provide a bulwark against developing and deploying technologies that would, *prima facie*, benefit humanity’s estate (to use Bacon’s phrase). For, while Cherlow advances the ethic that one “may not become a creator himself” – essentially echoing the voice of the Golem in our mystical Midrash – one cannot ignore the fact that “many medieval and premodern sources did not ‘listen’ to the voice of the Golem” (Idel 2019: 391). Accordingly, starting just around the time of the Scientific

Revolution (16c.), stories of creating actual Golems began to proliferate within Jewish literature (see, e.g., Scholem 1969: 198; Idel 2019, Pt. 4).

On the one hand, the mere discussion of making actual humanoids, notwithstanding the fact that no one succeeded in making one with human-level consciousness, certainly seems to sanction the attempts. On the other hand, the fact that these Golems largely ran amok seems to call into question the propriety of such endeavors. As a result, these Golem stories are brought in ethical discussions as cautionary tales, neither condoning nor condemning similar (albeit technologically based) undertakings (see e.g., Sherwin 2007; Zoloth 2008; Rubin 2013; Ambrus 2020; Goltz, Zeleznikow, and Dowdeswell 2020; Vudka 2020).

Rava's Gavra

There is, however, a more significant source text that can provide categorical direction. That is to say, the Golem stories mentioned so far – whether of the post Scientific Revolution genre or of the mystical Midrash genre from centuries prior – all fall into the category of Kabbalah (Jewish Mysticism). Accordingly, though mystical texts are considered to be important works within the corpus of Jewish literature, at times even taken into account in both ethical and legal discussions (see, e.g., Tzitz Eliezer 21:5), their weight is limited in the face of the foundational texts of the Bible, Midrash and Talmud that ground Jewish Thought.²⁹

And here it is critical to understand the distinctions between these texts and literary genres in order to judge the significance of the ideas being considered according to their source. For, while in general an idea should be judged by its merit, in Jewish thought there is a hierarchy of values. This is because the Bible is considered divine in origin and, though open to interpretation, cannot be easily overridden by human reason.³⁰ Following the Bible in import, the Talmud is considered to be the suppository of divinely

²⁹ Indeed, Bleich (1998: 60) notes that the Jeremiah Midrash is not, in and of itself, to be regarded halachically.

³⁰ This is not to say that the Bible contradicts human reason, but that human reason is limited in comparison to that of the divine (see, e.g., Navon 2014).

inspired teachings and, though filled with discussions and debates, is not easily contested.³¹

And this brings us to Midrash. First, we must distinguish between the “loose” versus “strict” use of the term “Midrash.” In its loose sense, the term applies to any extra-Biblical exposition on Biblical themes – the story of Jeremiah and Ben Sira being a case in point. In its strict sense, the term refers quite specifically to the extra-Biblical expositions on Biblical themes written from the time of Ezra through the eleventh-century (Epstein 1983: xviii) and restricted to specific rabbinic compilations, e.g., Midrash Rabbah, Midrash Tanchuma, Pirkei DeRebbi Eliezer, etc. It is these rabbinic compilations, as opposed to other “Midrashim,” that carry the most significant weight in Jewish Thought.

But even here there is another distinction to be made, and that is between Midrashim found within the Talmud and those found without. The Talmudic Midrashim are generally referred to as “Aggadot” and held to imply greater import than extra-Talmudic Midrashim (Bernstein 2015). Accordingly, R. Asher Ben Yehiel (medieval Talmudic scholar) writes that anything found in the Talmud has legal status (Ned. 9:2). That said, R. Yehezkel Landau (eighteenth century Talmudic scholar) makes no such distinction, writing: “When it comes to Midrash and Aggadot, their primary intention was to impart ethical lessons, through allusion and allegory, and indeed all of these constitute fundamentals of our religion” (Noda BeYehuda, Tinyana YD 161). As such, while all Midrashim carry ethical import, Aggadot may carry legal import – making the ethical values they express, enforceable.³²

With this introduction understood, we can now turn to the seminal Aggadah describing, what Idel (2019: 27) calls, “The most influential passage treating the possibility to create an artificial human being”:

³¹ See, e.g., R. Avraham Yeshaya Karelitz (Kovetz Igrot Chazon Ish, Vol. II, Ch. 24).

³² Note: while the discussion surrounding the relationship between ethics and law, be it secular or Jewish approaches, is beyond the scope here, it is important to note that legislation of a moral value makes the value enforceable – entailing a right if positive, a penalty if negative. On the Jewish approach to law and ethics see Ch. 4 “Polemics on Perfection.”

Rava said: If the righteous wished, they could create a world, for it is written, “But your iniquities have separated between you and your God” (Is. 59:2). Rava created a man (*Rava Bara Gavra*) and sent him to R. Zeira. He [i.e., R. Zeira] spoke to him but received no answer. Then [R. Zeira] said: “You are from my pietist friends.³³ Return to your dust” (San. 65b).

What are we to make of this text?

To begin, Rava’s claim that “the righteous could create a world” is wholly unique, novel and without precedent, in short: a hypothesis. To support, or perhaps attempt to prove, the veracity of his hypothesis, he brings an axiom in the form of a biblical text which teaches that what separates humans from God, *the Creator*, are iniquities. Rava’s logic being that if one were free of iniquity then one would be like God, *the Creator*, and hence also be able to create (see, e.g., Maharal, *Hidushei Aggada*, ad loc., s.v. *ee ba’u*).

Having proven his hypothesis, at least theoretically, Rava is now ready to demonstrate its truth, empirically, by attempting to create a human being.³⁴ In this he apparently succeeds, as the text attests: “Rava Bara Gavra.” But Rava is not yet satisfied and seeks to put his creation to the test. He thus sends the Gavra to his longtime friend R. Zeira, who, in a test prefiguring the now famous “Turing Test” (Turing 1950),³⁵ attempts to

³³ Lit. *harraya*, variously translated as “friends” (see, e.g., Rashi; *Yad Rama* ad loc.) but indicating the group of sages that could be referred to as “pietists” (see, e.g., Maharsha ad loc.; Steinsaltz ad loc.; Idel 2019: 27). The suggestion that the term refers to “magicians” is rejected by Idel (*ibid.*).

³⁴ See, e.g., *Sidrei Taharot* (Ohalot 5), Aruch LeNer (San. 65b) who explain that Rava made a Gavra to demonstrate his claim. The word “world” is ambiguous and can be understood literally as “world” (see, e.g., Idel 2019: 31) or figuratively as “human” (see, e.g., Idel 2019: 110). Indeed, there is a notion that a human comprises or reflects the whole of creation (Idel 2019: 353).

³⁵ For the sake of completeness, Turing proposed that if a machine could hold a conversation with a human without being discerned as a machine it could be said that the machine is intelligent. The “intelligence” he aimed at was clearly intelligence supported by ZOPC (see, e.g., Penrose 1991: 9; Oppy and Dowe 2021: 2.4). Nevertheless, he realized his test could not verify subjective experience (Turing 1950: 447) and was thus expressing a kind of behaviorist approach – i.e., the most we can hope to observe from the test is behavior consistent with human-like intelligence, not the underlying causes of the behavior (see,

engage Rava's Gavra in conversation. He does so because speech has been considered, long before Turing made it the centerpiece of his test, the distinguishing feature of human intelligence (see, e.g., Aristotle, *Politics* I:II; Descartes, *Discourse* V; Hobbes, *Leviathan* 1:4). Indeed, as noted above, it is the potential to speak, what modern thinkers refer to as "inner speech," that is understood to be the expression of second-order phenomenal consciousness, or in religious terminology, the soul (*neshama*).³⁶

Today, following Artificial Intelligence's success with Large Language Models (see, e.g., Bender et al. 2021, Wiggers 2022), it is widely acknowledged that such a test is not enough to determine second-order phenomenal consciousness (see, e.g., Churchland and Churchland 1990: 37; Grau 2011: 458; Lumberras 2018: 163; Haikonen 2019: 194; Brand 2020: 213). Consequently, many a modified "Turing Test" have been devised to determine consciousness (see catalogs in, e.g., Hales 2014: ch. 12; Elamrani and Yampolskiy 2019; Haikonen 2019: 194-200). Yet even these are found wanting, leaving many to maintain that there is simply no conclusive test to determine consciousness (see, e.g., Putman 1964: 690; Nagel 1974: 436; Kurzweil 2006: 78; Bringsjord 2010: 309).

But if this is true, how can we explain R. Zeira's conversation-based test? I suggest there are two possibilities. One is that R. Zeira was somehow, supernaturally, able to use the notion of conversation to determine not simply if it could speak (for killing a mute is murder),³⁷ but if it had "the potential to speak" (*koach hadibbur*). Alternatively, R. Zeira already knew, as maintains R. Gershon Chanoch Leiner (nineteenth century Chasidic Rebbi in his *Sidrei Taharot*, *Ohalot* 5), that the Gavra was conjured by his friends – the conversation attempt being a simple verification before casting his spell "back to your dust" (which, by the way, would have done nothing to a real human – Weiss 2003, *Noah* 12:2).

e.g., Allen, Varner, and Zinser 2000: 254; Tallis 2012: 196; Parthemore and Whitby 2013: 12; Sloman 2013; Floridi 2014: 189; Soraker 2014: 34; Gerdes and Ohrstrom 2015: 99; Chalmers 2018; Andreotta 2021: fn. 7).

³⁶ For a discussion on relationship between speech and soul, see Ch. 3 "To Make a Mind."

³⁷ See, e.g., *Mishna Berura* (*Biur Halacha* 329, s.v. *ela*); *Sheilat Yavetz* (2:82); *Sidrei Taharot* (*Ohalot* 5).

Either way, Rava's Gavra failed. It was not, after all, the 2OPC being that Rava had set out to create but a 1OPC being, "an animal in the form of a man."³⁸ There are two prominent possibilities for the failure:

- (1) It could be that Rava wasn't the righteous *tzaddik* he had hoped he was but rather still had some iniquity separating him from God. This is the position of the Bahir (#196) as well as commentators like Leiner (Sidrei Taharot, Ohalot 5) who writes that while "there was no one so great as Rava, nevertheless he did not reach the level ... of being utterly free of sin."³⁹ It is this reading that provided the impetus for later Golem attempts recorded in the mystical literature (see, e.g., Idel 2019: 106-107).
- (2) Alternatively, it could be that the hypothesis is simply false – i.e., the righteous cannot, if they want, create a human being. Such is the position of R. Moshe Cordovero (sixteenth century kabbalist), who writes incredulously, "How could one even imagine that it is possible to bring down a soul (i.e., "*neshamah, nefesh and ruach*") in to such a body?!" (Pardes Rimonim 24:10).⁴⁰ Accordingly, Rava's support verse – i.e., what separates man from God is iniquity – does not come teach that man could become Godlike, a creator, but only that man could connect to God in a spiritual sense (*deveikut*). This is the position of the R. Judah Loew ben Bezalel (sixteenth century Talmudic scholar, a.k.a., "Maharal") who explains:

Rava ... purified himself, performed the procedures of the Book of Creation (Sefer Yetzirah), connected [spiritually] to God and created a Gavra. But it had

³⁸ R. Yaakov Emden (Sheilat Yavetz 2:82); R. Chaim Azulai (Marit HaAyin, San. 65b, s.v. *Rava*); R. Leiner (Sidrei Taharot, Ohalot 5). Interestingly, the deficiency of Rava's Gavra is noted by R. Aryeh Kaplan who explains, in his commentary to Sefer Yetzirah, that the numerical value of "Rava Bara Gavra" is 612, one less than the 613 typological limbs and veins of a full humans (Kaplan 1997: xxi). For completeness, the anomalous position of R. Tzadok MiLublin (Divrei Halomot 6) should be noted: the Gavra is above an animal, having 2OPC, but without a *neshamah*.

³⁹ Similarly, R. Isaac the Blind (Scholem 1969: 193), R. Abraham Yagel (Idel 2019: 216), R. Natan (Idel 2019: 106), R. Jacob Ettlinger (Aruch LeNer, San. 65b), R. Bleich (1998: 58), Pseudo-Saadya Gaon (Sefer Yetzirah 2:5).

⁴⁰ R. Cordovero's position may be viewed as extreme only in the sense that he does not believe any spiritual essence can be imbued, as opposed to others in this camp who hold that only the highest-level human soul (*neshamah*) cannot be imbued.

not speech for [Rava] had not the power to bring a speaking soul into a man to make a being like himself. And this is obvious, since how can a man create a being like himself, when it is impossible for God, Who is above all, to create a being like Himself!⁴¹

Now, while one may question the soundness of the statement that “a being can’t create one like himself,” indeed Bleich calls it “curious” (1998: 81 fn. 58), nevertheless, the claim that man “has not the power to bring a speaking soul into a man,” does have broad consensus.⁴² R. Chaim Yosef David Azulai (eighteenth century Talmudic scholar), for example, justifies this claim based on the creation verse (Gen. 2:7) as follows: “The potential (*koach*) for intelligence and speech is from God alone, as it says, ‘and [God] breathed into his nostrils the breath of life (*nishmat hayim*)’.”⁴³ Accepting this, R. Meir Abulafia (medieval Talmudic scholar) defends Rava’s original claim by suggesting that, while man cannot draw the soul, God will grant the soul to the Golem provided its creator is perfectly righteous.⁴⁴

⁴¹ Maharal (ad loc., s.v. *rava*). It is important to note a seeming contradiction in the Maharal’s commentary. On Rava’s hypothesis (ibid, s.v. *ee’ban*), he explains that it is entirely *possible* for the righteous to create a human being, yet here on the test of the hypothesis (ibid, s.v. *rava bara*) he writes that it is entirely *impossible*. The contradiction can be reconciled by understanding that the first comment explains the hypothesis, as is; while the second comment explains why the hypothesis failed – precisely as I am suggesting in my reading of the passage.

⁴² In defense of the Maharal, his point is raised as a possibility by a modern robotics professor: “there could be a fundamental theorem of the universe ... that no creature is smart enough to build a copy of itself” (Richardson 2015: 55).

⁴³ Marit HaAyin (San. 65b, s.v. *rava*). Similarly, Maharsha (Hidushei Aggada, ad loc., s.v. *v’lo hava*), R. Tzadok MiLublin (Divrei Halomot 6, s.v. *af shekatan*), R. Joseph Ashkenazi (Idel 2019: 71), R. Pinhas Eliyahu Horwitz (ibid: 238), R. Elazar of Worms (ibid.: 55). R. Bleich (1998: 81 fn. 58) understands the Hesed le-Avraham (Ein Yaakov, Mayan 4, Nahar 30) to agree that man cannot draw a soul to a Golem, but this only because of the limited power of Sefer Yetzirah, thus perhaps leaving open modern attempts. Rosenfeld (1977: 61) also makes this conjecture.

⁴⁴ Yad Rama (San. 65b, s.v. *amar*). So too R. Hannanel (San. 67a), R. Weiss (2003, Vaera 9:3). In this vein, it is important to note a rabbinic Midrash that teaches, “if all the creatures in the world gathered together to make a single gnat and put a soul (*neshamah*) into it, they would not succeed” (Gen. R. 39:14). R. Kaplan (1997: xx) explains that this statement is not denying the theoretical possibility but only the practical one due to the loss of knowledge. R. Weiss (2003, Vaera 9:3) understands this to mean, quite simply, that man cannot create synthetic life (*yesh mi’ayin*) without divine intervention. Interestingly, Turing makes this very

And that brings us back to Rava's claim. Here, against the understanding of a number of modern readers,⁴⁵ I suggest that this Aggadah was canonized in order to lay to rest the hypothesis that a human can create a 2OPC humanoid.⁴⁶ Indeed, its refutation is inherent in its very claim. The hypothesis, "if the righteous wanted ...," is explained by the great medieval biblical commentator, R. Shlomo Yitzhaki (Rashi), to mean that said "righteous" individual not be simply "righteous," but "completely free of sin" (Rashi, ad loc.) Yet we know from Ecclesiastes that "there is not a righteous man upon earth, that doeth good, and sinneth not" (7:20).⁴⁷ And this is borne out by all the failed attempts to make a human-level Golem brought in the mystical literature,⁴⁸ the singular success of Jeremiah and Ben Sira coming not to affirm the possibility but admonish against it – i.e., even if you could do it, you shouldn't.⁴⁹

claim in arguing for the possibility of a conscious machine: "We are ... instruments of His will providing mansions for the souls that He creates" (Turing 1950: 443).

⁴⁵ Rosenfeld 1966: 26; Rosenfeld 1977: 61; Liebes 1991; Bleich 1998: 58; Rakover 2002: 113; Sherwin 2007: 137.

⁴⁶ It should be noted that Peter Schafer (1995: 253) makes this claim but brings no support to defend it. Idel (2019: 31, 413) argues, from a literary perspective, that the passage is a Talmudic polemic against creating a sentient being (albeit by magic not science) and teaching that endowing a body with a soul (*neshamah*) is impossible (ibid.: 17).

⁴⁷ R. Leiner (Sidrei Taharot, Ohalot 5) makes precisely this same argument using Ecclesiastes (7:20), but nevertheless leaves open the possibility, given that there were four saintly individuals recorded in the Gemara (BB 17a). This, however, begs the question (noted by Aruch LeNer, San. 65b, s.v. *ee ban*): if the possibility truly existed for these saintly four, why did they not indeed actualize this ultimate potential and create a world?!

⁴⁸ See, e.g., R. Yaakov Emden (Sheilat Yavetz 2:82); Scholem (1969: 202-203).

⁴⁹ It should be mentioned here that R. Isaac ben Samuel of Acre, as part of the school of Ecstatic Kabbalah, reread the Jeremiah midrash as not only affirming the *possibility* of creating a full-fledged human-level Golem but *permitting* it (Idel 2019: 346-352). That said, such sentiments must be understood in their mystical context – i.e., the endeavor to cleave to God by mystically imitating God's creative act. Accordingly, as R. Isaac never writes that he succeeded in making an actual Golem, we too should not be moved to include his account in our normative analysis. In addition, Abraham Abulafia, founder of the Ecstatic school of thought, absolutely forbid creating a Golem (Idel 2019: 345), seeing it as purely a mystical endeavor (ibid.: 102).

And it is precisely this understanding that brings R. Zeira to send it back to its origins. I propose that R. Zeira, in administering his language test to the Gavra, is not checking for second-order phenomenal consciousness in order to then accept it into the family of persons. Rather, he wants to know whether, if it fails the test, he can eliminate it with impunity like any dangerous animal; or, if it passes the test, he will have to express his indignation, once again, at the “shenanigans” of his old friend Rava who had once killed him and brought him back to life (Meg. 7b), much like the life he has now given to the Gavra.

R. Zeira’s position, then, is clear: it is categorically forbidden to make a synthetic sentient being in the form of a human, whether it has first-order or second-order consciousness. Furthermore, I suggest that this is precisely the position of the Gemara itself. For, while interpreting a narrative is more of an art than a science, the evidence presented here more than reasonably supports reading the Aggadah as favoring R. Zeira over Rava. Indeed, if Rava’s position was to be considered ascendant, it would be reasonable to expect that he should have been cited as resurrecting his Gavra (just as he had done with R. Zeira [Meg. 7b]). The very fact that Rava’s actions are defeated and never returned to again – not here, nor anywhere else in the Gemara – surely bodes ill for his position.

Some may argue that R. Zeira’s position is not the last word on the issue, as there is a continuation of the Aggadah:

R. Hanina and R. Oshaia spent every Sabbath eve in studying the ‘Book of Creation’, by means of which they created a third-grown calf and ate it.

Here, in contrast to Rava’s Gavra that is eliminated with impunity, the calf created by R. Hanina and R. Oshaia is used for the purpose, holy purpose I might add, for which it was created. The Gemara clearly approves of the creation here, because, besides the fact that the creation of the calf was not a one-off event but repeated “every Sabbath eve,” if the calf was forbidden food it could not rightly be used for the religious purpose of a mitzvah meal with the recitation of blessings upon its consumption (Shulchan Aruch, OH 196:1). Interestingly, it is R. Oshaia himself who makes this ruling (JM Challah 1:5)! And if one might think that the calf itself was kosher but not its creation, the Talmud

goes on to explain, explicitly, that the actions of R. Hanina and R. Oshaia were “categorically permissible” (*mutar le’chatchila*).

Now, while some see this explicit permit to create a calf as approval-by-association to create a Gavra,⁵⁰ I would argue precisely the opposite.⁵¹ For, whereas the Talmud tells the two stories together in succession (65b), it is only after four pages (67b) that it reintroduces the calf story to give it its *hechsber* (stamp of religious approval).⁵² The separation is telling. It graphically demonstrates the fact that permitting the synthetic creation of an animal is very far from permitting the synthetic creation of a human. Indeed, to realize this one need look no further than to the gravity of the issues surrounding the cloning of animals versus those of the cloning of humans (see, e.g., Kass 2014). The potential ethical dilemmas inherent in the former pale, radically, in comparison to those inherent in the latter.

Conclusion

In conclusion, I have argued herein that if technology could achieve the creation of a 2OPC being, regardless of substrate (i.e., silicon, carbon, other), such technology should be banned. And, more restrictively, given the great epistemological concern that, while many an enhanced Turing Test has been proposed, no one has come up with a “R. Zeira (Supernatural) Test” that would ensure we know if a being is conscious or not, I herein endorse the calls to ban even the attempt to develop sentient robots.⁵³ This is in

⁵⁰ Aruch LeNer (San. 65b); Bleich 1998: 58; Rakover 2002: 113; Weiss 2003, Vaera 10:2.

⁵¹ R. Shem Tov ibn Shaprut (Pardes Rimmonim 13a) also sees the two stories as contrasts between permitted and prohibited.

⁵² For the sake of completeness, the technical reason for the separation is that the discussion of permitted versus forbidden creations was not yet raised until four pages later. That said, if the creation of a Gavra (i.e., human) was permissible, it should have been cited explicitly alongside the calf, just as calf and Gavra were cited together originally. For one cannot infer a permit to create a Gavra from a permit to create a calf, the difference between human and animal being too vast, indeed, as vast as the difference between 2OPC and 1OPC.

⁵³ See, e.g., McCarthy 1999: 15 cited in Gunkel 2018: 112; Metzinger 2003: 622; Gamez 2008: 906; Grau 2011: 457; Yampolskiy 2013: 392; Miller 2017; Schwitzgebel and Garza 2020: 466. For sources on banning conscious robots see above fn. 18.

opposition to some who say that we should work to make such a being but then desist (e.g., Atlan in Idel 1996: 28).

To support this position, I brought the Golem, in its various forms, to serve as a moral paradigm. Beginning with Jeremiah's Golem, who condemned the creation of a synthetic human for fear that such will lead humanity to declare that "God is dead," I set this source aside for, in our era awash in atheism, the apparent demise of God is old news and, as such, creating synthetic humans today will not significantly change humanity's faith in God.

But here it is important to note that the theological concern voiced by the Golem, in fact, remains relevant. For, if "God is dead," then, as Dostoyevsky wrote, "all is permitted."⁵⁴ That is to say, the Golem's concern – that humanity will leave God and follow man – was not merely a petty concern over power in a physical sense (i.e., man will create just like God), but a deep concern over power in a metaphysical sense (i.e., man will assume moral authority instead of God). It is stunning just how prescient this ninth-century Golem was, for indeed, both concerns were realized as the Scientific Revolution gave birth to what might be called the moral revolution wherein Kant, for the first time in history, shifted moral authority from God to man.⁵⁵ And it was this shift that ultimately brought about the observation that "all is permitted," that the world has lost its absolute moral ground.⁵⁶ The Golem's admonition, then, while mute as a call to stop science and technology, remains in all its vigor as a call to return to the original Architect; for even if humans can build the same building, they cannot provide it the same ground.

This point was made contemporary by Bill Joy, former chief scientist of Sun Microsystems, who noted that Nietzsche followed his famous declaration with the warning that there is a danger in substituting science for God. And this because science pushes forward despite the "disutility" – i.e., despite the detriments it may entail for humanity (Nietzsche [1887] 2001: 201). Accordingly, Joy expresses great concern over the "control" problem: "The truth that science seeks can certainly be considered a

⁵⁴ Dostoyevsky [1880] 2019. See esp. Jean-Paul Sartre ([1947] 2007: 28-29).

⁵⁵ See, e.g., Verbeek 2011: 12. Calverley (2011: 214) explains that the shift came with the Enlightenment.

⁵⁶ See, e.g., Berkovits 2004: 106.

dangerous substitute for God if it is likely to lead to our extinction” (Joy 2000). Joy calls for a ban of AI technologies – like conscious robots – that have the potential to extinct humanity. In this he is like the prophet Jeremiah, who forewarns of the destruction to come as a result of shirking divine morality and “worshipping the work of their own hands” (Jer. 1:16). Perhaps it is for this very reason that the mystical Midrash is told in Jeremiah’s name. And how apt for us today are Jeremiah’s words at the end of that Midrash: “it is worthwhile to study these matters for the sake of knowing the power and dynamis of the creator of the world, but not in order to do [them]. You shall study them in order to comprehend and teach.”

Following the analysis of Jeremiah and his Golem, I demonstrated that the position they voiced took on more normative significance in the Aggadah of Rava’s Gavra. Here I argued that this narrative, when looked at as an organic whole, comes to prohibit the making of a synthetic conscious being.⁵⁷ But perhaps I should be more exact and claim that the Aggadah comes to voice the notion that, while it is actually impossible to make a second-order phenomenally conscious being (i.e., ensouled with *neshamah*), making even a first-order phenomenally conscious being (i.e., ensouled with *nefesh behemit*) that looks like a human, although possible, is forbidden. Now, irrespective of whether such creations (1OPC or 2OPC) are possible, the ethical directive is unambiguous: it is forbidden to create sentient beings (1OPC or 2OPC) in human form.⁵⁸ This is, of course, in contradistinction to the explicit permit, made by the Talmud, to create a 1OPC being in the form of an animal. Though even here there is reticence to allow it (see, e.g., Shach YD 179:18).

Now, if my reading of this Talmudic Aggadah is correct, then the prohibition against making a sentient being in human form is of deontological value, it being adduced neither from consequences nor virtues but from the actions of great teachers – in

⁵⁷ Noteworthy is the fact that even R. Bleich (1998: 75), who didn’t read the Aggadah as prohibitive, did acknowledge that neither does it “encourage” such creation.

⁵⁸ Note: the position that artificially intelligent synthetic beings (with or without consciousness) should not be made in human form has been voiced in both philosophical and legal contexts. See, e.g., Bryson (2020: 22) who notes this is the position of the UK and OECD; see also Johnson and Verdicchio (2018: 299).

rabbinic parlance, “*maaseh rav*.”⁵⁹ And it is precisely the fact that this teaching is sourced in rabbinic action that gives it not only ethical value but legal value. For, according to the principle of “*halacha al pi maaseh rav*,” the actions of a rabbi “are recounted specifically for purposes of learning halacha from them” (Bernstein 2015: 53).⁶⁰ And indeed, this narrative is brought in numerous legal discussions.⁶¹

Accordingly, this deontological prohibition presents us with a legally binding ethical duty. It is a duty, however, squarely grounded in Jewish law and lore, thus making it incumbent upon those who accept Jewish law and lore. That said, I would argue that it could and, indeed, should apply to all peoples. For the *maaseh rav* is neither contingent on Jewish beliefs (i.e., there is nothing explicitly “Jewish” in the act of R. Zeira), nor is the issue itself a parochial one (i.e., the propriety of creating a sentient humanoid is of concern to all). Accordingly, given that the question is universal, and the response is universal, the duty, then, is universal (Mendelssohn [1782] 2017: 34-36).

Nonetheless, coming as it does from within the corpus of Jewish Thought, the duty could be said to hold of little significance to non-Jews. Yet nothing could be further from the truth. Indeed, the Torah itself commands that it be written in seventy languages (Sotah 7:5) in order to be made relevant to all the typological seventy nations (see, e.g., Rashi Sotah 35b, s.v. *he’ach*). And so it has, as noted most emphatically by sixth American President John Quincy Adams: “The law given from Sinai was a civil and municipal as well as a moral and religious code; ... [its] laws are essential to the existence

⁵⁹ Note that it is entirely irrelevant whether this “*maaseh*” (act) ever actually took place or not, for it is brought as if it did.

⁶⁰ See also, Encyc. Talmudit (“halacha” s.v. *al pi maaseh rav*). Noteworthy is the fact that even without invoking the “*maaseh rav*” principle, the narrative would still have legal weight according to those, like R. Asher Ben Yehiel (Ned. 9:2), who consider any Aggadah to be halachically significant.

⁶¹ On the status of clones, see, e.g., Bleich 1998; Steinberg and Loike 1998; Loike 2000; Rakover 2002; Loike and Tendler 2003. On the status of humanoids, see, e.g., Hacham Tzvi (93); Sheilat Yaavetz (2:82); Lehorot Natan (7:11); Sidrei Taharot (Ohalot 5); Divrei Halomot (6); Hashukei Hemed (San. 65b); Marit HaAyin (San. 65b); Tzafnat Paneach (2:7); Darkei Teshuva (7:11). Birkei Yosef (OH 55:4 s.v. *u'lmai*); Machazik Bracha (ad loc); Ikarei HaDat (OH 3:15); Kaf HaChaim (OH 55:12); Rivevot Efraim (7:385); Weiss (2003, Noah 12:2, Vaera 9:3-5); Shu”t Yehuda Ya’aleh (1:26 s.v. *v’da*); Gilonei Hashas (San. 19b, s.v. *sham maale*).

of men in society, *and most which have been enacted by every nation which ever professed any code of laws.*”⁶² Now, the “law given from Sinai” includes both written and oral, both Torah and Talmud (Ber. 5a), such that, if the written Torah is of value to the world at large, then the Talmud is no less, for it contains the values of the written Torah as applied over time (see, e.g., Berkovits 1983: 83; Eisen 1991: 85).⁶³ Accordingly, I suggest that the Talmudic-based deontological ban on the development of sentient humanoids applies universally to all of humanity.

That said, many argue that imposing a ban on such a tantalizing technology – “the greatest technological achievement of our species” – will be difficult, if not impossible, to enforce.⁶⁴ Indeed, Bleich laments, “It is a truism that, in the usual course of human events, that which *can* be done *will* be done” (1998: 48). This attitude, known as “technological determinism” (see, e.g., Johnson and Miller 2008: 13), has caused great consternation among all who realize that technology, while holding the great promise of creating a world perfect, also holds the great peril of bringing about its demise (see, e.g., Jonas 1984). As a result, ever since the Nuclear Bomb was dropped in the mid-twentieth century, the field of Technology Ethics has been growing with ever greater urgency to guide, and if need be, limit, the development of all our tantalizing technologies (see, e.g., Mitcham and Nissenbaum 1998; Vallor 2016: 28-32). Interestingly, the celebrated first Chief Rabbi of pre-state Israel, Abraham Isaac Hacoen Kook, made this point already at the beginning of the twentieth century:

⁶² Adams 1850: 61 (*emphasis added*). Similarly: “The Bible is the rock on which our republic rests” (Andrew Jackson). “The teachings of the Bible are so interwoven and so entwined with our civic and our social life, it would be impossible for us to figure what life would be if these teachings were removed” (Teddy Roosevelt). “The existence of the Bible, as a book for the people, is the greatest benefit which the human race has ever experienced” (Immanuel Kant, unbound pages in the Konigsberg library [Convolut G. i., ii]). See also Johnson 1987: 585.

⁶³ It should be noted that, while there is a lot of discussion regarding the propriety of teaching non-Jews the oral Torah (see, e.g., Bleich 1983b), R. Bleich writes that “Jews should certainly not hesitate to make the teachings of Judaism, as they bear upon contemporary mores, more readily accessible to fellow citizens” (ibid.: 339).

⁶⁴ See, e.g., Grau 2011: 458; Neely 2014: 105; Hernandez-Orallo 2017: 448; Gunkel 2018: 113; Danaher 2019b: 2032.

Behold, if human abilities will increase mightily⁶⁵ and [yet] human goodwill will not develop according to pure ethics,⁶⁶ will not such of powers, then, benefit only man's material being ... selfish concern overriding moral concern ... leading, perforce, to an onerous life for all?! ... And in contrast, by applying ethics to guide these powers [of ability and will], great good and blessing will be achieved by both the individual and the world. For truly, the complete good will come specifically from the perfect joining of these two forces – the ability and the will – [with the intent] toward the good end. ... And as humanity continues to increase in knowledge, so will it recognize the unity in all the various forces of nature [to the point that]: “If the righteous wanted they could create a world” (*ee ba'u tzadikvei baru alma* - San. 65b). This is the ultimate approach to human endeavor, wherein “ability and will” are unified toward the fulfillment of that sublime goal: the perfection of man (Kook 1906: 24).

Stunningly, Kook encapsulates his approach on how to perfect our world via science and technology in Rava's great claim. How are we to create our world? How are we to reach “our greatest achievement”? Only by joining our “will” (*ba'u*) and our creative “ability” (*baru*) under the guidance of “pure ethics” (*tzadikvei*).⁶⁷ And as ethics is measured not only by what one does, but by what one does not do, in the case of a conscious humanoid (*gavra*), our greatest achievement will be in the ethical act of refraining from that achievement.⁶⁸

In the words of R. Zeira, “Return to your dust!”

⁶⁵ Literally, “sevenfold,” which is a biblical expression for “greatly” (see Lev. 26, esp. Rashbam, Lev. 26:18).

⁶⁶ Note that there can be misguided goodwill! For humanity can believe it is doing great good, yet without reference to a “pure ethic,” do great damage.

⁶⁷ See Rakover 2002: 117.

⁶⁸ For “the golem is less a celebration of human power than a reminder of its limits” (Rubin 2013). Also, Ambrus 2020: 283.

4. Conclusion

The goal of this thesis has been to investigate the moral status of artificial intelligence, especially as embodied in human-like form, but with implications for any system that engages us in a human-like manner. It is an investigation that seeks to answer the ethical questions posed by today's cutting-edge technologies. Indeed, in 2019, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems posed the following quandaries: "Is AI ... a product that can be bought and sold? A domesticated animal with more rights than a simple piece of property, but less than a human? A person? Something new?"¹

To respond, I analyzed the ontological characteristics that constitute a human being and demonstrated that it is second-order phenomenal consciousness (2OPC), what religious people refer to as the human soul, that uniquely defines us.

With this understanding, I divided my investigation into two parts. The first asks, "What is the moral status of artificial intelligence without consciousness (i.e., mindless AI)?" The second asks, "What is the moral status of artificial intelligence with consciousness (i.e., mindful AI)?" This bifurcation led me to define three essential dilemmas: (1) the Mindless Robot Dilemma, which I called the VAD dilemma, that asks, how we can maintain our virtuous disposition in interacting with a mindless machine while simultaneously maintaining our appreciation for authentic relationships; (2) the Mindful Robot Dilemma [1], which asks if we can, ethically, create a human-level conscious being to relieve us of our burdens; and (3) The Mindful Robot Dilemma [2], which asks if it is ethical to create a synthetic human-level conscious being simply to join us in the family of human beings.

Regarding mindless AI, we saw that while it has the moral status of a toaster, it nevertheless demands our moral consideration for the sake of our own moral integrity. Here we encountered thinkers on both extremes of the VAD dilemma – some arguing that we must treat mindless machines for what they are (i.e., mindless machines), others arguing that we must treat them as they appear to us (i.e., humanoid beings with beliefs,

¹ IEEE 2019.

desires and intentions). I demonstrated that both extremes are untenable and thus proposed a middle approach, based on the Virtuous Servant Owner paradigm, suggesting that one can preserve their virtue by maintaining ethical interactions with any being that presents as human-like, but nevertheless maintain the emotional/social distance that one would keep with a slave. In so doing, the moral integrity of both the individual and society as a whole is preserved.

As for mindful AI, I showed that, unlike its mindless cousin, human-level consciousness grants it the moral status of a human being – i.e., personhood. To be clear, personhood is a subset of humanhood that recognizes an entity to possess the moral status of a human being, despite not being a human being. The standard, or ontological, approach to determining moral personhood, as noted in Ch. 2: The Virtuous Servant Owner (sec. The Dilemma), asks “what the entity ‘is’ [i.e., its salient ontology] in order to determine how we ‘ought’ [i.e., morally] to treat it.” In Ch. 3: To Make a Mind (sec. Consciousness), I compiled a great many criteria put forth to determine humanhood, ultimately demonstrating that it is “consciousness [i.e., 2OPC] that is arguably *the* defining ontological feature of human beings.” Thus, an entity possessing 2OPC, despite lacking many other human traits (such as being born of a human mother or the ability to reproduce with a human), would still be accorded personhood. I then explained (*ibid.*, sec. Machine Consciousness) that it is this understanding of personhood that is driving the field of machine consciousness in its quest for “the holy grail of artificial intelligence” – a conscious machine (as noted in Ch. 5: Eudemonia of a Machine, Introduction).

Regarding the moral dilemma of manufacturing such beings to serve us, I brought to bear the classic moral approaches to show that making such beings is beyond the pale. And while my research also bore new arguments from within these moral approaches, it was my use of Maimonides’ “Jeremiah Glory of Man” paradigm that uniquely, and in my opinion unambiguously, argued against the manufacture of mindful AI to serve. Moreover, beyond the inherent injustice of enslaving a conscious humanoid, the ill effects upon the slave-owners themselves, as evidenced by Frederick Douglass (Ch. 2: The Virtuous Servant Owner, sec. VSO), and by extension, an entire society participating in or even condoning slavery, as articulated by G. W. F. Hegel, Alexis de Tocqueville, and Harriet Ann Jacobs (see Gunkel 2018: 127-129), cannot be overstated.

Having rejected the possibility of creating mindful AI to serve us, I addressed the moral dilemma of creating mindful AI simply to live in fellowship with us. Here, to avoid the speculative nature of applying consequentialist and virtue ethics to a non-existent entity, I made a deontological argument based on the Golem paradigm, in particular, the Talmudic Gavra. I argued that the prohibition of creating mindful AI, learned as it is from the Talmudic recounting of a rabbinic action (*maaseh rav*), is a wholly deontological ethic that is legally binding.

That said, it is worth emphasizing the role virtue plays within a “wholly” deontological approach. As I explained (Ch. 1, Ch. 2, Ch. 4), both Maimonides and Kant can be shown to value virtue though their moral outlooks are strongly deontological. As for Maimonides, Wurzbarger articulates his synthesis of virtue and deontology as follows:

“Jewish piety involves more than meticulous adherence to the various rules and norms of religious law; it also demands the cultivation of an ethical personality... We are commanded to engage in a never-ending quest for moral perfection, which transcends the requirements of an ‘ethics of obedience’...[T]he halakhic [i.e., Jewish legal] system serves merely as the foundation of Jewish piety” (1994: 3-4; see also *ibid.* Ch. 5).

Kant’s value of virtue is explained in a similar manner by Loudon:

“Kant’s notion of action *aus Pflicht* [“out of duty”] means in the most fundamental sense not that one performs a specific act for the sake of a specific rule which prescribes it ... but rather that one strives for a way of life in which all of one’s acts are a manifestation of a character which is in harmony with moral law. Action *aus Pflicht* is action motivated by virtue, albeit virtue in Kant’s sternly rationalist sense [i.e., acting virtuously from rational understanding]” (1986: 485-486).

Accordingly, it is the very adherence to the deontological that serves to fashion the virtuous personality. In Loudon’s words: “The virtuous agent is one who consistently ‘follows the rules’ out of respect for the idea of rationally legislated law. But ‘the rules’, while they do serve as action-guides, are intended most fundamentally as life-guides”

(ibid.: 479). Similarly, Nancy Sherman elucidates Kant's view that virtue not only affords the individual with the fortitude to adhere to deontological principles, but also cultivates within the individual the desire to adhere to them, ultimately leading to the individual's perfection (1997: 136f).

My analysis brought to bear thought from all corners of moral philosophy but was ultimately based on Jewish philosophy. For it is my sincere belief that Jewish philosophy has much to contribute to the ongoing conversation over the ethical concerns arising from new technologies in general and artificial intelligence in particular. This does not mean that Jewish philosophy has "the answers" to every moral dilemma, but that it provides a unique prism through which every moral dilemma can be reasoned and resolved.² It is a prism that was handed to Moses on Sinai and forged in the crucible of Jewish history. For believers, the moral imperatives derived through this prism are grounded in the objectivity of divine will. Yet even for those who do not believe, the moral imperatives so derived are grounded in a 3300-year-old tradition that has grappled with moral dilemmas, not simply as thought experiments, but as living dilemmas that are often, as R. Yaakov Medan aptly puts it, "written in blood."

The divide between believers and non-believers in how they approach philosophical arguments rooted in religious texts can be best understood through the contrasting approaches of Bertrand Russell (1912) and Margaret Macdonald (1953). For believers, arguments derived from religious texts lead to conclusions that are regarded as objectively true, based as they are in the word of God. This aligns with Russell's scientific approach to philosophy that aims to uncover objective truths about reality through rigorous reasoning and empirical evidence. Non-believers, on the other hand, can find the arguments from religious texts compelling, if devoid of objective truths, as narratives that illuminate aspects of possible lived experiences, akin to how MacDonald viewed philosophical theories. "For Macdonald, philosophy's value lies not in providing us with new facts about the world, but rather in helping to see the familiar in a new light, in drawing attention to features of experience that might ordinarily pass us by, and by

² By "resolved" I do not imply that dilemmas do not remain dilemmas, but that we walk away with an understanding of the issues, as well as clear directives for action.

providing us with stories that can help make better sense of the world around us” (Warburton 2024).

It is in this sense that my use of the prism of Jewish philosophy to analyze the modern dilemmas posed by artificial intelligence will, it is hoped, shed new light on the familiar dilemmas of AI that are becoming more acute with every passing day. It has been an analysis that not only sought to clarify the specific technological dilemmas described, but the very human dilemmas implicit in them: what is it to be human and how are we to be human.³ These questions were addressed in various ways, but one cannot help but notice that the prophet Jeremiah appears in this context throughout my thesis. He is the orator of the human ideal as brought by Maimonides in my chapters, *Finding Virtue in a Law on Slaves*, *Polemics on Perfection*, *Eudemonia of a Machine*, and he is the voice of reason and restraint when it comes to venturing over the line in our creative activities to fix the world, as seen in my chapter, *Let Us Make Man in Our Image*. Perhaps it is then fitting to conclude with his words:

Thus saith the Lord: Let not the wise man glory in his wisdom, neither let the mighty man glory in his might, let not the rich man glory in his riches; but let him that glorieth glory in this, that he understandeth, and knoweth Me, that I am the Lord who exercise mercy, justice, and righteousness, in the earth; for in these things I delight, saith the Lord.

³ Not a few researchers in the field have noted that AI has brought us to ask these great existential and ethical questions (see, e.g., McCorduck ([1979] 2004: 244); Kaiser (1989); Herzfeld (2002a: 5); Guizzo 2010; Turkle 2011a: 30; Ambrus (2020: 288-9); Bryson (2020b: 23: 63); Coeckelbergh (2020b: 9); Kingwell (2020: 339-40)).

5. Bibliography

- Abbott, Ryan Benjamin, and Elizabeth Shubov. 2023. "The Revolution Has Arrived: AI Authorship and Copyright Law." *Florida Law Review* 75 (6). <https://doi.org/10.2139/ssrn.4185327>.
- Adamatzky, Andrew. 2016. "Twenty Five Uses of Slime Mould in Electronics and Computing." *Int. Journ. Of Unconventional Computing* 11: 449–71. https://www.researchgate.net/publication/299462623_Twenty_five_uses_of_slime_mould_in_electronics_and_computing_Survey.
- Adams, John. 1809. "Letter from John Adams to François Adriaan van Der Kemp." Founders Online. February 16, 1809. <https://founders.archives.gov/documents/Adams/99-02-02-5302>.
- Adams, John Quincy. 1850. *Letters of John Quincy Adams to His Son, on the Bible and Its Teachings*. Auburn: James M. Alden. <https://archive.org/download/lettersofjohnqui00adam/lettersofjohnqui00adam.pdf>.
- Agar, Nicholas. 2019. "How to Treat Machines That Might Have Minds." *Philosophy & Technology* 33 (2): 269–82. <https://doi.org/10.1007/s13347-019-00357-8>.
- Aiken, Lisa. 2009. *The Baal Teshuva Survival Guide*. Beverly Hills, CA: Rossi Publications.
- Allen, Colin, Gary Varner, and Jason Zinser. 2000. "Prolegomena to Any Future Artificial Moral Agent." *Journal of Experimental & Theoretical Artificial Intelligence* 12 (3): 251–61. <https://doi.org/10.1080/09528130050111428>.
- Altmann, Alexander. 1972. "Maimonides' Four Perfections." In *Israel Oriental Studies II*, 15–24. Tel Aviv: Tel Aviv University.
- Ambrus, Gabor. 2020. "Image, Servitude, Partnership." In *Artificial Intelligence: Reflections in Philosophy, Theology, and the Social Sciences*, edited by Benedikt Paul Gocke and Astrid Rosenthal-Von der Putten. Boston: Brill.
- Amital, Yehuda. 2002. "Perfecting Nature." *VBM*. Gush Etzion: Yeshivat Har Etzion. <https://www.etzion.org.il/en/tanakh/torah/sefer-vayikra/parashat-tazria/perfecting-nature>.
- Anderson, David L. 2013. "Machine Intentionality, the Moral Status of Machines, and the Composition Problem." In *Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemology and Rational Ethics*, edited by Vincent C. Müller, 321–34. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-31674->

6.

- Anderson, Micheal, and Susan Leigh Anderson, eds. 2011. *Machine Ethics*. New York: Cambridge University Press.
- Anderson, Susan Leigh. 2011a. "Philosophical Concerns with Machine Ethics." In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- . 2011b. "The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics." In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Anderson, Susan Leigh, and Michael Anderson. 2014. "Towards a Principle-Based Healthcare Agent." *Machine Medical Ethics* 74 (September): 67–77. https://doi.org/10.1007/978-3-319-08108-3_5.
- Andreotta, Adam J. 2021. "The Hard Problem of AI Rights." *AI & SOCIETY* 36. <https://doi.org/10.1007/s00146-020-00997-x>.
- Annas, Julia. 2008. *Plato's Ethics*. *Oxford Handbooks Online*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195182903.003.0011>.
- Anthes, Gary. 2001. "Computer Consciousness." *Computerworld*, November 5, 2001. <https://www.computerworld.com/article/2584729/computer-consciousness.html>.
- Arico, Adam, Brain Fiala, Robert F. Goldberg, and Shaun Nichols. 2011. "The Folk Psychology of Consciousness." *Mind & Language* 26 (3): 327–52. <https://doi.org/10.1111/j.1468-0017.2011.01420.x>.
- Aristotle. 2004. *Nicomachean Ethics*. Translated by Roger Crisp. Cambridge: Cambridge University Press.
- . 2013. *Politics*. Translated by Carnes Lord. 2nd ed. University Of Chicago Press.
- Arkin, Ronald C. 2010. "The Case for Ethical Autonomy in Unmanned Systems." *Journal of Military Ethics* 9 (4): 332–41. <https://doi.org/10.1080/15027570.2010.536402>.
- Asaro, Peter M. 2006. "What Should We Want from a Robot Ethic?" *International Review of Information Ethics* 6 (12): 9–16. <https://peterasaro.org/writing/Asaro%20IRIE.pdf>.
- Baars, Bernard J. 2019. *On Consciousness: Science & Subjectivity - Updated Works on Global Workspace Theory*. Nautilus Press.
- Bar-Asher Siegal, Michal, and Avi Shmidman. 2018. "Reconstruction of the Mekhilta Deuteronomy Using Philological and Computational Tools." *Journal of Ancient*

- Judaism* 9 (1): 2–25. <https://doi.org/10.30965/21967954-00901002>.
- Barresi, John, and Raymond Martin. 2012. *Naturalization of the Soul*. Routledge.
- Barron, Andrew B., and Colin Klein. 2016. “What Insects Can Tell Us about the Origins of Consciousness.” *Proceedings of the National Academy of Sciences* 113 (18): 4900–4908. <https://doi.org/10.1073/pnas.1520084113>.
- Basl, John. 2014. “What to Do about Artificial Consciousness.” In *Ethics and Emerging Technologies*, edited by Ronald L. Sandler. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137349088>.
- Beasley, Yaakov. 2019. “The Morality of Slavery.” Yeshivat Har Etzion. 2019. <https://www.etzion.org.il/en/tanakh/torah/sefer-shemot/parashat-mishpatim/morality-slavery>.
- Bedau, Mark, and Mark Triant. 2014. “Social and Ethical Implications of Creating Artificial Cells.” In *Ethics and Emerging Technologies*, edited by Ronald L. Sandler. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137349088>.
- Bedzow, Ira. 2017. *Maimonides for Moderns*. Switzerland: Springer International Publishing. <https://doi.org/10.1007/978-3-319-44573-1>.
- Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots.” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event, Canada: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.
- Bentham, Jeremy. (1789) 2019. *An Introduction to the Principles of Morals and Legislation*. Sydney NSW: Wentworth Press.
- Bergen, Jan Peter, and Peter-Paul Verbeek. 2020. “To-Do Is to Be: Foucault, Levinas, and Technologically Mediated Subjectivation.” *Philosophy & Technology* 34 (January): 325–48. <https://doi.org/10.1007/s13347-019-00390-7>.
- Berkovits, Eliezer. 1983. *Not in Heaven: The Nature and Function of Jewish Law*. New York: KTAV.
- . 2002. *Essential Essays on Judaism*. Edited by David Hazony. Jerusalem: Shalem Press.
- . 2004. *God, Man and History*. Edited by David Hazony. Jerusalem: Shalem Press.
- Bernstein, Immanuel. 2015. “Learning Halacha from Aggadah.” *Journal of Halacha and Contemporary Society* LXX.
- Bertolini, Andrea. 2018. “Human-Robot Interaction and Deception.” *Osservatorio Del*

- Diritto Civile E Commerciale, Rivista Semestrale* 2 (December): 645–59.
<https://doi.org/10.4478/91898>.
- Bertolini, Andrea, and Shabahang Arian. 2020. “3 Do Robots Care?” In *Aging between Participation and Simulation: Ethical Dimensions of Social Assistive Technologies*, edited by Joschka Haltaufderheide, Johanna Hovemann, and Jochen Vollmann, 35–52. Berlin: De Gruyter. <https://doi.org/10.1515/9783110677485-003>.
- Biba, Jacob. 2023. “Top 20 Humanoid Robots in Use Right Now.” Built In. September 13, 2023. <https://builtin.com/robotics/humanoid-robots>.
- Birhane, Abeba, and Jelle van Dijk. 2020. “Robot Rights? Let’s Talk about Human Welfare Instead.” *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, February, 207–13. <https://doi.org/10.1145/3375627.3375855>.
- Bishop, John Mark, and Slawomir Nasuto. 2013. “Of (Zombie) Mice and Animats.” In *Philosophy and Theory of Artificial Intelligence*, edited by Vincent C. Muller. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-31674-6>.
- Blau, Yitzchak. 2000. “The Implications of a Jewish Virtue Ethic.” *The Torah U-Madda Journal* 9: 19–41. <https://www.jstor.org/stable/40914639>.
- . 2002. “Ivan Karamazov Revisited: The Moral Argument for Religious Belief.” *The Torah U-Madda Journal* 11: 50–60. <http://www.jstor.org/stable/40914687>.
- Bleich, J. David. 1976. *Contemporary Halakbic Problems*. Vol. 1. NY: KTAV.
- . 1983a. “Status of the Deaf-Mute in Jewish Law.” In *Contemporary Halakbic Problems*. Vol. 2. NY: KTAV.
- . 1983b. “Teaching Torah to Non-Jews.” In *Contemporary Halakbic Problems*. Vol. 2. NY: KTAV.
- . 1998. “Cloning: Homologous Reproduction and Jewish Law.” *Tradition* 32 (3). <https://www.jstor.org/stable/23261122>.
- . 2013. “Is There an Ethic beyond Halakhah?” In *The Philosophical Quest: Of Philosophy, Ethics, Law and Halakhah*, 125–42. CT: Maggid Books.
- . 2015. “Mitochondrial DNA Replacement: How Many Mothers?” *Tradition* 48 (4): 60–84. <https://www.jstor.org/stable/44821375>.
- Block, Ned. 1995. “On a Confusion about a Function of Consciousness.” *Behavioral and Brain Sciences* 18 (02). <https://doi.org/10.1017/s0140525x00038188>.
- Bloom, Paul, and Sam Harris. 2018. “It’s Westworld. What’s Wrong with Cruelty to Robots?” *The New York Times*, April 23, 2018, sec. Opinion. <https://www.nytimes.com/2018/04/23/opinion/westworld-conscious-robots->

- morality.html.
- Boden, Margaret A. 1985. "Wonder and Understanding." *Zygon* 20 (4): 391–400. <https://doi.org/10.1111/j.1467-9744.1985.tb00605.x>.
- Boden, Margret, Joanna Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, et al. 2010. "Principles of Robotics." 2010. <https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>.
- Bostrom, Nick. 2003a. "Ethical Issues in Advanced Artificial Intelligence." Nickbostrom.com. 2003. <https://nickbostrom.com/ethics/ai>.
- . 2003b. "Are We Living in a Computer Simulation?" *The Philosophical Quarterly* 53 (211): 243–55. <https://doi.org/10.1111/1467-9213.00309>.
- . 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Bostrom, Nick, and Vincent C. Muller. 2016. "Future Progress in Artificial Intelligence: A Survey of Expert Opinion." In *Fundamental Issues of Artificial Intelligence*, edited by Vincent C. Muller. Berlin: Springer.
- Boucher, Philip. 2019. "How Artificial Intelligence Works." *European Parliament Think Tank*. European Parliamentary Research Service. [http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634420/EPRS_BRI\(2019\)634420_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634420/EPRS_BRI(2019)634420_EN.pdf).
- Brand, Lukas. 2020. "Why Machines That Talk Still Do Not Think, and Why They Might Nevertheless Be Able to Solve Moral Problems." In *Artificial Intelligence: Reflections in Philosophy, Theology, and the Social Sciences*, edited by Paul Gocke Benedikt and Astrid Rosenthal-Von der Putten, 203–17. Boston: Brill.
- Breazeal, Cynthia L. 2002. *Designing Sociable Robots*. Cambridge, Mass.: MIT Press.
- Brey, Philip. 2014. "From Moral Agents to Moral Factors: The Structural Ethics Approach." In *Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek, 125–42. Berlin: Springer. doi:10.1007/978-94-007-7914-3_8.
- Bringsjord, Selmer. 2007. "Ethical Robots: The Future Can Heed Us." *AI & SOCIETY* 22 (4): 539–50. <https://doi.org/10.1007/s00146-007-0090-9>.
- . 2010. "Meeting Florida's Challenge to Artificial Intelligence from the Knowledge-Game Test for Self-Consciousness." *Metaphilosophy* 41 (3): 292–312. <http://www.jstor.org/stable/24439827>.
- Brinkman, Bo. 2014. "Augmented Reality, and Its Philosophical and Ethical Challenges." In *Emerging Pervasive Information and Communication Technologies (PICT)*, edited by

- Kenneth D. Pimple. Dordrecht: Springer Netherlands.
<https://doi.org/10.1007/978-94-007-6833-8>.
- Broadie, Sarah. 1999. "Aristotle's Elusive Summum Bonum." *Social Philosophy and Policy* 16 (1): 233–51. <https://doi.org/10.1017/s0265052500002314>.
- . 2002. "Philosophical Introduction." In *Nicomachean Ethics*. Oxford: Oxford University Press.
- Broudy, Harry S. 1941. "Kierkegaard's Levels of Existence." *Philosophy and Phenomenological Research* 1 (3): 294–312. <https://doi.org/10.2307/2102760>.
- Bryson, Joanna. 2010. "Robots Should Be Slaves." In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues (Natural Language Processing: Vol. 8)*, edited by Y. Wilk, 63–74. John Benjamins Publishing Company. <https://researchportal.bath.ac.uk/en/publications/robots-should-be-slaves>.
- . 2012. "Patience Is Not a Virtue." In *The Machine Question: AI, Ethics and Moral Responsibility*, edited by David Gunkel, Joanna Bryson, and Steve Torrance, 73–77. Society for the Study of Artificial Intelligence and the Simulation of Behaviour. <http://events.cs.bham.ac.uk/turing12/>.
- . 2016. "Robots Are Owned. Owners Are Taxed. Internet Services Cost Information." *Adventures in NI*. June 23, 2016. <https://joanna-bryson.blogspot.com/2016/06/robots-are-owned-owners-are-taxed.html>.
- . 2020a. "The Coexistence of Artificial and Natural Intelligence." *Digital Future Society*. March 2, 2020. <https://digitalfuturesociety.com/interviews/the-coexistence-of-artificial-and-natural-intelligence-interview-with-joanna-bryson/>.
- . 2020b. "The Artificial Intelligence of the Ethics of Artificial Intelligence." In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das, 3–25. NY: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.013.1>.
- Buber, Martin. 1970. *I and Thou*. Translated by Walter Arnold Kaufmann. NY: Charles Scribner's Sons.
- Burdett, Michael S. 2020. "Personhood and Creation in an Age of Robots and AI: Can We Say 'You' to Artifacts?" *Zygon*® 55 (2): 347–60. <https://doi.org/10.1111/zygo.12595>.
- Burke, James. 1978. "Connections Episode 1: The Trigger Effect." BBC. https://archive.org/details/james-burke-connections_s01e01.
- Calverley, David J. 2011. "Legal Rights for Machines." In *Machine Ethics*, edited by

- Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Cappuccio, Massimiliano L., Anco Peeters, and William McDonald. 2019. "Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition." *Philosophy & Technology* 33 (1): 9–31. <https://doi.org/10.1007/s13347-019-0341-y>.
- Cappuccio, Massimiliano L., Eduardo B. Sandoval, Omar Mubin, Mohammad Obaid, and Mari Velonaki. 2020. "Can Robots Make Us Better Humans?" *International Journal of Social Robotics* 13 (October): 7–22. <https://doi.org/10.1007/s12369-020-00700-6>.
- Carmy, Shalom. 1996. "Pluralism and the Category of the Ethical." *Tradition: A Journal of Orthodox Jewish Thought* 30 (4): 145–63. <http://www.jstor.org/stable/23261241>.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- . 2010. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17. <https://consc.net/papers/singularityjcs.pdf>.
- . 2018. "Episode 25: Sean Carroll Interviews David Chalmers on Consciousness." Sean Carroll's Mindscape. December 3, 2018. <https://www.preposterousuniverse.com/podcast/2018/12/03/episode-25-david-chalmers-on-consciousness-the-hard-problem-and-living-in-a-simulation/>.
- Charpa, Ulrich. 2012. "Synthetic Biology and the Golem of Prague: Philosophical Reflections on a Suggestive Metaphor." *Perspectives in Biology and Medicine* 55 (4): 554–70. <https://doi.org/10.1353/pbm.2012.0036>.
- Cherlow, Yuval. 2016. *In His Image*. Jerusalem: Maggid.
- Choi, Charles Q. 2013. "Brain Scans Show Humans Feel for Robots." *IEEE Spectrum: Technology, Engineering, and Science News*, April 24, 2013. <https://spectrum.ieee.org/robotics/artificial-intelligence/brain-scans-show-humans-feel-for-robots>.
- Chomanski, Bartek. 2019. "What's Wrong with Designing People to Serve?" *Ethical Theory and Moral Practice* 22 (4): 993–1015. <https://doi.org/10.1007/s10677-019-10029-3>.
- Churchland, Paul M., and Patricia Churchland. 1990. "Could a Machine Think?" *Scientific American* 262. <https://www.jstor.org/stable/24996642>.
- Coeckelbergh, Mark. 2009. "Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics." *International Journal of Social Robotics* 1 (3): 217–21. <https://doi.org/10.1007/s12369-009-0026-2>.

- . 2010a. “Humans, Animals, and Robots: A Phenomenological Approach to Human-Robot Relations.” *International Journal of Social Robotics* 3 (2): 197–204. <https://doi.org/10.1007/s12369-010-0075-6>.
- . 2010b. “Moral Appearances: Emotions, Robots, and Human Morality.” *Ethics and Information Technology* 12 (3): 235–41. <https://doi.org/10.1007/s10676-010-9221-y>.
- . 2010c. “Robot Rights? Towards a Social-Relational Justification of Moral Consideration.” *Ethics and Information Technology* 12 (3): 209–21. <https://doi.org/10.1007/s10676-010-9235-5>.
- . 2013. “The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics.” *Philosophy & Technology* 27 (1): 61–77. <https://doi.org/10.1007/s13347-013-0133-8>.
- . 2014. “Robotic Appearances and Forms of Life. A Phenomenological-Hermeneutical Approach to the Relation between Robotics and Culture.” In *Robotics in Germany and Japan: Philosophical and Technical Perspectives*, edited by Michael Funk and Bernhard Irrgang. Frankfurt Am Main: Peter Lang Edition.
- . 2015. “The Tragedy of the Master: Automation, Vulnerability, and Distance.” *Ethics and Information Technology* 17 (3): 219–29. <https://doi.org/10.1007/s10676-015-9377-6>.
- . 2016. “Care Robots and the Future of ICT-Mediated Elderly Care: A Response to Doom Scenarios.” *AI & Society* 31 (4): 455–62. <https://doi.org/10.1007/s00146-015-0626-3>.
- . 2020a. *AI Ethics*. Cambridge, Massachusetts: MIT Press.
- . 2020b. “How to Use Virtue Ethics for Thinking about the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance.” *International Journal of Social Robotics* 13 (October): 31–40. <https://doi.org/10.1007/s12369-020-00707-z>.
- . 2020c. “Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots, with Implications for Thinking about Animals and Humans.” *Minds and Machines*, December. <https://doi.org/10.1007/s11023-020-09554-3>.
- . 2021. “Three Responses to Anthropomorphism in Social Robotics: Towards a Critical, Relational, and Hermeneutic Approach.” *International Journal of Social Robotics*, March. <https://doi.org/10.1007/s12369-021-00770-0>.

- Cole, David. 2020. "The Chinese Room Argument." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Winter 2020. Metaphysics Research Lab, Stanford University.
- Conradie, Ernst M. 2017. "Do Only Humans Sin? In Conversation with Frans de Waal." In *Issues in Science and Theology: Are We Special?*, edited by Michael Fuller, Dirk Evers, Anne Runehov, and Knut-Willy Saether, 117–33. Springer. https://doi.org/10.1007/978-3-319-62124-1_9.
- Curry, Oliver Scott. 2016. "Morality as Cooperation: A Problem-Centred Approach." In *The Evolution of Morality*, edited by Shackelford T. and Hansen R., 27–51. Cham: Springer. https://doi.org/10.1007/978-3-319-19671-8_2.
- Damasio, Antonio R. 2010. *Self Comes to Mind: Constructing the Conscious Brain*. London: Vintage.
- Danaher, John. 2018. "Why We Should Create Artificial Offspring: Meaning and the Collective Afterlife." *Science and Engineering Ethics* 24 (4): 1097–1118. <https://doi.org/10.1007/s1194801799320>.
- . 2019a. "The Philosophical Case for Robot Friendship." *Journal of Posthuman Studies* 3 (1): 5. <https://doi.org/10.5325/jpoststud.3.1.0005>.
- . 2019b. "Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism." *Science and Engineering Ethics* 26 (4): 2023–49. <https://doi.org/10.1007/s11948-019-00119-x>.
- Darling, Kate. 2016. "Extending Legal Protection to Social Robots." In *Robot Law*, edited by Ryan Calo, A. Michael Froomkin, and Ian Kerr, 213–31. MA: Edward Elgar. DOI 10.4337/9781783476732.
- . 2017. "Who's Johnny? Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy." In *Robot Ethics 2.0*, edited by Patrick Lin, Ryan Jenkins, and Keith Abney, 173–88. NY: Oxford University Press.
- Darling, Kate, Palash Nandy, and Cynthia Breazeal. 2015. "Empathic Concern and the Effect of Stories in Human-Robot Interaction." In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication*, 770–75. IEEE.
- Davenport, David. 2014. "Moral Mechanisms." *Philosophy & Technology* 27 (1): 47–60. <https://doi.org/10.1007/s13347-013-0147-2>.
- Davidson, Herbert A. 1987. "The Middle Way in Maimonides' Ethics." *Proceedings of the American Academy for Jewish Research* 54: 31–72. <https://doi.org/10.2307/3622580>.
- Denis, Lara. 2000. "Kant's Conception of Duties Regarding Animals: Reconstruction and

- Reconsideration.” *History of Philosophy Quarterly* 17 (4): 405–23.
<https://www.jstor.org/stable/27744866>.
- Dennett, Daniel. 1976. “Conditions of Personhood.” In *The Identities of Persons*, edited by A. Rorty. CA: UC Press.
<https://dl.tufts.edu/downloads/w0892p348?filename=sf268h06v.pdf>.
- . 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- . 2007. *Consciousness Explained*. New York: Little, Brown And Company.
- Descartes, Rene. (1637) 2017. *Discourse on the Method of Rightly Conducting One’s Reason and Seeking Truth in the Sciences*. Translated by Jonathan Bennett. Early Modern Philosophy. <https://www.earlymoderntexts.com/assets/pdfs/descartes1637.pdf>.
- . (1641) 2017. *Meditations on First Philosophy*. Translated by Jonathan Bennett. Early Modern Philosophy. <https://www.earlymoderntexts.com/assets/pdfs/descartes1641.pdf>.
- Diamond, James. 2016. “The Treatment of Non-Israelite Slaves: From Moses to Moses.” The Torah.com. Project TABS. April 19, 2016.
<https://www.thetorah.com/article/the-treatment-of-non-israelite-slaves-from-moses-to-moses>.
- Dietrich, Eric. 2011. “Homo Sapiens 2.0: Building the Better Robots of Our Nature.” In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Dihal, Kanta. 2020. “Enslaved Minds: Artificial Intelligence, Slavery, and Revolt.” In *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, edited by Stephen Cave, Kanta Dihal, and Sarah Dillon, 189–212. Oxford: Oxford University Press.
- Domingos, Pedro. 2018. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books, a Member of the Perseus Books Group.
- Dorff, Elliot N., and Jonathan K. Crane, eds. 2012. *The Oxford Handbook of Jewish Ethics and Morality*. Oxford: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199736065.013.0001>.
- Dostoyevsky, Fyodor. (1880) 2019. *The Brothers Karamazov*. Translated by Constance Garnett. New Delhi: Om Books International.
- Douglass, Frederick. 1845. *Narrative of the Life of Frederick Douglass, an American Slave*.

http://www.ibiblio.org/ebooks/Douglass/Narrative/Douglass_Narrative.pdf.

- . 1848. “The Myth of the Happy Slaves.” *The North Star*, April 28, 1848.
<http://www.accessible-archives.com/2012/04/the-myth-of-the-happy-slaves-in>.
- Douglas, Thomas. 2014. “Moral Enhancement.” In *Ethics and Emerging Technologies*, edited by Ronald L. Sandler. London: Palgrave Macmillan UK.
<https://doi.org/10.1057/9781137349088>.
- Dreyfus, Hubert L. 1972. *What Computers Can't Do*. NY: Harper & Row.
- Dreyfus, Hubert L. 1992. *What Computers Still Can't Do*. MIT Press.
- Dubber, Markus Dirk, Frank Pasquale, and Sunit Das, eds. 2020. *The Oxford Handbook of Ethics of AI*. New York: Oxford University Press.
- Duffy, Brian R. 2003. “Anthropomorphism and the Social Robot.” *Robotics and Autonomous Systems* 42 (3-4): 177–90. [https://doi.org/10.1016/s0921-8890\(02\)00374-3](https://doi.org/10.1016/s0921-8890(02)00374-3).
- Dyson, George. 2012. *Darwin among the Machines: The Evolution of Global Intelligence*. Basic Books.
- Eisen, Chaim. 1991. “Mosheh Rabbeinu and Rabbi Akiva.” *Jewish Thought* 1 (2).
- . 1992. “You Will Be like God.” *Jewish Thought* 2 (1).
- Elamrani, A., and R. Yampolskiy. 2019. “Reviewing Tests for Machine Consciousness.” *Journal of Consciousness Studies* 26 (5-6). <https://philpapers.org/rec/ELARTF>.
- Epstein, Isadore. 1983. “Foreword.” In *Midrash Rabbah*, edited by Harry Freedman and Maurice Simon. Vol. One: Genesis. NY: Soncino.
- Eskens, Romy. 2017. “Is Sex with Robots Rape?” *Journal of Practical Ethics* 5: 62–76.
- Falk, Zeev. 1961. *Marriage and Divorce: Reforms in the Family Law of German-French*. Jerusalem: Mifal Hashichpul.
- Feinberg, Joel. 1986. “Wrongful Life and the Counterfactual Element in Harming.” *Social Philosophy and Policy* 4 (01): 145. <https://doi.org/10.1017/s0265052500000467>.
- Feinberg, Todd E., and Jon Mallatt. 2016. *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambridge, Massachusetts: The MIT Press.
- Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford, Calif.: Stanford University Press.
- Fitzgerald, McKenna, Aaron Boddy, and Seth D. Baum. 2020. “2020 Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy.” Global Catastrophic Risk Institute Technical Report 20-1. <https://gcrinstitute.org/2020-survey-of->

- artificial-general-intelligence-projects-for-ethics-risk-and-policy/.
- Fjelland, Ragnar. 2020. “Why General Artificial Intelligence Will Not Be Realized.” *Humanities and Social Sciences Communications* 7 (1). <https://doi.org/10.1057/s41599-020-0494-4>.
- Floridi, Luciano. 2008. “Artificial Intelligence’s New Frontier: Artificial Companions and the Fourth Revolution.” *Metaphilosophy* 39 (4/5): 651–55. <https://www.jstor.org/stable/24439697>.
- . 2014. “Artificial Agents and Their Moral Nature.” In *Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek. Springer.
- . 2016. “Should We Be Afraid of AI?” *Aeon*, 2016. <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-improbable>.
- Foerst, Anne. 1998. “Cog, a Humanoid Robot, and the Question of the Image of God.” *Zygon* 33 (1): 91–111. <https://doi.org/10.1111/0591-2385.1291998129>.
- . 2009. “Robots and Theology.” *EWE* 20 (2): 181–93. https://www.researchgate.net/publication/273886034_Robots_and_Theology.
- Fossa, Fabio. 2018. “Artificial Moral Agents: Moral Mentors or Sensible Tools?” *Ethics and Information Technology* 20 (2): 115–26. <https://doi.org/10.1007/s10676-018-9451-y>.
- Fox, Marvin. 1994. *Interpreting Maimonides: Studies in Methodology, Metaphysics, and Moral Philosophy*. Chicago: University Of Chicago Press.
- Frank, Daniel H. 1985. “The End of the Guide: Maimonides on the Best Life of Man.” *Judaism* 34 (4): 485–95.
- Frankl, Viktor. (1946) 2006. *Man’s Search for Meaning*. Boston: Beacon Press.
- Franklin, Stan. 2003. “Ida: A Conscious Artifact?” *Journal of Consciousness Studies* 10: 47–66. <https://ccrg.cs.memphis.edu/assets/papers/IDA-ConsciousArtifact.pdf>.
- Friedberg, Albert Dov. 2019. “‘... Hasidut Leads to Ruah Haqodesh ...’ – a New Reading of the Closing Chapters of Maimonides’ Guide.” *JSIJ* 17: 1–32. <http://jewish-faculty.biu.ac.il/files/jewish-faculty/shared/JSIJ17/friedberg.pdf>.
- Gamez, David. 2008. “Progress in Machine Consciousness.” *Consciousness and Cognition* 17 (3): 887–910. <https://doi.org/10.1016/j.concog.2007.04.005>.
- . 2018. *Human and Machine Consciousness*. Open Book Publishers.
- Garcia, Tamara, and Ronald Sandler. 2014. “Enhancing Justice?” In *Ethics and Emerging Technologies*, edited by Ronald L. Sandler. London: Palgrave Macmillan UK. <https://doi.org/10.1057/9781137349088>.

- Gaon, Saadia. (933AD) 1976. *The Book of Beliefs and Opinions*. Translated by Samuel Rosenblatt. New Haven: Yale University Press.
- Gerdes, Anne. 2016. "The Issue of Moral Consideration in Robot Ethics." *ACM SIGCAS Computers and Society* 45 (3): 274–79. <https://doi.org/10.1145/2874239.2874278>.
- Gerdes, Anne, and Peter Ohrstrom. 2015. "Issues in Robot Ethics Seen through the Lens of a Moral Turing Test." *Journal of Information, Communication and Ethics in Society* 13 (2): 98–109. <https://doi.org/10.1108/jices-09-2014-0038>.
- Ghiglino, Davide, and Agnieszka Wykowska. 2020. "When Robots (Pretend To) Think." In *Artificial Intelligence: Reflections in Philosophy, Theology, and the Social Sciences*, edited by Benedikt Paul Göcke and Astrid Rosenthal-Von der Putten, 49–74. Boston: Brill.
- Gilbert, David. 2024. "Google's 'Woke' Image Generator Shows the Limitations of AI." *Wired*. February 22, 2024. <https://www.wired.com/story/google-gemini-woke-ai-image-generation/>.
- Gocke, Benedikt Paul. 2020. "Could Artificial General Intelligence Be an End-In-Itself?" In *Artificial Intelligence: Reflections in Philosophy, Theology, and the Social Sciences*, edited by Benedikt Paul Gocke and Astrid Rosenthal-Von der Putten. Boston: Brill.
- Gocke, Benedikt Paul, and Astrid Rosenthal-Von, eds. 2020. *Artificial Intelligence : Reflections in Philosophy, Theology, and the Social Sciences*. Boston: Brill.
- Goldberg, Judah. 2014. "Is There an Ethic beyond Formal Jewish Law? ." *VBM. Yeshivat Har Etzion*. December 11, 2014. <https://www.etzion.org.il/en/halakha/studies-halakha/philosophy-halakha/there-ethic-beyond-formal-jewish-law>.
- Goltz, Nachshon, John Zeleznikow, and Tracey Dowdeswell. 2020. "From the Tree of Knowledge and the Golem of Prague to Kosher Autonomous Cars: The Ethics of Artificial Intelligence through Jewish Eyes." *Oxford Journal of Law and Religion* 9 (February): 132–56. <https://doi.org/10.1093/ojlr/rwaa015>.
- Goodman, Lenn. 2009. "Happiness." In *The Cambridge History of Medieval Philosophy*, 1:455–71. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CHOL9780521762168.035>.
- Gordon, John-Stewart. 2020. "What Do We Owe to Intelligent Robots?" *AI & SOCIETY* 35 (1): 209–23. <https://doi.org/10.1007/s00146-018-0844-6>.
- Graaf, Maartje M.A de, and Bertram F Malle. 2019. "People's Explanations of Robot

- Behavior Subtly Reveal Mental State Inferences.” In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 239–48. IEEE. <https://doi.org/10.1109/HRI.2019.8673308>.
- Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2018. “Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts.” *Journal of Artificial Intelligence Research* 62 (July): 729–54. <https://doi.org/10.1613/jair.1.11222>.
- Grau, Christopher. 2011. “There Is No ‘I’ in ‘Robot.’” In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Gray, Kurt, and Chelsea Schein. 2012. “Two Minds vs. Two Philosophies: Mind Perception Defines Morality and Dissolves the Debate between Deontology and Utilitarianism.” *Review of Philosophy and Psychology* 3 (3): 405–23. <https://doi.org/10.1007/s13164-012-0112-5>.
- Green, Erin Elizabeth. 2018. “Robots and AI: The Challenge to Interdisciplinary Theology.” University of St. Michael’s College. https://tspace.library.utoronto.ca/bitstream/1807/93393/1/Green_Erin_E_201811_PhD_thesis.pdf.
- Grodzinsky, Frances S., Keith W. Miller, and Marty J. Wolf. 2014. “Developing Automated Deceptions and the Impact on Trust.” *Philosophy & Technology* 28 (1): 91–105. <https://doi.org/10.1007/s13347-014-0158-7>.
- Grossl, Johannes. 2020. “Artificial Intelligence and Polygenic Scoring.” In *Artificial Intelligence: Reflections in Philosophy, Theology, and the Social Sciences*, edited by Benedikt Paul Gocke and Astrid Rosenthal-Von der Putten. Boston: Brill.
- Grossman, Avraham. 2001. *Hasidot U-Mordot: Nashim Yehudiyot Be-Eropah Bi-Yeme-Ha-Benayim*. Jerusalem: Merkaz Zalman Shazar.
- Grunfeld, Isidor. 1975. *The Jewish Dietary Laws*. London: Soncino Press.
- Guizzo, Erico. 2010. “Hiroshi Ishiguro: The Man Who Made a Copy of Himself.” IEEE Spectrum: Technology, Engineering, and Science News, April 22, 2010. <https://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself>.
- Gunkel, David J. 2012. *The Machine Question Critical Perspectives on AI, Robots, and Ethics*. The MIT Press.
- . 2017. “The Other Question: Can and Should Robots Have Rights?” *Ethics and Information Technology* 20 (2): 87–99. <https://doi.org/10.1007/s10676-017-9442-4>.

- . 2018. *Robot Rights*. Cambridge, Mass: MIT Press.
- Gunkel, David J., and Jordan Joseph Wales. 2021. “Debate: What Is Personhood in the Age of AI?” *AI & SOCIETY* 36 (January): 473–86. <https://doi.org/10.1007/s00146-020-01129-1>.
- Ha-Levi, Aaron. (1523) 1978. *Sefer HaHinnuch: The Book of [Mizvah] Education*. Translated by Charles Wengrov. New York: Feldheim.
- Haikonen, Pentti O. 2019. *Consciousness and Robot Sentience*. 2nd ed. Singapore: World Scientific Publishing.
- Hales, Colin G. 2014. *The Revolutions of Scientific Structure*. Singapore: World Scientific Publishing.
- Hameroff, Stuart, and Roger Penrose. 1996. “Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness.” *Mathematics and Computers in Simulation* 40 (3-4): 453–80. [https://doi.org/10.1016/0378-4754\(96\)80476-9](https://doi.org/10.1016/0378-4754(96)80476-9).
- Hampton, Gregory Jerome. 2015. *Imagining Slaves and Robots in Literature, Film, and Popular Culture : Reinventing Yesterday’s Slave with Tomorrow’s Robot*. New York: Lexington Books.
- Harari, Yuval Noah. 2015. *Sapiens: A Brief History of Humankind*. London: Vintage.
- . 2019. *21 Lessons for the 21st Century*. Random House.
- Harbron, Patrick. 2000. “The Future of Humanoid Robots.” *Discover Magazine*, 2000. <https://www.discovermagazine.com/technology/the-future-of-humanoid-robots>.
- Hartman, David. 1986. *Maimonides: Torah and Philosophic Quest*. Philadelphia: Jewish Publication Society.
- Harvey, Warren Zev. 1990. “Maimonides on Human Perfection, Awe, and Politics.” In *The Thought of Moses Maimonides: Philosophical and Legal Studies*, edited by Ira Robinson. NY: Mellen.
- Haugeland, John. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press.
- Hauskeller, Michael. 2013. *Better Humans?* Routledge.
- . 2017. “Automatic Sweethearts for Transhumanists.” In *Robot Sex*, edited by John Danaher and Neil McArthur. MIT Press. <https://doi.org/10.7551/mitpress/9780262036689.001.0001>.
- . 2020. “What Is It like to Be a Bot? SF and the Morality of Intelligent Machines.” In *Minding the Future. Contemporary Issues in Artificial Intelligence*, edited by Barry

- Dainton, Will Slocombe, and Attila Tanyi. NY: Springer.
- Hawley, Scott. 2019. "Challenges for an Ontology of Artificial Intelligence." *Perspectives on Science and Christian Faith* 71 (2): 83–95.
- Hayun, Yehudah. 1996. *Talmudic Principles, Concepts and Expressions and Concepts*. Bnei Brak: Condensed Torah Encyclopedia.
- Hegel, Georg Wilhelm Friedrich. (1807) 2019. *The Phenomenology of Spirit*. Edited and translated by Terry Pinkard. Cambridge University Press.
- Heil, John. 2004. *Philosophy of Mind: A Guide and Anthology*. Oxford: Oxford University Press.
- Hernandez-Orallo, Jose. 2017. *The Measure of All Minds*. Cambridge: Cambridge University Press.
- Herzfeld, Noreen. 2002a. *In Our Image: Artificial Intelligence and the Human Spirit*. Minneapolis: Fortress Press.
- . 2002b. "Creating in Our Own Image: Artificial Intelligence and the Image of God." *Zygon* 37 (2): 303–16. <https://doi.org/10.1111/0591-2385.00430>.
- Heschel, Abraham Joshua. 1965. *Who Is Man?* Stanford, CA: Stanford University Press.
- Hirsch, Samson Raphael. (1867) 1989. *The Pentateuch*. Translated by Isaac Levy. Gateshead: Judaica Press.
- Hobbes, Thomas. (1651) 2017. *Leviathan*. Translated by Jonathan Bennett. Early Modern Philosophy. <https://www.earlymoderntexts.com/assets/pdfs/hobbes1651part1.pdf>.
- Hoffman, Guy. 2012. "Embodied Cognition for Autonomous Interactive Robots." *Topics in Cognitive Science* 4 (4): 759–72. <https://doi.org/10.1111/j.1756-8765.2012.01218.x>.
- Holland, Nancy. 2018. *Heidegger and the Problem of Consciousness*. Bloomington, Indiana: Indiana University Press.
- Horgan, Terence. 2013. "Original Intentionality Is Phenomenal Intentionality." Edited by Sherwood J. B. Sugden. *Monist* 96 (2): 232–51. <https://doi.org/10.5840/monist201396212>.
- Horowitz, Isaac. 1873. *Beer Yitzhak*. Lvov: A. N. Suss. <https://hebrewbooks.org/31492>.
- Huebner, Bryce. 2009. "Commonsense Concepts of Phenomenal Consciousness: Does Anyone Care about Functional Zombies?" *Phenomenology and the Cognitive Sciences* 9 (1): 133–55. <https://doi.org/10.1007/s11097-009-9126-6>.
- Hume, David. 1739. *A Treatise of Human Nature*. Clarendon Press. <https://oll>

- resources.s3.us-east-2.amazonaws.com/oll3/store/titles/342/0213_Bk.pdf.
- Idel, Moshe. 1996. *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid (HEBREW)*. Jerusalem: Schocken.
- . 2019. *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid*. Brooklyn, N.Y.: KTAV.
- IEEE. 2019. “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.” Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/>.
- Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington, Ind.: Indiana University Press.
- Irwin, Terence. 2006. “Introduction.” In *Nicomachean Ethics*. Ind.: Hackett.
- Ishiguro, Kazuo. 2021. *Klara and the Sun*. Vintage.
- Jacobs, Harriet. 2020. *Incidents in the Life of a Slave Girl*. S.L.: Modern Library.
- Jakobovits, Julian (Yoel). 2000. “Cloning and Its Challenges.” *The Torah U-Madda Journal* 9: 195–98. <https://www.jstor.org/stable/40914654>.
- Jefferson, Geoffrey. 1949. “The Mind of Mechanical Man.” *The British Medical Journal* 1 (4616): 1105–10. <https://www.jstor.org/stable/25372573>.
- Jefferson (Bernard), P.J. 1980. *Curlender v. Bio-Science Labs*, 106. Court of Appeals of California, Second Appellate District, Division One. <https://law.justia.com/cases/california/court-of-appeal/3d/106/811.html>.
- Jiménez-Rodríguez, Luis O. 2017. “Human Uniqueness or Anthropocentrism? Semantic, Anthropological and Theological Clarifications in Dialogue with Damasio’s Neuroscience.” In *Issues in Science and Theology: Are We Special?*, edited by Michael Fuller, Dirk Evers, Anne Runehov, and Knut-Willy Saether, 191–207. Springer. https://doi.org/10.1007/978-3-319-62124-1_14.
- Johnson, Aaron M., and Sidney Axinn. 2014. “Acting vs. Being Moral: The Limits of Technological Moral Actors.” *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*, May. <https://doi.org/10.1109/ethics.2014.6893396>.
- Johnson, David Kyle. 2013. “Do Souls Exist?” *Think* 12 (35): 61–75. <https://doi.org/10.1017/s1477175613000195>.
- . 2020. *Black Mirror and Philosophy Dark Reflections*. NJ: Wiley-Blackwell.
- Johnson, Deborah G. 2011. “Computer Systems: Moral Entities but Not Moral Agents.”

- In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Johnson, Deborah G., and Keith Miller. 2008. *Computer Ethics: Analyzing Information Technology*. Upper Saddle River, N.J.: Prentice Hall.
- Johnson, Deborah G., and Mario Verdicchio. 2018. "Why Robots Should Not Be Treated like Animals." *Ethics and Information Technology* 20 (4): 291–301. <https://doi.org/10.1007/s10676-018-9481-5>.
- Johnson, Paul. 1987. *A History of the Jews*. NY: Harper.
- Jonas, Hans. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: The University of Chicago Press.
- Jones, Cynthia M. 2014. "Preserving Life, Destroying Privacy: PICT and the Elderly." In *Emerging Pervasive Information and Communication Technologies (PICT)*, edited by Kenneth D. Pimple. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-007-6833-8>.
- Jones, Meg Leta, and Jason Millar. 2017. "Hacking Metaphors in the Anticipatory Governance of Emerging Technology." In *The Oxford Handbook of Law, Regulation and Technology*, edited by Roger Brownsword, Eloise Scotford, and Karen Yeung. Oxford: University Press. <https://doi.org/10.1093/oxfordhb/9780199680832.013.34>.
- Joy, Bill. 2000. "Why the Future Doesn't Need Us." *Wired*. April 2000. <https://www.wired.com/2000/04/joy-2/>.
- Kagan, Shelly. 2015. "What's Wrong with Speciesism?" *Journal of Applied Philosophy* 33 (1): 1–21. <https://doi.org/10.1111/japp.12164>.
- Kaiser, Christopher. 1989. "How Can a Theological Understanding of Humanity Enrich Artificial Intelligence Work?" *The Asbury Journal* 44 (2). <https://place.asburyseminary.edu/asburyjournal/vol44/iss2/6/>.
- Kamm, Francis. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford: University Press. <https://doi.org/10.1093/acprof:oso/9780195189698.001.0001>.
- Kant, Immanuel. 1996. *Lectures on Ethics*. Edited by Peter Heath and J. B. Schneewind. New York: Cambridge University Press.
- . (1785) 2006. *Groundwork of the Metaphysics of Morals*. Edited and translated by Mary J Gregor. Cambridge: Cambridge University Press.
- . (1797) 2017. *The Metaphysics of Morals*. Edited by Lara Denis. Translated by Mary Gregor. *Cambridge Texts in the History of Philosophy*. 2nd ed. Cambridge: Cambridge

- University Press. <https://doi.org/10.1017/9781316091388>.
- Kaplan, Aryeh. 1989. *Babir*. Weiser Books.
- . 1997. *Sefer Yetzirah - Book of Creation*. SF: Weiser.
- Kass, Leon. 2006. *The Beginning of Wisdom: Reading Genesis*. Chicago: University of Chicago Press.
- . 2014. “Preventing a Brave New World.” In *Ethics and Emerging Technologies*, edited by Ronald L. Sandler. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137349088>.
- Kastrup, Bernardo. 2017. “An Ontological Solution to the Mind-Body Problem.” *Philosophies* 2 (4): 10. <https://doi.org/10.3390/philosophies2020010>.
- Kautz, Henry A. 2022. “The Third AI Summer: AAAI Robert S. Englemore Memorial Lecture.” *AI Magazine* 43 (1): 105–25. <https://doi.org/10.1002/aaai.12036>.
- Kedar, Shmuel. 2007. *Torat Obel (HEBREW)*. Vol. 1. Ofra: Kedar. <https://download.hebrewbooks.org/downloadhandler.ashx?req=56253>.
- Kellner, Menachem. 1990. *Maimonides on Human Perfection*. GA: Scholars Press.
- . 2009. *Science in the Bet Midrash: Studies in Maimonides*. MA: Academic Studies Press.
- Keynes, John Maynard. (1930) 2009. “Economic Possibilities for Our Grandchildren.” In *Essays in Persuasion*. New York: Classic House Books. <http://www.econ.yale.edu/smith/econ116a/keynes1.pdf>.
- Kierkegaard, Søren. 1985. *Fear and Trembling*. Harmondsworth: Penguin.
- Kim, Min-Sun, and Eun-Joo Kim. 2012. “Humanoid Robots as ‘the Cultural Other’: Are We Able to Love Our Creations?” *AI & SOCIETY* 28 (3): 309–18. <https://doi.org/10.1007/s00146-012-0397-z>.
- Kingwell, Mark. 2020. “Are Sentient AIs Persons.” In *The Oxford Handbook of Ethics of AI*, edited by Markus Dirk Dubber, Frank Pasquale, and Sunit Das. New York: Oxford University Press.
- Koch, Christof. 2019. *Feeling of Life Itself*. Cambridge, Massachusetts: The MIT Press.
- Kogan, Barry S. 1989. “‘What Can We Know and When Can We Know It?’ Maimonides on the Active Intelligence and Human Cognition.” In *Moses Maimonides and His Time*, edited by Eric L. Ormsby, 121–37. Washington, D.C.: Catholic University of America Press.
- Kohler, Sebastian. 2023. “Can We Have Moral Status for Robots on the Cheap?” *Journal of Ethics and Social Philosophy* 24 (1). <https://doi.org/10.26556/jesp.v24i1.1659>.
- Kook, Abraham Isaac. 1906. *Eder Hayakar*.

- <http://www.daat.ac.il/daat/v1/tohen.asp?id=334>.
- Koplin, Julian, and Dominic Wilkinson. 2019. "Moral Uncertainty and the Farming of Human-Pig Chimeras." *Journal of Medical Ethics* 45 (7): 440–46. <https://doi.org/10.1136/medethics-2018-105227>.
- Korn, Eugene. 2002. "Legal Floors and Moral Ceilings: A Jewish Understanding of Law and Ethics." *The Edab Journal* 2 (2). <https://library.yctorah.org/files/2016/09/Legal-Floors-and-Moral-Ceilings-A-Jewish-Understanding-Of-Law-and-Ethics.pdf>.
- Kreisel, Howard. 1992. "Individual Perfection vs. Communal Welfare and the Problem of Contradictions in Maimonides' Approach to Ethics." *Proceedings of the American Academy for Jewish Research* 58: 107–41. <https://doi.org/10.2307/3622631>.
- Kurzweil, Ray. 2006. *The Singularity Is Near: When Humans Transcend Biology*. Penguin Books.
- Ladak, Ali. 2022. "Is Artificial Consciousness Possible? A Summary of Selected Books." Sentience Institute. 2022. <https://www.sentienceinstitute.org/blog/is-artificial-consciousness-possible>.
- LaGrandeur, Kevin. 2011. "The Persistent Peril of the Artificial Slave." *Science Fiction Studies* 38 (2): 232–52. <https://doi.org/10.5621/sciefictstud.38.2.0232>.
- . 2013. *Androids and Intelligent Networks in Early Modern Literature and Culture Artificial Slaves*. NY: Routledge.
- Lamm, Norman. 1965. "The Religious Implications of Extraterrestrial Life." *Tradition Online* 7 (4). <https://traditiononline.org/the-religious-implications-of-extraterrestrial-life/>.
- . 2007. "Amalek and the Seven Nations: A Case of Law vs. Morality." In *War and Peace in the Jewish Tradition*, edited by Lawrence H Schiffman and Joel B Wolowelsky. New York: Michael Scharf Publication Trust of the Yeshiva University Press.
- Lavender, Isiah. 2011. *Race in American Science Fiction*. Indiana: Indiana University Press.
- Lee, Lisa M. 2014. "Health Information in the Background: Justifying Public Health Surveillance without Patient Consent." In *Emerging Pervasive Information and Communication Technologies (PICT)*, edited by Kenneth D. Pimple. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-007-6833-8>.
- Leibowitz, Nehama. (1976) 1986. *Studies in Shemot (Exodus)*. Translated by Aryeh Newman. Sixth. Jerusalem: World Zionist Organization.

- Leibowitz, Yeshayahu. 1992. *Judaism, Human Values, and the Jewish State*. Edited by Eliezer Goldman. Translated by Yoram Navon. Cambridge, Mass.: Harvard University Press.
- Leibtag, Menachem. 2003. "The Akeda and Miscellaneous Topics." *VBM*. Gush Etzion: Yeshivat Har Etzion. <https://www.etzion.org.il/en/tanakh/torah/sefer-bereshit/parashat-vayera/vayera-akeda-and-miscellaneous-topics>.
- Leong, Brenda, and Evan Selinger. 2019. "Robot Eyes Wide Shut." *Proceedings of the Conference on Fairness, Accountability, and Transparency*, January. <https://doi.org/10.1145/3287560.3287591>.
- Leveringhaus, Alex. 2016. *Ethics and Autonomous Weapons*. London: Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-52361-7>.
- Levy, David. 2008. *Love and Sex with Robots*. New York: Harper Perennial.
- . 2009. "The Ethical Treatment of Artificially Conscious Robots." *International Journal of Social Robotics* 1 (3): 209–16. <https://doi.org/10.1007/s12369-009-0022-6>.
- Liao, S. Matthew. 2010. "The Basis of Human Moral Status." *Journal of Moral Philosophy* 7 (2): 159–79. <https://doi.org/10.1163/174552409x12567397529106>.
- . 2014. "Selecting Children: The Ethics of Reproductive Genetic Engineering." In *Ethics and Emerging Technologies*, edited by Ronald Sandler. London: Palgrave Macmillan.
- , ed. 2020a. *Ethics of Artificial Intelligence*. Oxford University Press. <https://doi.org/10.1093/oso/9780190905033.001.0001>.
- . 2020b. "A Short Introduction to the Ethics of AI." In *Ethics of Artificial Intelligence*, edited by S. Matthew Liao. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190905033.001.0001>.
- . 2020c. "The Moral Status and Rights of Artificial Intelligence." In *Ethics of Artificial Intelligence*, edited by S. Matthew Liao. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190905033.001.0001>.
- Lichtenstein, Aharon. 2002. "The Human and Social Factor in Halakha." *Tradition* 36 (1). <https://traditiononline.org/the-human-and-social-factor-in-halakha/>.
- . 2004a. "Does Judaism Recognize an Ethic Independent of Halakhah?" In *Leaves of Faith - Vol. 2*. NJ: KTAV.
- . 2004b. *Leaves of Faith*. Vol. 2. NJ: KTAV.
- . 2016. "Human Dignity in Halakha." *VBM*. Yeshivat Har Etzion. December 25,

2016. <https://www.etzion.org.il/en/halakha/studies-halakha/philosophy-halakha/human-dignity-halakha>.
- Liebes, Yehuda. 1991. "Golem Is Numerically Hochmah." *Kiryat Sefer* 63. <https://liebes.huji.ac.il/yehudaliebes/files/golem.pdf>.
- Lin, Patrick. 2012. "Introduction." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George Bekey. MIT Press.
- Lin, Patrick, Keith Abney, and George A Bekey, eds. 2012. *Robot Ethics : The Ethical and Social Implications of Robotics*. MIT Press.
- , eds. 2017. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press.
- Locke, John. 1690. *An Essay Concerning Human Understanding*. The Project Gutenberg eBook. <https://www.gutenberg.org/files/10615/10615-h/10615-h.htm>.
- Loike, John D. 2000. "Is a Human Clone a Golem?" *The Torah U-Madda Journal* 9: 236–44. <https://www.jstor.org/stable/40914661>.
- Loike, John D., and Moshe D. Tendler. 2003. "Ma Adam Va-Teda-Ehu: Halakhic Criteria for Defining Human Beings." *Tradition* 37 (2). <https://traditiononline.org/ma-adam-va-teda-ehu-halakhic-criteria-for-defining-human-beings/>.
- . 2007. "Ethical Dilemmas in Stem Cell Research: Human-Animal Chimeras." *Tradition: A Journal of Orthodox Jewish Thought* 40 (4): 28–49. <http://www.jstor.org/stable/23263518>.
- . 2014. "Tampering with the Genetic Code of Life: Comparing Secular and Halakhic Ethical Concerns." *Hakira* 18. <https://hakirah.org/Vol18LoikeTendler.pdf>.
- Lori, Dajose. 2021. "How to Read a Jellyfish's Mind." *California Institute of Technology*, November 24, 2021. <https://www.caltech.edu/about/news/how-to-read-a-jellyfishs-mind>.
- Lorrimar, Victoria. 2017. "Human Uniqueness and Technology: Are We Co-Creators with God?" In *Issues in Science and Theology: Are We Special?*, edited by Michael Fuller, Dirk Evers, Anne Runehov, and Knut-Willy Saether, 169–79. Springer. https://doi.org/10.1007/978-3-319-62124-1_12.
- Louden, Robert B. 1986. "Kant's Virtue Ethics." *Philosophy* 61 (238): 473–89. <https://doi.org/10.1017/s0031819100061246>.
- Lumbreras, Sara. 2018. "Strong Artificial Intelligence and Imago Hominis." In *Issues in*

- Science and Theology: Are We Special?*, edited by Michael Fuller, Dirk Evers, Anne Runehov, and Knut-Willy Sæther. Springer.
- Luzzatto, Moshe Hayyim. (1734) 1983. *The Way of God (Derech Hashem)*. Translated by Aryeh Kaplan. Fourth. Jerusalem: Feldheim Publishers.
- Macdonald, Margaret. 1953. "Linguistic Philosophy and Perception." *Philosophy* 28 (107): 311–24. <https://www.jstor.org/stable/3748146>.
- Mackenzie, Robin. 2018. "Sexbots: Customizing Them to Suit Us versus an Ethical Duty to Created Sentient Beings to Minimize Suffering." *Robotics* 7 (4): 70. <https://doi.org/10.3390/robotics7040070>.
- Maimonides, Moses. (1199) 1872. "Translation of an Epistle Addressed by R. Moses Maimonides to R. Shmuel Ibn Tibbon." In *Miscellany of Hebrew Literature: Volume I*. London: N. Trubner.
- . (1190) 1956. *The Guide for the Perplexed*. Translated by Michael Friedlander. New York: Dover.
- . (1190) 1963. *The Guide of the Perplexed*. Translated by Shlomo Pines. Chicago: University of Chicago Press.
- . 1981. *Mishnah Torah: The Book of Knowledge*. Translated by Moses Hyamson. Jerusalem: Feldheim.
- . (1168) 1984. *The Commandments*. Translated by Charles Ber Chavel. London: Soncino Press.
- Margonelli, Lisa. 2023. "The Obligations of Knowledge." *Issues in Science and Technology*, January 11, 2023. <https://issues.org/obligations-of-knowledge-margonelli/>.
- Marr, Bernard. 2017. "The 4 Ds of Robotization: Dull, Dirty, Dangerous and Dear." *Forbes*. October 16, 2017. <https://www.forbes.com/sites/bernardmarr/2017/10/16/the-4-ds-of-robotization-dull-dirty-dangerous-and-dear/?sh=79eec3e03e0d>.
- Maslow, Abraham H. 1943. "A Theory of Human Motivation." *Psychological Review* 50 (4): 370–96. <https://doi.org/10.1037/h0054346>.
- Mayor, Adrienne. 2018. *Gods and Robots: The Ancient Quest for Artificial Life*. Princeton Princeton Univ Press.
- McCarthy, John, Marvin Minsky, and Claude Shannon. 1955. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>.
- McCorduck, Pamela. (1979) 2004. *Machines Who Think*. CRC Press.

- McDermott, Drew. 2011. "What Matters to a Machine?" In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- McFadden, Johnjoe. 2020. "Integrating Information in the Brain's EM Field: The Cemi Field Theory of Consciousness." *Neuroscience of Consciousness* 2020 (1). <https://doi.org/10.1093/nc/niaa016>.
- Mendelssohn, Moses. (1782) 2017. *Jerusalem*. Translated by Jonathan Bennett. Early Modern Philosophy. <https://www.earlymoderntexts.com/assets/pdfs/mendelssohn1782.pdf>.
- Mercer, Calvin. 2015. "Whole Brain Emulation Requires Enhanced Theology, and a 'Handmaiden.'" *Theology and Science* 13 (2): 175–86. <https://doi.org/10.1080/14746700.2015.1023527>.
- Metzinger, Thomas. 2003. *Being No One*. Cambridge, MA: MIT Press.
- Metzler, Theodore. 2007. "Viewing Assignment of Moral Status to Service Robots from the Theological Ethics of Paul Tillich: Some Hard Questions." In *AAAI Workshop Technical Report WS-07-07*, 15–20. Menlo Park, California: The AAAI Press. <https://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-004.pdf>.
- Mill, John Stuart. (1859) 2011. *On Liberty*. Project Gutenberg. <https://www.gutenberg.org/cache/epub/34901/pg34901-images.html>.
- . (1863) 2017. *Utilitarianism*. Edited by Jonathan Bennett. Early Modern Texts. <https://www.earlymoderntexts.com/assets/pdfs/mill1863.pdf>.
- Miller, Arthur I. 2020. *Artist in the Machine: The World of AI-Powered Creativity*. MIT Press.
- Miller, Keith W. 2010. "It's Not Nice to Fool Humans." *IT Professional* 12 (1): 51–52. <https://doi.org/10.1109/mitp.2010.32>.
- Miller, Lantz Fleming. 2017. "Responsible Research for the Construction of Maximally Humanlike Automata: The Paradox of Unattainable Informed Consent." *Ethics and Information Technology* 22 (4): 297–305. <https://doi.org/10.1007/s10676-017-9427-3>.
- Minsky, Marvin. 1985. *The Society of Mind*. New York: Simon And Schuster.
- Mitcham, Carl, and Helen Nissenbaum. 1998. "Technology and Ethics." <https://doi.org/10.4324/9780415249126-L102-1>.
- Moor, James H. 1995. "Is Ethics Computable?" *Metaphilosophy* 26 (1/2): 1–21. <https://www.jstor.org/stable/24439044>.
- . 2011. "The Nature, Importance, and Difficulty of Machine Ethics." In *Machine*

- Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge: Harvard University Press.
- Moreham, N. A. 2008. "The Right to Respect for Private Life in the European Convention on Human Rights: A Re-Examination." *European Human Rights Law Review* 1 (January). <https://ssrn.com/abstract=2383507>.
- Moreland, James P. 2009. *The Recalcitrant Imago Dei: Human Persons and the Failure of Naturalism*. London: SCM Press.
- Morin, Alain. 2006. "Levels of Consciousness and Self-Awareness: A Comparison and Integration of Various Neurocognitive Views." *Consciousness and Cognition* 15 (2): 358–71. <https://doi.org/10.1016/j.concog.2005.09.006>.
- Motzkin, Aryeh Leo. 2012. "On the Limitations of Human Knowledge." In *Philosophy and the Jewish Tradition*, edited by Yehuda Halper. Leiden: Brill.
- Muehlhauser, Luke, and Louie Helm. 2012. "The Singularity and Machine Ethics." In *Singularity Hypotheses: A Scientific and Philosophical Assessment*, edited by Amnon H. Eden and James H. Moor. Berlin: Springer. https://www.researchgate.net/publication/233935716_The_Singularity_and_Machine_Ethics.
- Muller, Vincent C., ed. 2013. *Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemology and Rational Ethics*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-31674-6>.
- . 2016. "Autonomous Killer Robots Are Probably Good News." In *Drones and Responsibility*, edited by Ezio Di Nucci and Filippo Santonio de Si, 67–81. London: Ashgate. <https://doi.org/10.4324/9781315578187-4>.
- . 2021. "Is It Time for Robot Rights? Moral Status in Artificial Entities." *Ethics and Information Technology*. <https://doi.org/10.1007/s10676-021-09596-w>.
- Musial, Maciej. 2017. "Designing (Artificial) People to Serve – the Other Side of the Coin." *Journal of Experimental & Theoretical Artificial Intelligence* 29 (5): 1087–97. <https://doi.org/10.1080/0952813x.2017.1309691>.
- . 2022. "Can We Design Artificial Persons without Being Manipulative?" *AI & Society*. <https://doi.org/10.1007/s00146-022-01575-z>.
- Nachmanides, Moses. 1976. *Ramban (Nachmanides): Commentary on the Torah*. Translated by Charles Chavel. Vol. Deuteronomy. New York: Shilo Publishing House.

- Nadler, Steven. 2021. "Maimonides on Human Perfection and the Love of God." In *Maimonides' Guide of the Perplexed - a Critical Guide*, edited by Daniel Frank and Aaron Segal, 266–85. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108635134.021>.
- Nagel, Thomas. 1974. "What Is It like to Be a Bat?" *The Philosophical Review* 83 (4): 435–50. <https://doi.org/10.2307/2183914>.
- Natsoulas, Thomas. 1991. "The Concept of Consciousness: The Personal Meaning." *Journal for the Theory of Social Behaviour* 21 (3): 339–67. <https://doi.org/10.1111/j.1468-5914.1991.tb00200.x>.
- Navon, Chaim. 2007. "The Image of God." *Alei Etzion* 15. <https://www.gush.net/alei/15-09cn-image%20final.rtf>.
- Navon, Mois. 2008. "Halacha, Ethics and Aesthetics." *Everett Journal of Jewish Ethics* 1. <http://www.divreinavon.com/pdf/HalachaEthicsAesthetics.pdf>.
- . 2014. "The Binding of Isaac." *Hakirah: The Flatbush Journal of Jewish Law and Thought* 17: 233–56. <https://hakirah.org/Vol17Navon.pdf>.
- . 2021. "The Virtuous Servant Owner—a Paradigm Whose Time Has Come (Again)." *Frontiers in Robotics and AI* 8 (September). <https://doi.org/10.3389/frobt.2021.715849>.
- . 2023. "Autonomous Weapons Systems and Battlefield Dignity – a Jewish Perspective." In *"Alexa, How Do You Feel about Religion?" Technology, Digitization and Artificial Intelligence in the Focus of AI*, edited by Anna Puzio, Hendrik Klinge, and Nicole Kunkel. Darmstadt: WBG.
- . 2024a. "A Jewish Theological Perspective on Technology (Orthodox)." In *St Andrews Encyclopaedia of Theology*, edited by Brendan N. Wolfe et al. University of St. Andrews. <https://www.saet.ac.uk/Judaism/AJewishTheologicalPerspectiveonTechnologyOrthodox>.
- . 2024b. "The Trolley Problem Just Got Digital - Ethical Dilemmas in Programming Autonomous Vehicles." *B.D.D. - Bekhol Derakbekha Daehu* 38. http://www.divreinavon.com/pdf/EthicalDilemmasinProgrammingAutonomousVehicles_MoisNavon.pdf.
- Neely, Erica L. 2014. "Machines and the Moral Community." *Philosophy & Technology* 27 (1): 97–111. <https://doi.org/10.1007/s13347-013-0114-y>.
- Nietzsche, Friedrich. (1887) 2001. *The Gay Science*. Edited by Bernard Arthur. Translated

- by Josefine Nauckhoff. Cambridge: Cambridge University Press.
- Noguerol, Teodoro Martin, Felix Paulano-Godino, María Teresa Martín-Valdivia, Christine O. Menias, and Antonio Luna. 2019. “Strengths, Weaknesses, Opportunities, and Threats Analysis of Artificial Intelligence and Machine Learning Applications in Radiology.” *Journal of the American College of Radiology* 16 (9, Part B): 1239–47. <https://doi.org/10.1016/j.jacr.2019.05.047>.
- Nørskov, Marco. 2017. “Technological Dangers and the Potential of Human–Robot Interaction.” In *Social Robots*, edited by Marco Nørskov. Taylor & Francis.
- Nyholm, Sven. 2020. *Humans and Robots : Ethics, Agency, and Anthropomorphism*. NY: Rowman & Littlefield Publishing Group.
- Oppy, Graham, and David Dowe. 2021. “The Turing Test.” Edited by Edward N. Zalta. Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/entries/turing-test/>.
- Ortony, Andrew. 1975. “Why Metaphors Are Necessary and Not Just Nice.” *Educational Theory* 25 (1): 45–53. <https://doi.org/10.1111/j.1741-5446.1975.tb00666.x>.
- Parthemore, Joel, and Blay Whitby. 2013. “What Makes Any Agent a Moral Agent?” *International Journal of Machine Consciousness* 05 (02): 105–29. <https://doi.org/10.1142/S1793843013500017>.
- Penrose, Roger. 1991. *The Emperor’s New Mind : Concerning Computers, Minds and the Laws of Physics*. NY: Penguin.
- . 1995. “Beyond the Doubting of a Shadow.” *Psyche* 2: 89–129. <https://journalpsyche.org/files/0xaa2c.pdf>.
- Peters, Ted. 2005. “The Soul of Trans-Humanism.” *Dialog: A Journal of Theology* 44 (4): 381–95. <https://doi.org/10.1111/j.0012-2033.2005.00282.x>.
- Petersen, Steve. 2007. “The Ethics of Robot Servitude.” *Journal of Experimental & Theoretical Artificial Intelligence* 19 (1): 43–54. <https://doi.org/10.1080/09528130601116139>.
- . 2012. “Designing People to Serve.” In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George Bekey. MIT Press.
- . 2017. “Is It Good for Them Too? Ethical Concern for the Sexbots.” In *Robot Sex: Social Implications and Ethical*, edited by John Danaher and Neil McArthur, 155–71. Cambridge, USA: MIT Press.
- Picard, Rosiland. 1995. “Affective Computing.” Cambridge, MA: M.I.T Media

- Laboratory Perceptual Computing Section Technical Report No. 321.
<https://affect.media.mit.edu/pdfs/95.picard.pdf>.
- Pines, Shlomo. 1979. "The Limitations of Human Knowledge according to Al-Farabi, Ibn Bajja, and Maimonides." In *Studies in Medieval Jewish History and Literature*, edited by Isadore Twersky, 1:82–109. MA: Harvard University Press.
- Plato. (399BC) 1997. *Plato: Complete Works*. Edited by John M. Cooper and D. S. Hutchinson. Indianapolis, Ind.: Hackett Publishing Company.
- Prescott, Tony J. 2017. "Robots Are Not Just Tools." *Connection Science* 29 (2): 142–49.
<https://doi.org/10.1080/09540091.2017.1279125>.
- Pruss, Alexander R. 2009. "Artificial Intelligence and Personal Identity." *Faith and Philosophy* 26 (5): 487–500. <https://doi.org/10.5840/faithphil200926550>.
- Purves, Duncan, Ryan Jenkins, and Bradley J. Strawser. 2015. "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons." *Ethical Theory and Moral Practice* 18 (4): 851–72. <https://doi.org/10.1007/s10677-015-9563-y>.
- Putman, Hilary. 1964. "Robots: Machines or Artificially Created Life?" *The Journal of Philosophy* 61 (21). <https://doi.org/10.2307/2023045>.
- Rabinovitch, Nahum. 2003. "The Way of Torah." *The Edab Journal* 3 (1).
<https://library.yctorah.org/files/2016/09/The-Way-of-Torah.pdf>.
- Rachels, Stuart, and James Rachels, eds. 2015. *The Right Thing to Do: Basic Readings in Moral Philosophy*. New York: Mcgraw-Hill Education.
- Rakover, Nahum. 2002. "Cloning - Competition with God? (HEBREW)." *Shana BeShana*, 105–17.
- . 2011. "Man as a Synthesis of Body and Spirit: A Jewish Perspective." In *The Jewish Law Annual*, edited by Hanina Ben-Menahem and Berachyahu Lifshitz. NY: Routledge for The Institute For Research In Jewish Law Faculty Of Law, The Hebrew University Of Jerusalem.
- Ravitsky, Aviram. 2014. "The Balanced Path and the Path of Asceticism: The Unity of Maimonides' Ethics." *Tradition* 47 (1): 28–47.
- Rawidowicz, Simon. 1974. *Studies in Jewish Thought*. Philadelphia: Jewish Publication Society.
- Redstone, Josh. 2014. "Making Sense of Empathy with Social Robots." In *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy 2014*, edited by Johanna Seibt, Marco Nørskov, and Raul Hakli, 171–78. Amsterdam: IOS Press.
- Reeve, CDC. 2014. "Beginning and Ending with Eudaimonia." In *The Cambridge*

- Companion to Aristotle's Nicomachean Ethics*, edited by Ronald Polansky, 14–33. Cambridge University Press.
- Reeves, Byron, Jeff Hancock, and Xun Liu. 2020. “Social Robots Are like Real People: First Impressions, Attributes, and Stereotyping of Social Robots.” *Technology, Mind, and Behavior* 1 (1). <https://doi.org/10.1037/tmb0000018>.
- Reining, Stefan. 2020. “Revisiting the Dancing-Qualia Argument for Computationalism.” In *Artificial Intelligence: Reflections in Philosophy, Theology, and the Social Sciences*, edited by Benedikt Paul Gocke and Astrid Rosenthal-Von der Putten. Boston: Brill.
- Richards, Neil M., and William D. Smart. 2016. “How Should the Law Think about Robots?” In *Robot Law*, edited by Ryan Calo, A. Michael Froomkin, and Ian Kerr, 3–24. MA: Edward Elgar. DOI 10.4337/9781783476732.
- Richardson, Kathleen. 2015. *An Anthropology of Robots and AI: Annihilation Anxiety and Machines*. New York, NY: Routledge.
- . 2016. “Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines.” *IEEE Technology and Society Magazine* 35 (2): 46–53. <https://doi.org/10.1109/mts.2016.2554421>.
- Rodogno, Raffaele. 2016. “Robots and the Limits of Morality.” In *Social Robots: Boundaries, Potential, Challenges*, edited by Marco Nørskov. Routledge.
- Rohlf, Michael. 2023. “Immanuel Kant.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta and Uri Nodelman, Fall 2023. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2023/entries/kant/>.
- Rosenfeld, Azriel. 1966. “Religion and the Robot.” *Tradition Online* 8 (3). <https://traditiononline.org/religion-and-the-robot/>.
- . 1977. “Human Identity: Halakhic Issues.” *Tradition* 16 (3): 58–74. <https://www.jstor.org/stable/23258438>.
- Rubin, Charles. 2013. “The Golem and the Limits of Artifice.” *The New Atlantis*, 2013. <https://www.thenewatlantis.com/publications/the-golem-and-the-limits-of-artifice>.
- Russell, Bertrand. 1912. “The Philosophy of Bergson.” *The Monist* 22 (3): 321–47. <https://www.jstor.org/stable/27900381>.
- Samsonovich, Alexei. 2010. “Biologically Inspired Cognitive Architectures for AI.” BICA.AI. 2010. <https://bica.ai/wp-content/architectures.xls>.
- Sandberg, Anders. 2013. “Feasibility of Whole Brain Emulation.” In *Philosophy and Theory of Artificial Intelligence*, edited by Vincent C. Muller. Berlin, Heidelberg: Springer.

- <https://doi.org/10.1007/978-3-642-31674-6>.
- Sandberg, Anders, and Nick Bostrom. 2008. "Whole Brain Emulation: A Roadmap." Oxford University: Future of Humanity Institute. <https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf>.
- Sandler, Ronald L., ed. 2014. *Ethics and Emerging Technologies*. London: Palgrave Macmillan UK. <https://doi.org/10.1057/9781137349088>.
- Saperstein, Marc. 2012. *Jewish Preaching in Times of War 1800-2001*. Oxford Littman Library of Jewish Civilization.
- Sartre, Jean-Paul. (1947) 2007. *Existentialism Is a Humanism*. New Haven: Yale University Press.
- Savulescu, Julian. 2001. "Procreative Beneficence: Why We Should Select the Best Children." *Bioethics* 15 (5-6): 413–26. <https://doi.org/10.1111/1467-8519.00251>.
- Savulescu, Julian, and Ingmar Persson. 2012. "Moral Enhancement, Freedom, and the God Machine." *The Monist* 95 (3): 399–421. <https://www.jstor.org/stable/42751159>.
- Schafer, Peter. 1995. "The Magic of the Golem: The Early Development of the Golem Legend." *Journal of Jewish Studies* 46 (1-2): 249–61. <https://doi.org/10.18647/1802/jjs-1995>.
- Scheler, Gabriele. 2023. "Sketch of a Novel Approach to a Neural Model." *PREPRINT*. <https://doi.org/10.13140/RG.2.2.25578.39368>.
- Scheutz, Matthias. 2014a. "Artificial Emotions and Machine Consciousness." In *The Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William M Ramsey. Cambridge, United Kingdom: Cambridge University Press.
- . 2014b. "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots." In *Robot Ethics: The Ethical and Social Implications of Robotics*. London, England: MIT Press.
- Schneider, Susan. 2020. "How to Catch an AI Zombie." In *Ethics of Artificial Intelligence*, edited by S. Matthew Liao. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190905033.001.0001>.
- Scholem, Gershom. 1966. "The Golem of Prague & the Golem of Rehovoth." *Commentary Magazine*, January 1, 1966. <https://www.commentary.org/articles/gershom-scholem/the-golem-of-prague-the-golem-of-rehovoth/>.
- . 1969. *On the Kabbalah and Its Symbolism*. Translated by Ralph Manheim. NY:

Schocken Books.

- Schwab, Klaus. 2017. *The Fourth Industrial Revolution*. New York: Crown Business.
- Schwarzschild, Steven S. 1990. *The Pursuit of the Ideal: Jewish Writings of Steven Schwarzschild*. Edited by Menachem Marc Kellner. Albany: State University Of New York Press.
- . 2007. “Justice.” In *Encyclopaedia Judaica*, edited by Fred Skolnik and Michael Berenbaum, 2nd ed. Detroit: Macmillan Reference in Ass. with Keter.
- Schwitzgebel, Eric, and Mara Garza. 2015. “A Defense of the Rights of Artificial Intelligences.” *Midwest Studies in Philosophy* 39 (1): 98–119. <https://doi.org/10.1111/misp.12032>.
- . 2020. “Designing AI with Rights, Consciousness, Self- Respect, and Freedom.” In *Ethics of Artificial Intelligence*, edited by S. Matthew Liao. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190905033.001.0001>.
- Searle, John. 1980. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3 (03): 417–57. <https://doi.org/10.1017/s0140525x00005756>.
- . 2007. “Biological Naturalism.” In *The Blackwell Companion to Consciousness*, edited by Max Velmans and Susan Schneider. Malden, MA: Blackwell Publishing.
- Seele, Peter, Claus Dierksmeier, Reto Hofstetter, and Mario D. Schultz. 2019. “Mapping the Ethicality of Algorithmic Pricing: A Review of Dynamic and Personalized Pricing.” *Journal of Business Ethics* 170 (December): 697–719. <https://doi.org/10.1007/s10551-019-04371-w>.
- Shalev-Shwartz, Shai, Shaked Shammah, and Amnon Shashua. 2018. “On a Formal Model of Safe and Scalable Self-Driving Cars.” ArXiv.org. October 27, 2018. <https://doi.org/10.48550/arXiv.1708.06374>.
- . 2020. “On the Ethics of Building AI in a Responsible Manner.” *ArXiv (Cornell University)*, January. <https://doi.org/10.48550/arxiv.2004.04644>.
- Shanahan, Murray. 2010. *Embodiment and the Inner Life*. New York: Oxford University Press.
- . 2016. “Beyond Humans, What Other Kinds of Minds Might Be out There?” *Aeon*, October 19, 2016. <https://aeon.co/essays/beyond-humans-what-other-kinds-of-minds-might-be-out-there>.
- Shapira, Haim. 2018. “The Virtue of Mercy according to Maimonides: Ethics, Law, and Theology.” *Harvard Theological Review* 111 (4): 559–85. <https://doi.org/10.1017/s0017816018000275>.

- Sharkey, Amanda. 2018. "Autonomous Weapons Systems, Killer Robots and Human Dignity." *Ethics and Information Technology* 21 (December). <https://doi.org/10.1007/s10676-018-9494-0>.
- Sharkey, Amanda, and Noel Sharkey. 2010. "Granny and the Robots: Ethical Issues in Robot Care for the Elderly." *Ethics and Information Technology* 14 (1): 27–40. <https://doi.org/10.1007/s10676-010-9234-6>.
- Shatz, David. 1996. "Beyond Obedience: Walter Wurzbürger's Ethics of Responsibility." *Tradition* 30 (2): 74–95. <https://traditiononline.org/beyond-obedience-walter-wurzburgers-ethics-of-responsibility/>
- . 2005. "Maimonides' Moral Theory." In *The Cambridge Companion to Maimonides*, edited by Kenneth Seeskin, 167–92. Cambridge: Cambridge University Press. <https://doi.org/10.1017/ccol0521819741>.
- . 2012. "Ethical Theories in the Orthodox Movement." In *The Oxford Handbook of Jewish Ethics and Morality*, edited by Elliot N. Dorff and Jonathan K. Crane. Oxford University Press. 10.1093/oxfordhb/9780199736065.013.0015.
- Sherman, Nancy. 1997. *Making a Necessity of Virtue*. Cambridge: Cambridge University Press.
- Sherwin, Byron L. 2007. "Golems in the Biotech Century." *Zygon* 42 (1): 133–44. <https://doi.org/10.1111/j.1467-9744.2006.00810.x>.
- Shilat, Yitzhak. 1998. "Genetic Reproduction in Light of Halakha (HEBREW)." *Techumin* 18.
- Shmalo, Gamliel. 2012. "Orthodox Approaches to Biblical Slavery." *The Torah U-Madda Journal* 16. <https://www.jstor.org/stable/23596054>.
- Signorelli, Camilo Miguel. 2018. "Can Computers Become Conscious and Overcome Humans?" *Frontiers in Robotics and AI* 5 (October). <https://doi.org/10.3389/frobt.2018.00121>.
- Singer, Peter. 2009. *Animal Liberation: The Definitive Classic of the Animal Movement*. New York, N.Y.: Harper Collins.
- Sinnott-Armstrong, Walter. 2023. "Consequentialism." Edited by Edward N. Zalta and Uri Nodelman. *Stanford Encyclopedia of Philosophy*. Stanford University. <https://plato.stanford.edu/entries/consequentialism>.
- Slack, Gordy. 2023. "What DALL-E Reveals about Human Creativity." Stanford HAI. January 17, 2023. <https://hai.stanford.edu/news/what-dall-e-reveals-about-human-creativity>.

- Sloman, Aaron. 2013. "Aaron Sloman Absolves Turing of the Mythical Turing Test." In *Alan Turing: His Work and Impact*, edited by S. Barry Cooper and Jan Van Leeuwen. Waltham, MA: Elsevier. <https://www.cs.bham.ac.uk/research/projects/cogaff/sloman-turing-test.pdf>.
- Smids, Jilles. 2020. "Danaher's Ethical Behaviourism: An Adequate Guide to Assessing the Moral Status of a Robot?" *Science and Engineering Ethics* 26 (5): 2849–66. <https://doi.org/10.1007/s11948-020-00230-4>.
- Smirnova, Lena, Brian Caffo, David Gracias, Qi Huang, Itzy Morales Pantoja, Bohao Tang, Donald Zack, et al. 2023. "Organoid Intelligence (OI): The New Frontier in Biocomputing and Intelligence-In-a-Dish." *Frontiers in Science* 1. <https://doi.org/10.3389/fsci.2023.1017235>.
- Smith, David Harris, and Guido Schillaci. 2021. "Why Build a Robot with Artificial Consciousness?" *Frontiers in Psychology* 12 (April). <https://doi.org/10.3389/fpsyg.2021.530560>.
- Soloveitchik, Joseph. 1978a. "Majesty and Humility." *Tradition* 17 (2): 25–37.
- . 1978b. "Redemption, Prayer, Talmud Torah." *Tradition* 17 (2). <https://traditiononline.org/redemption-prayer-talmud-torah/>.
- . 1983. *Halakhic Man*. Philadelphia: Jewish Publication Society.
- . (1979) 1993. *Reflections of the Rav: Lessons in Jewish Thought - Vol. 1*. Edited by Abraham R. Besdin. NJ: KTAV.
- . 2000. *Fate and Destiny: From Holocaust to the State of Israel*. Hoboken, N.J.: KTAV.
- . 2003. *Worship of the Heart: Essays on Jewish Prayer*. Edited by Shalom Carmy. NJ: KTAV.
- . 2006. *Festival of Freedom: Essays on Pesah and the Haggadah*. Edited by Joel B Wolowelsky and Reuven Ziegler. NJ: KTAV.
- . 2008a. *Abraham's Journey: Reflections on the Life of the Founding Patriarch*. Edited by David Shatz, Joel B. Wolowelsky, and Reuven Ziegler. NJ: KTAV.
- . 2008b. *And from There You Shall Seek*. Jersey City, NJ: KTAV.
- . 2012. *Halakhic Positions of Rabbi Joseph B. Soloveitchik*. Edited by Aharon Ziegler. KTAV.
- . (1965) 2012. *The Lonely Man of Faith*. Jerusalem: Maggid. <https://traditiononline.org/wp-content/uploads/2019/09/LMOF.pdf>.
- . 2016. *Maimonides - between Philosophy and Halakbah: Rabbi Joseph B. Soloveitchik's Lectures on the Guide of the Perplexed at the Bernard Revel Graduate School (1950-51)*.

- Edited by Lawrence J Kaplan. New York: Urim Publications.
- . 2017. *Halakbic Morality: Essays on Ethics and Masorah*. Edited by Joel B. Wolowelsky and Reuven Ziegler. CT: Maggid Books.
- Soraker, Johnny Hartz. 2014. “Continuities and Discontinuities between Humans, Intelligent Machines, and Other Entities.” *Philosophy & Technology* 27 (1): 31–46. <https://doi.org/10.1007/s13347-013-0132-9>.
- Sparrow, Robert. 2007. “Killer Robots.” *Journal of Applied Philosophy* 24 (1): 62–77. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>.
- . 2016. “Robots and Respect: Assessing the Case against Autonomous Weapon Systems.” *Ethics & International Affairs* 30 (1): 93–116. <https://doi.org/10.1017/s0892679415000647>.
- . 2017. “Robots, Rape, and Representation.” *International Journal of Social Robotics* 9 (4): 465–77. <https://doi.org/10.1007/s12369-017-0413-z>.
- . 2020. “Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might It Be Explained?” *International Journal of Social Robotics* 13 (1): 23–29. <https://doi.org/10.1007/s12369-020-00631-2>.
- . 2021. “Why Machines Cannot Be Moral.” *AI & Society* 36 (3): 685–93. <https://doi.org/10.1007/s00146-020-01132-6>.
- Sparrow, Robert, and Linda Sparrow. 2006. “In the Hands of Machines? The Future of Aged Care.” *Minds and Machines* 16 (2): 141–61. <https://doi.org/10.1007/s11023-006-9030-6>.
- Spero, Shubert. 2003. “Rabbi Joseph Dov Soloveitchik and the Role of the Ethical.” *Modern Judaism* 23 (1): 12–31. <https://www.jstor.org/stable/1396556>.
- . 2016. “The Good, the Right, and the Morality of Judaism.” *The Torah U-Madda Journal* 17: 202–17. <https://www.jstor.org/stable/26203067>.
- Steinberg, Avraham. 2000. “Human Cloning—Scientific, Moral and Jewish Perspectives.” *The Torah U-Madda Journal* 9: 199–206. <https://www.jstor.org/stable/40914655>.
- Steinberg, Avraham, and John D. Loike. 1998. “Human Cloning: Scientific, Ethical and Jewish Perspectives.” *Assia-Jewish Medical Ethics* 3 (2): 11–19. <https://pubmed.ncbi.nlm.nih.gov/11657947/>.
- Stern, Josef. 2013. *The Matter and Form of Maimonides’ Guide*. MA: Harvard University Press.
- Straiton, Jenny. 2019. “Grow Your Own Brain.” *BioTechniques* 66 (3): 108–12.

- <https://doi.org/10.2144/btn-2019-0019>.
- Strauss, Leo. 2013. *Leo Strauss on Maimonides: The Complete Writings*. Edited by Kenneth Hart Green. Chicago: The University Of Chicago Press.
- Susser, Daniel. 2013. "Artificial Intelligence and the Body: Dreyfus, Bickhard, and the Future of AI." In *Philosophy and Theory of Artificial Intelligence*, edited by Vincent C. Muller. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-31674-6>.
- Swinburne, Richard. 2019. *Are We Bodies or Souls?* Oxford: Oxford University Press.
- Tallis, Raymond. 2012. *Aping Mankind: Neuromania, Darwinitis and the Misrepresentation of Humanity*. Durham: Acumen.
- Tavani, Herman. 2018. "Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights." *Information* 9 (4): 73. <https://doi.org/10.3390/info9040073>.
- Thaler, Richard H, and Cass R Sunstein. 2021. *Nudge: Improving Decisions about Money, Health, and the Environment*. New York: Penguin Books.
- The Medical Futurist. 2018. "The Top 12 Social Companion Robots." The Medical Futurist. July 31, 2018. <https://medicalfuturist.com/the-top-12-social-companion-robots/>.
- Thorstensen, Erik. 2017. "Creating Golems: Uses of Golem Stories in the Ethics of Technologies." *NanoEthics* 11 (2): 153–68. <https://doi.org/10.1007/s11569-016-0279-9>.
- Tirosh-Samuels, Hava. 2012. "Transhumanism as a Secularist Faith." *Zygon* 47 (4): 710–34. <https://doi.org/10.1111/j.1467-9744.2012.01288.x>.
- Tocqueville, Alexis de. (1835) 2013. *Democracy in America, Part I*. Translated by Henry Reeve. www.gutenberg.org. Project Gutenberg. <https://www.gutenberg.org/files/815/815-h/815-h.htm>.
- Toivakainen, Niklas. 2015. "Machines and the Face of Ethics." *Ethics and Information Technology* 18 (4): 269–82. <https://doi.org/10.1007/s10676-015-9372-y>.
- Tollon, Fabio. 2020. "The Artificial View: Toward a Non-Anthropocentric Account of Moral Patiency." *Ethics and Information Technology* 23 (June): 147–55. <https://doi.org/10.1007/s10676-020-09540-4>.
- Tonkens, Ryan. 2012. "Out of Character: On the Creation of Virtuous Machines." *Ethics and Information Technology* 14 (2): 137–49. <https://doi.org/10.1007/s10676-012-9290-1>.

- Tononi, Giulio, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. "Integrated Information Theory: From Consciousness to Its Physical Substrate." *Nature Reviews Neuroscience* 17 (7): 450–61. <https://doi.org/10.1038/nrn.2016.44>.
- Torrance, Steve. 2008. "Ethics and Consciousness in Artificial Agents." *AI & Society* 22 (4): 495–521. <https://doi.org/10.1007/s00146-007-0091-8>.
- . 2011. "Machine Ethics and the Idea of a More-Than-Human Moral World." In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- . 2013. "Artificial Consciousness and Artificial Ethics: Between Realism and Social Relationism." *Philosophy & Technology* 27 (1): 9–29. <https://doi.org/10.1007/s13347-013-0136-5>.
- Turing, A. M. 1950. "Computing Machinery and Intelligence." *Mind* LIX (236): 433–60. <https://doi.org/10.1093/mind/lix.236.433>.
- Turkle, Sherry. 2011a. *Alone Together: Why We Expect More From Technology and Less from Each Other*. New York: Basic Books.
- . 2011b. "Authenticity in the Age of Digital Companions." In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson. NY: Cambridge University Press.
- Twersky, Isadore. 1980. *Introduction to the Code of Maimonides*. New Haven: Yale University Press.
- UCSB Scientists. 2020. "Which Animals Don't Have Blood?" UCSB Science Line. 2020. <http://scienceline.ucsb.edu/getkey.php?key=4817>.
- Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.
- Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University Of Chicago Press.
- . 2014. "Some Misunderstandings about the Moral Significance of Technology." In *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek, 75–88. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-7914-3_5.
- Veruggio, Gianmarco, and Keith Abney. 2012. "Roboethics: The Applied Ethics for a New Science." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George A Bekey, 347–63. Cambridge, Mass.: MIT Press.

- Vudka, Amir. 2020. "The Golem in the Age of Artificial Intelligence." *NECSUS* 9 (1). <https://doi.org/10.25969/mediarep/14326>.
- Wales, Jordan. 2020. "Empathy and Instrumentalization: Late Ancient Cultural Critique and the Challenge of Apparently Personal Robots." In *Culturally Sustainable Social Robotics: Proceedings of Robo-Philosophy 2020*, edited by Johanna Seibt, Marco Nørskov, and Oliver Santiago Quick, 114–24. Amsterdam: IOS Press.
- Walker, Mark. 2006a. "A Moral Paradox in the Creation of Artificial Intelligence: Mary Poppins 3000s of the World Unite!" In *Human Implications of Human-Robot Interaction*, edited by Ted Metzler. AAAI.
- . 2006b. "Viewing Assignment of Moral Status to Service Robots from the Theological Ethics of Paul Tillich: Some Hard Questions." In *AAAI Workshop Technical Report WS-06-09*, 23–28. Menlo Park, California: The AAAI Press. <https://www.aaai.org/Library/Workshops/2006/ws06-09-005.php>.
- Wallach, Wendell, and Colin Allen. 2010. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.
- Wang, Nyu, and Michael Yuan Tian. 2022. "'Intelligent Justice': Human-Centered Considerations in China's Legal AI Transformation." *AI and Ethics*, August. <https://doi.org/10.1007/s43681-022-00202-3>.
- Warburton, Nigel. 2024. "Philosophical Theories Are like Good Stories: Margaret Macdonald." *Aeon*. <https://aeon.co/essays/philosophical-theories-are-like-good-stories-margaret-macdonald>.
- Warwick, Kevin. 2012. "Robots with Biological Brains." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George Bekey. MIT Press.
- Weiss, Asher. 2003. *Minhat Asher (HEBREW)*. Jerusalem: Machon Minhat Asher.
- Weiss, Raymond L. 1991. *Maimonides' Ethics: The Encounter of Philosophic and Religious Morality*. Chicago: University of Chicago Press.
- Weiss, Raymond L., and Charles E. Butterworth. 1975. *Ethical Writings of Maimonides*. New York: Dover Publications.
- Weiss, Tzachi. 2013. "The Reception of Sefer Yetsirah and Jewish Mysticism in the Early Middle Ages." *The Jewish Quarterly Review* 103 (1): 26–46. <https://www.jstor.org/stable/43298679>.
- Weissenbacher, Alan. 2018. "Moral Enhancement and Deification through Technology?" *Theology and Science* 16 (3): 243–46.

<https://doi.org/10.1080/14746700.2018.1488465>.

- Weizenbaum, Joseph. 1976. *Computer Power and Human Reason*. San Francisco: W.H. Freeman.
- Whang, Oliver. 2023. “‘Consciousness’ in Robots Was Once Taboo. Now It’s the Last Word.” *The New York Times*, January 6, 2023, sec. Science. <http://nytimes.com/2023/01/06/science/robots-artificial-intelligence-consciousness.html>.
- Whitby, Blay. 2008. “Sometimes It’s Hard to Be a Robot: A Call for Action on the Ethics of Abusing Artificial Agents.” *Interacting with Computers* 20 (3): 326–33. <https://doi.org/10.1016/j.intcom.2008.02.002>.
- Wiggers, Kyle. 2022. “The Emerging Types of Language Models and Why They Matter.” TechCrunch. April 28, 2022. <https://techcrunch.com/2022/04/28/the-emerging-types-of-language-models-and-why-they-matter/>.
- Wurzbarger, Walter S. 1994. *Ethics of Responsibility: Pluralistic Approaches to Covenantal Ethics*. Philadelphia; Jerusalem: Jewish Publication Society.
- . 1996. “The Centrality of Creativity in the Thought of Rabbi Joseph B. Soloveitchik.” *Tradition: A Journal of Orthodox Jewish Thought* 30 (4): 219–28. <https://www.jstor.org/stable/23261246>.
- . 2008. *Covenantal Imperatives: Essays by Walter S. Wurzbarger on Jewish Law, Thought and Community*. Edited by Eliezer L Jacobs and Shalom Carmy. Jerusalem: Urim Publications.
- Yampolskiy, Roman V. 2013. “Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach.” In *Philosophy and Theory of Artificial Intelligence*, edited by Vincent C. Muller. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-31674-6>.
- Yancy, George, and Peter Singer. 2015. “Peter Singer: On Racism, Animal Rights and Human Rights.” *Opinionator*. October 8, 2015. <https://opinionator.blogs.nytimes.com/2015/05/27/peter-singer-on-speciesism-and-racism/>.
- Zlatev, Jordan. 2001. “The Epigenesis of Meaning in Human Beings, and Possibly in Robots.” *Minds and Machines* 11: 155–95.
- Zoloth, Laurie. 2008. “Go and Tend the Earth: A Jewish View on an Enhanced World.” *Journal of Law, Medicine & Ethics* 36 (1): 10–25. <https://doi.org/10.1111/j.1748-720x.2008.00233.x>.

6. Appendix: The Moral Status of a Soul

I claim that a humanoid robot with second-order phenomenal consciousness will have the same moral status as a human being. This is because the most widely accepted approach to determining moral status is the ontological approach, which states that if the fundamental ontological characteristic of an entity whose moral status is unfamiliar (e.g., humanoid robot) corresponds to that of an entity whose moral status is familiar (e.g., a human being), then their moral status corresponds as well. In simple terms, if the fundamental characteristic of a robot corresponds to that of a human being, then morally speaking, the robot must be treated as one treats a human being (see Ch. 2: The Virtuous Servant Owner). And as discussed (Ch. 3: To Make a Mind), the fundamental ontological characteristic that defines a human being is second-order phenomenal consciousness (i.e., 2OPC), be it of the mind or the soul.¹

This approach is consistent with the Jewish method of determining moral status, wherein a being with a soul (i.e., 2OPC) would have the moral status of a human being. Such can be seen implicitly in the commentaries on the creation of Adam – i.e., a human being is characterized by having a *nefesh adam* or *neshamah* as exhibited in the capacity for intellect and speech (*deah v'dibbur*).² It can also be seen explicitly in the halachic discussions surrounding the Golem (i.e., a mystically conjured, synthetic humanoid),³ all of which refer to the notions of soul, intellect and speech, in order to classify the Golem's moral status.⁴ To take but one example, R. Shmuel Eidels (Maharsha, San. 65b) rejects the

¹ For the sake of completeness, two important side issues: (1) epistemology: some argue that since it is difficult, if not impossible, to know with certainty if a being has consciousness, we should accord it moral status based on its behavior (see the “appearances camp” in Ch. 2 “The Virtuous Servant Owner”) and (2) unconsciousness: since there are cases where human beings lack consciousness (e.g., anencephalic infants and individuals in an irreversible coma), we should accord moral status if the being in question has the physical basis for the development of consciousness (Liao 2020c). Note that neither of these issues refute consciousness as the defining feature, but rather seek to address anomalies.

² See sources in Ch. 3 “To Make A Mind.”

³ For a thorough examination of the Golem see, e.g., Idel 2019.

⁴ See, e.g., Sheilat Yavetz (2:82); Sidrei Taharot (Ohalot 5); Marit HaAyin (San 65b, s.v. *rava*); Maharsha (Hidushei Aggada, ad loc., s.v. *v'lo hava*), Maharal (San. 65b, s.v. *rava*), R. Tzadok MiLublin (Divrei Halomot

possibility of a particular Golem having human moral status, “due to its lack of the power of the human soul (*koach haneshama*) as expressed in [the capacity for] speech.”⁵

That said, there are those who disagree with granting human moral status to an entity simply because it shares with human beings that defining ontological characteristic, a soul (i.e., 2OPC). R. Tzvi Ashkenazi (Hacham Tzvi 93), anomalously,⁶ argues that only one born of a human mother could have human status; and R. Joseph Rosen (Tzafnat Paneach 2:7) argues that a being created “miraculously” via the Book of Creation (*Sefer Yetzira*) cannot have human status.⁷ Such claims imply that the creature could be killed with impunity, bringing R. Gershon Chanoch Leiner (Sidrei Taharot, Ohalot 5) to exclaim, “How could you kill a creature that has the spiritual life-force and physical form of a man just because it wasn’t born of a mother ... [ay, neither] was Adam HaRishon (i.e., the first man)!” Accordingly, R. Bleich (1998: 82 fn. 66) conjectures that these opinions were expressed only with regard to a 1OPC being (i.e., having animal status), such that even R. Ashkenazi, et al., would agree that a 2OPC being would indeed have human moral status.

6, s.v. *af shekatan*), Hashukei Hemed (San. 65b). For a more detailed review of the Golem and its moral implications, see Ch. 6 “Let Us Make Man In Our Image.”

⁵ Noteworthy here is that while the Maharsha did not write “*capacity* for speech,” but simply “speech,” it is clear from many other commentaries that such was his intent, as it is undisputed that a mute is considered to have a human soul and full human status (see, e.g., Mishna Berura, Biur Halacha 329, s.v. *ela*; Sheilat Yavetz 2:82; Sidrei Taharot, Ohalot 5).

⁶ The claim was echoed by, e.g., R. Natan Gestetner (Lehorot Natan 7:11), R. Tzvi Hirsch Shapira (Darkei Teshuva 7:11), R. Daniel Tirani (Ikarei HaDat, OH 3:15). Nevertheless, R. Asher Weiss (2003: Vaera 9:5, fn. 7) notes that the claim is entirely novel and without precedent, and R. Bleich (1998: 62-3) explains why it is, in fact, invalid.

⁷ The idea of “miraculously” created beings having no moral status is found in the Gem. (San. 59b), and while it is adopted by some, Meiri (*ad loc.*, s.v. *barbel*) explains that it does not apply to things made by “natural” means. And while R. Rosen held creation by Sefer Yetzirah to be “miraculous,” R. Shem Tov ibn Shaprut (Pardes Rimmonim 13a) explains that Sefer Yetzirah is to be considered “natural science.” Be that as it may, R. Asher Weiss (2022: 13) rejects the whole category, explaining (on synthetic meat) that we apply the ontological approach to determine the status of an entity in question, “its method of creation being entirely irrelevant.”

7. Appendix: Other Ethical Issues Inherent in AI

The explicitly stated objective of this thesis has been to discuss “the moral status of artificial intelligence,” specifically, entities that engage us human-like – be they chatbots (e.g., ChatGPT, Claude, Gemini), voicebots (e.g., Siri, Alexa), digital persons (e.g., D-ID, Synthesia) or humanoid robots (e.g., Atlas, Optimus, MenTeeBot, Ameca, Sophia) – whether mindless or mindful. Having done so, it is nonetheless worth mentioning that AI, in all its many forms and applications, raises numerous moral issues¹ that, while beyond the scope of this thesis, would benefit from a Jewish ethical approach. Accordingly, I list here the more prominent moral issues inherent in AI, without attempting to address them (as this remains a project beyond the scope of this work, but one which I have begun).

Autonomous Weapons Systems – AI can be used to enable machines to make the decision to kill people, whether as part of a military campaign on the battlefield or as part of police efforts in the civil arena. Many argue that even if such systems perform better than human beings, reducing death and injury overall, there remains the question as to the ethical propriety of allowing a machine to take the life of a human (see Navon 2023).

Autonomous Vehicles – AI can be used to take complete control of a car (a.k.a. level 5 autonomy). While it is argued that such systems will radically reduce traffic deaths in comparison to human drivers, there are still ethical questions that must be addressed.

One of the most celebrated questions is that of the Trolley Problem that asks: in the event of an unavoidable accident, should the system stay the course, remaining *passive* yet resulting in more casualties; or should the system change course, *actively* choosing to sacrifice those not initially in harm’s way yet resulting in fewer casualties (see, e.g., Navon 2024b)? There are, however, more pressing questions, such as: what is the appropriate balance between safety and utility – i.e., too safe and the car goes nowhere, too assertive and the car moves more efficiently yet endangers people in its path (Shalev-Shwartz et al. 2020).

¹ For a review of the many dilemmas raised by AI, see, e.g., Anderson and Anderson 2011; Lin, Abney, and Bekey 2012; Muller 2013; Sandler 2014; Lin, Abney, and Bekey 2017; Coeckelbergh 2020a; Dubber, Pasquale, and Das 2020; Gocke and Rosenthal-Von 2020; Liao 2020a.

Transhumanism – AI can enable human beings to surpass the limitations of their natural abilities. Transhumanism proposes to integrate AI – e.g., in the form of brain chips – to enhance not only memory, but cognitive capacities and even moral behavior. While a society of super smart, super ethical people may sound super utopian, ethical dilemmas abound. Socially, such technology will likely be divisive between haves and have-nots (see, e.g., Garcia and Sandler 2014; Ishiguro 2021). Ethically, moral enhancement has generated significant debate (see, e.g., Savulescu and Persson 2012; Hauskeller 2013; Douglas 2014; Weissenbacher 2018; Gross 2020). From a Jewish perspective, the Talmud teaches that, “all is in the hands of heaven except for the fear of heaven” (Ber. 33b). Understanding “fear of heaven” as “ethical behavior,”² this means that human beings exercise their freewill solely in their ethical decisions.³ If so, implanting moral-behavior-regulators would make human beings essentially automatons. This would clearly be antithetical to the Maimonidean striving for perfection discussed in this thesis (esp., Chs. 4, 5). And there are many more dilemmas inherent in this technology (see, e.g., Tirosh-Samuelsan 2012, Sandler 2014: esp. Part IV).

Posthumanism – AI is being developed to allow human beings to discard their corporeal bodies and upload precise models of their neural networks, a.k.a. patternism (e.g., Moravec 1988, Kurzweil 2006). If it works, it will allow an individual to live for eternity – or until the power goes out. Myriad are the ethical questions raised by technological posthumanism.⁴ Is this really good for human beings, for humanity? From a Jewish perspective, is this what “*olam haba*” (the world to come) is supposed to look like? In the belief that there is another spiritual realm, should one anchor oneself to this world indefinitely? And then there is the more radical posthumanist approach which argues that, since humans are fatally flawed, “if we could implement in machines the better angels of our nature, then morally we have a duty to, and then we should exit, stage left” (Dietrich 2011: 536).

² Indeed, this is how the term is used in the Bible (see, e.g., Leibtag 2003).

³ As argued Kant (Rohlf 2023: 5.4) and Leibovitch (1992: 21).

⁴ Note that technological posthumanism is distinct from philosophical/cultural posthumanism (see, e.g., Tirosh-Samuelsan 2012).

Social Robot Applications – While we have discussed mindless social robots in the context of how we should relate to them (Ch. 2: The Virtuous Servant Owner) and have noted some of the issues they raise (Ch. 3: To Make a Mind, Introduction), here are some dilemmas occasioned by specific applications:

Elder Care Robots - Robots are being designed for elder care. Is it moral to have our elderly cared for by entities that are only capable of maintaining unidirectional relationships (e.g., Bertolini and Arian 2020)? From a Jewish perspective, we could ask if the commandments like “*vehadarta pnei zaken*” (respect your elders) and “*kibud av v'em*” (honor your father and mother) can be fulfilled via robot care. On the other hand, with the explosion of the elderly population around the world, there are simply not enough people to give one-on-one care (e.g., Liao 2020b). What is the ethical balance?

Nannybots – Robots are being designed to care for our children. What kind of “humanity” will they learn from robots (e.g., Turkle 2011a)? And there are questions of manipulation, as children are vulnerable to the influence of those close to them (e.g., Bertolini 2018). From a Jewish perspective, the commandment of “*pru u'r'vu*” (be fruitful and multiply) is understood to include not simply the physical engendering of offspring, but entails raising them with love and values (e.g., R. S. R. Hirsch, Gen. 1:28). Can a robot do that? Should a robot do that? On the other hand, there is a great need to offload the duties of childrearing, certainly at times. Again, what is the ethical balance?

Lovebots/Sexbots – Perhaps one of the greatest drivers of the robotics industry is the development of robots for relationships. But, as I wrote above, “true love cannot be unidirectional, impersonal, programmed” (Ch. 3: To Make a Mind, Introduction). On the other hand, some argue that there are people who simply cannot find love or maintain relationships (e.g., Richardson 2015: 16) – is there room to allow for such “use cases”? Socially, allowing for robot relationships could tear the very fabric of society as we know it (e.g., Turkle 2011a, Gunkel 2018: 129). From a Jewish perspective, “it is not good for man to be alone” – does the companionship of a mindless humanoid fulfill the biblical ethic to find a partner (*ezer kenegdo*)? Is there something more to this ethic than simple utility?

Social media – AI is being deployed in social media in ways too numerous to count and raising as many ethical questions. But social media is merely the more popular form of what is referred to as Pervasive Information and Communication Technologies (PICT), which encompasses a vast array of devices and systems designed to “collect, transfer, store, analyze, and use massive amounts of personal information” (Pimple 2013: 210). PICT raises issues of privacy, bias, discrimination, and manipulation. However, while these values are undeniably important, they are not so sacrosanct that other values don’t override them. Arguments for overriding privacy are made in the name of state security (e.g., Nissenbaum 2004: fn. 114), public health (e.g., Lee 2014), and monitoring the elderly (e.g., Jones 2014). Regarding bias and discrimination, work has been done on training AI models to ensure they are clean of these social scourges, yet such efforts run the risk of falsifying history itself (e.g., Gilbert 2024). Regarding manipulations, there exists a tension between manipulating for “good,” a.k.a. “nudging” (Thaler and Sunstein 2021) versus manipulating for “bad.”

Deepfakes – AI is being used to generate convincingly real images, audio and video representations of people. The two primary modes of this technology are: (1) avatar mode – AI is used to animate a character, say, Elvis Presley, to do or say whatever the creator desires; (2) face-swap mode – AI is used to take the face (and voice) of a character, say, Gal Gadot, and put her on the body of an unrelated image or full-length movie. Deepfakes, and their underlying image-processing algorithms (e.g., GAN, VAE, DDPM), are used for many positive purposes (e.g., art, cinema, entertainment, therapy), but are unfortunately used for very malevolent objectives (e.g., exploitation, manipulation, crime). Can we regulate the technology to be used only for good? If we cannot, should it be banned?

Virtual Reality (VR) – AI is being used to create compete virtual worlds in which, with the help of sophisticated goggles and gloves, an individual is immersed in a wholly interactive 3D environment. There are numerous positive applications for this technology – e.g., training in sports, business and the military, psychological therapies from PTSD to domestic violence, and the list goes on. Alas, for every positive use there is a corresponding negative abuse. So just as there are issues online with privacy, surveillance, manipulation, etc., so too do these issues arise in VR. Furthermore, because

VR provides a very realistic environment of oneself and others who join, all the crimes that can happen in the physical world can take place in the virtual world. Of course, if someone is beaten or abused – “bodily” – they do not sustain physical injuries, but the (same) psychological damage is done. What is real in a virtual world? What is ethical in a virtual world?

Jobs – It is estimated that within the next one-hundred years AI will take over every human job in the world (Grace et al. 2018). This may not come as a surprise to those familiar with the midrash which proclaims that in the end of days, ‘our work will be done by others’ (see Torah Temimah, Ex. 31:15 n. 34) – though it may not have been understood that the “others” would be AI. Be that as it may, the fundamental question, asked almost one hundred years ago by John Maynard Keynes ([1930] 2009), who predicted the end for human labor by 2030, is: What will we do with our free time? This is an existential question with social and ethical implications. For, how will we sustain, what Yuval Noah Harari calls, the useless class? And more importantly, how can we help people from being “useless”?

While not exhaustive, this list does provide a good picture of the primary applications and their prominent ethical dilemmas. Other issues worthy of mention are those related to Intellectual Property (e.g., Abbott and Shubov 2023), Augmented Reality (e.g., Brinkman 2014), Algorithmic Pricing (e.g., Seele et al. 2019), Personal Eugenics (e.g., Grossl 2020), Robo-Judges (e.g., Wang and Tian 2022), Robo-Doctors (e.g., Nogueroles et al. 2019), as well as the enabling of crime, slavery, and terrorism. In the Jewish world concerns include Robo-Rabbis, Shabbat observance, as well as the enabling of heresy and promiscuity (see, e.g., Navon 2024a: 3.7). And we would be remiss without highlighting the existential threat known as the “Alignment Problem” and related “Control Problem” (mentioned in Ch. 6: “Let Us Make Man in Our Image”, sec. Should We Create Them).

Looking at the many applications the technology affords, it can be seen that AI has the potential to be employed to habituate virtue or habituate vice. Take for example VR, one can enter a virtual world and be guided to be a better human being. For example, one system was developed to educate against discrimination, giving racist users a black avatar. A similar system was used to educate against domestic violence by giving the wife-beater user a female avatar. On the other hand, VR systems can be used to perpetuate abuse

and violence by enabling it or by simply allowing it. In a similar way, social robots can be programmed to reinforce virtue (through positive feedback) or to allow vice (by accepting it).

In short, while AI holds great promise it also poses great peril, and so, like all technologies that preceded it, AI demands our close attention to ensure it is designed, developed and deployed within ethical boundaries (Navon 2024a: 4).

Hebrew Abstract

תקציר התזה

בעשור האחרון, הידוע בחוגים טכנולוגיים בתור "הקיץ השלישי של הבינה המלאכותית (AI)", חוונו עלייה מרשימה של הבינה המלאכותית, הן ביכולותיה והן בהתפשטותה ברחבי העולם. בינה מלאכותית היום מפעילה טכנולוגיות שבעבר הלא רחוק נחשבו כמדע בדיוני. משואבי אבק למכונות אוטונומיות, מטלפונים חכמים לפצצות חכמות, עוזרים וירטואליים ומציאות וירטואלית, ועוד ועוד. אבל, למרות שרשימת החידושים המבוססים בינה מלאכותית ארוכה מאוד, בעיניי החידוש המעניין ביותר והגורם לתסבוכת אתית הגדולה ביותר הינו הרובוט החברתי. כי בעוד שרובוטים חברתיים נועדו לשחרר את האנושות מהנטלים שלה, בדומה לכל טכנולוגיה מאז המצאת המחרשה, הם מביאים איתם דילמות מוסריות שונות בתכלית מכל טכנולוגיה אחרת מאז המצאת המחרשה. הסיבה לכך היא שרובוט שנראה כמונו, מתנהג כמונו ומדבר כמונו, מאתגר אותנו גם בדרכים שלא נתקלנו בהן קודם לכן. בפעם הראשונה בתולדות האנושות אנו נדרשים להתמודד עם ישות שיש לה את היכולות של "דעה ודיבור", יכולות אשר היו, עד כה, נחלתם הבלעדית של אותו נזר הבריאה: האדם.

עם זאת, לא כל הרובוטים נוצרו שווים, ובדומה, גם לא הדילמות שהם מולידים. מצד אחד, יש לנו את הרובוטים של היום, אשר מופעלים בעזרת מה שנקרא "בינה מלאכותית חלשה". זו טכנולוגיה שמשמשת ברשתות נירונים מלאכותיות (ANN) כדי לנתח, להחליט ולבצע משימות בהסתמך על מתמטיקה בלבד. העיבוד הקוגניטיבי של רובוטים כאלה הוא פונקציונלי בלבד, מחוסר ההבנה הסובייקטיבית הנלווית לתודעה ברמה האנושית. רובוטים אלה הם מערכות מחשוב מתוחכמות ביותר בצורה דמויית אדם ולכן אני מכנה אותם "מכונות חסרי-נפש" (mindless machines). מצד שני, יש לנו את הרובוטים של המחר, שיבוססו על "בינה מלאכותית חזקה". זו טכנולוגיה המבקשת ליצור ולהפעיל תודעה ברמה אנושית כדי לנתח, להחליט ולבצע את משימותיה. ניתן לומר שרובוטים מהסוג הזה, הם בעלי יכולות קוגניטיביות דומות לאלו של בני אדם, כולל תודעה ברמה אנושית, ומשום כך, ניתן להגדירם כבעלי "נפש" (mind). הם, אם כן, רובוטים "מלאי-נפש" (mindful machines) - דמויי אדם במובן העמוק, אבל לא מהמין האנושי. כל אחד מסוגי הרובוטים האנושיים הללו – חסרי-נפש ומלאי-נפש - מציב דילמות מוסריות ייחודיות, ובהתאם, דורש גישה מוסרית ייחודית לו.

ראוי להזכיר שניתן ליישם את שני סוגי הבינה המלאכותית, תיאורטית, במספר עצום של צורות פיזיות (למשל, שעונים, טלפונים, מחשבים ניידים, ושרתים) אך, מה שמרתק במיוחד בתזה זו הוא התגלמותם בדימוי אדם – קריא: "רובוט". שכן, כאשר מערכת מחשב נראית, מתנהגת ומדברת כמו בני אדם, אנו, בני האדם, נעשים מבולבלים, מאוהבים ובסופו של דבר מעורבים בהם כאילו הם "אנושיים". עם זאת, כל התגלמות שמעסיקה אותנו כבני אדם, בין אם היא פשוטה כמו ChatGPT או Siri או Alexa או מתוחכמת כמו החברה הווירטואלית "Caryn.AI", היא רלוונטית לניתוח כאן.

מעניין שדווקא הגלגולים האחרונים הללו של בינה מלאכותית (למשל ChatGPT) העניקו דחיפות ניכרת לעבודה זו. בעוד שההתמקדות הראשונית שלי הייתה ברובוטים חברתיים (SRs), הריבוי של LLMs (Large Language Models) כמו ChatGPT קדמה לעליית ה-SRs. כתוצאה מכך, היישום המעשי של הגישות האתיות המוצעות שלי ל-SRs שבטוח יבואו למימוש בעתיד הקרוב, ניתנות ליישום על LLMs שכבר נמצאים כאן.

הדילמות

עכשיו, כפי שזכר, הדילמות האתיות שמעוררים רובוטים חסרי-הנפש (mindless robots) שונות באופן מהותי מהדילמות שמעוררים רובוטים מלאי-הנפש (mindful robots). בעוד שבמקרה של רובוט חסר-הנפש מעסיק אותנו כמה שנדמה כאנושי אך הוא אינו כזה, הרובוט מלא הנפש מעסיק אותנו, במובן מסוים, כמכונה אך כאמור איננו מכונה של ממש. בהתאם לכך, הדילמה משתנה לפי סוג הרובוט אליו נתייחס. ובכן, כאשר מדובר ברובוטים חסרי-הנפש, מוקד הדילמה הוא האדם והיחס שלנו כלפי המכונה. לעומת זאת, כאשר מדובר ברובוטים מלאי-הנפש מוקד השאלה הוא הרובוט עצמו והדילמות האתיות ביחס למכונה כאשר יש לה נפש אדם ממש.

ליתר דיוק, בעוד שכל טכנולוגיה מעוררת דילמות רבות, יש דילמה אחת בולטת שמציג הרובוט חסר הנפש, מה שאני מכנה "דיאלקטיקת אותנטיות-סגולה" (Virtue Authenticity Dialectic - VAD). כלומר, בהתחשב בכך שרובוט חברתי נראה, מדבר ומתנהג כמו בן אדם אמיתי, אנו באופן טבעי מגיבים ומתקשרים איתו כמו בן אדם. אם כן, ראוי להתייחס אליו באותו אופן סגולה שאנו צריכים לשאוף אליו כשמתקשרים עם בני אדם, שמא נבוא להתייחס אל רעינו בני אנוש כאל מכונות. עם זאת, על ידי עיסוק ברובוט כאילו הוא ישות חיה ומודעת, אנו מאבדים את ההערכה שלנו ליחסים אנושיים אותנטיים. אנו נלכדים במצב תפיסתי של יחסים בו הרובוט נתפס כאילו הוא ישות חיה ומודעת. מה שמביא אותנו לתפוס רובוט דמוי-האנוש כתחליף מתאים ואפילו עדיף לבני אדם, אשר להם קשרי גומלין מורכבים הרבה יותר. מערכות היחסים האדיבות שלנו עם מכונות עלולות לגרום לפטירת היחסים האותנטיים שלנו עם בני אדם.

בניגוד לרובוט חסר הנפש שרק נראה מודע (אבל הוא לא), הרובוט מלא הנפש נראה כמכונה במובן שנוצר כדי לשרת צורך מסוים, אבל הוא כן מודע. ישות כזו מולידה שתי דילמות עיקריות שבמרכזן עצם יצירתה: (1) האם מבחינה אתית מותר ליצור ישות מודעת שתכליתה היחידה ב"חיים" היא לשרת אותנו? הלא זו עבדות? (2) אם אנו דוחים את יצירתן של מכונות מלאי הנפש שיעשו את רצוננו, מה לגבי יצירתן להיות ילדינו, חברינו, שותפינו לחיים? האם יש משהו לא ראוי מבחינה אתית ביצירת ישויות מודעות סינתטיות?

על אף שלא מעט הוגים מודרניים התמודדו עם דילמות אלו תוך שימוש בגישות מוסריות שונות, טענתי היא שהפילוסופיה היהודית מעניקה לנו פרדיגמות עתיקות אך עמוקות שמהן ניתן לנסח תגובות מוסריות משמעותיות לדילמות האולטרה-מודרניות הללו. פרדיגמות אלו, המעוגנות כפי שהן במסורת היהודית, מספקות נקודות מבט פילוסופיות, אתיות ומשפטיות ייחודיות. ראשית, הפילוסופיה היהודית מסתכלת על העולם כתכליתי - כבעל התחלה, סוף ותכלית שיש להשיגה - וזה מיידע משמעותה של "חיים טובים". יתר על כן, חיים טובים תכליתיים כרוכים ומוגדרים על ידי הערכים האתיים של מסורת בת 3300 שנה שנשמרה בחיים על ידי עם, שלפי נשיא ארה"ב ג'ון אדמס, "עשה יותר לתרבת בני אדם מכל אומה אחרת." לבסוף, ערכים אתיים אלה אינם רק הצעות להתנהגות טובה, אלא מקודדים כנורמות משפטיות (כלומר, הלכתיות), ובכך מוסיפים את מרכיב החובה להתנהגות האתית האמורה. לפיכך, פרדיגמות אתיות יהודיות טעונות טלוס (telos) ומחוזקות על ידי חובה. ובעוד הפרדיגמות הללו, והגישות האתיות שהן יוצרות, מבוססות היטב בפילוסופיה היהודית, הן חורגות מגבולות דתיים ומוצגות כאן באופן שכל איש ואיש בכל עם ועם יכול להעריך וליישם.

הפרדיגמה הראשונה שיש להפעיל היא מה שאני מכנה "בעל משרת הסגולה" (Virtuous Servant Owner), הנגזרת מפרשנות שלי להלכה האחרונה ב"הלכות עבדים" של הרמב"ם. במבט ראשון, נראה שהטקסט הוא אוסף אקראי של ציטוטים ואנקדוטות שהובאו כדי לעודד מערכת יחסים רחמנית בבעל-עבד כלפי עבדו. אולם קריאה מעמיקה יותר מגלה טקסט בנוי כהוכחה לוגית – טקסט מכוונת ומכרעת שגם מגדירה וגם דורשת לא פחות משלמות מוסרית מלאה מבעל-עבד (וקל וחומר מאנשים "רגילים"). בתזה זו, אני איישם את הפרדיגמה הזו כגישה אתית לדילמת ה-VAD שנוצרה על ידי הרובוט חסר הנפש (כמו גם ישויות AI חסרות נפש אחרות, למשל, LLMs ואנשים וירטואליים). בעוד שיש פילוסופים שבחרו "לפתור" את הדילמה על ידי יחסי סגולה (מידות טובות) במחיר של אותנטיות, ופילוסופים אחרים בחרו "לפתור" את הדילמה על ידי יחסים אותנטיים במחיר של יחסי סגולה, אני טוען ש"בעל משרת הסגולה" יכול לשמור הן על מעלתו הסגולית והן על האותנטיות שלו.

הפרדיגמה השנייה אותה איישם היא מה שאפשר לכנות פרדיגמת "תהילת האדם של ירמיהו" אשר, כפי שהסביר הרמב"ם, מספרת על הטוב הנעלה (summum bonum) של האדם. הנביא ירמיהו מכריז: "כֹּה אָמַר ה' : ... בְּזֹאת יִתְהַלֵּל הַמִּתְהַלֵּל, הַשֶּׁפֶל וְיִדְעַ אֹתִי--כִּי אֲנִי ה', עֹשֶׂה חֶסֶד מְשֻׁפָּט וְצַדִּיקָה בְּאֶרֶץ ..." מכאן הרמב"ם מסביר כי תהילת האדם, ה-summum bonum שאליו כולם צריכים לשאוף, מורכבת מלדעת את ה' (כלומר, התפתחות שכלית) ולהדמות לו (כלומר, התפתחות מוסרית). לפיכך, לאדם חייב להיות החופש להתפתח, לשאוף להגשים את ה-summum bonum שלו. ובעוד שיש להודות, כפי שעושה הרמב"ם, שהטלוס הזה הוחזק על ידי "הפילוסופים הקדומים" (למשל, אריסטו), הוא נוסח הרבה לפניו על ידי הנביא ירמיהו. אני איישם את הפרדיגמה הזו, של "תהילת האדם של ירמיהו", כדי לבחון את הנכונות המוסרית שבתכנות רובוט מלא-הנפש על מנת שהוא יפטור אותנו מהנטלים שלנו. בצורה הזאת, אני מבהיר שתכנות כזה

יהפוך את חיי הישות הזו לחסרי משמעות ובכך יהווה הפרה מוסרית בוטה ביותר. ולמרות שהדילמה נדונה באמצעות גישות מוסריות חילוניות, פרדיגמה זו מציעה תובנות שאפילו בעלי השקפה חילונית יכולים להעריך.

לבסוף, הפרדיגמה השלישית אותה איישם, היא זו של "הגולם", יצור דמוי אנושי סינתטי שהועלה מעפר האדמה. הגולם מופיע בספרות היהודית לאורך ההיסטוריה בצורות שונות: (א) ישות הפועלת על ידי "חיות" - כלומר, כוח שמאפשר ניידות אך לא חוויה פנומנלית; (ב) ישות הפועלת על ידי נפש חיה – כלומר, היא בעלת חישה; ו(ג) ישות בעלת תודעה ברמה אנושית. הגולם בצורה האחרונה, בעל התודעה, הוא זה שמעורר עניין רב ביותר בהתייחסות לנכונות מוסרית של יצירת רובוט עם תודעה. עם זאת, ולמרות שכן אתייחס אליו בדיון, בעיני הגולם השני הוא המעניין יותר, שהוא בעצם הראשון בספרות היהודית, המצוי בתלמוד (סנהדרין סה:): ומכונה "גברא". גברא זה, על אף שלא הייתה לו תודעה ברמה אנושית, נוצר מתוך כוונה שתהיה לו תודעה ברמה אנושית. באמצעות הניתוח המקיף של נרטיב הגברא, אני מראה שהתלמוד מתנגד באופן מוחלט ליצירת דמוי אדם בעל תודעה (בין אם ברמת נפש אדם ובין אם ברמת נפש חיה). נוסף על כך, וכפי שיפורט ויובהר בתזה זו, לעמדת התלמוד חשיבות אתית ומשפטית כאחד, כך שאסור מבחינה הלכתית ליצור רובוטים עם תודעה.

תוכן העניינים

i	תקציר.....
1	1. הקדמה.....
8	2. תבנית התזה.....
13	3. התזה.....
172	4. סיכום.....
177	5. ביבליוגרפיה.....
215	6. נספח: מעמד המוסרי של נשמה.....
217	7. נספח: בעיות אתיות אחרות בבינה מלאכותית.....
א	תקציר בעבית.....

עבודה זו נעשתה בהדרכתו של

פרופ' חנוך בן פזי

מן המחלקה של פילוסופיה יהודית

של אוניברסיטת בר-אילן.

**המעמד המוסרי של בינה מלאכותית:
מבט מהאתיקה היהודית**

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"

מאת

מואיז נבון

המחלקה לפילוסופיה יהודית

הוגש לסנט של אוניברסיטת בר-אילן