Will Robots Too Be in the Image of God?
Dorobantu, M.

published in
Theology and Westworld
2020

document version
Version created as part of publication process; publisher's layout; not normally made publicly available

document license
Article 25fa Dutch Copyright Act

Link to publication in VU Research Portal

citation for published version (APA)

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 22. Dec. 2023
The HBO series *Westworld* smartly plays with explicit and implicit timeless questions regarding human nature and human distinctiveness: What does it mean to be human? Is there anything special about us? Could other entities, namely Artificial Intelligence (AI), ever be considered equal, or even superior, to us?

In this chapter we explore the challenges posed by the emergence of truly intelligent robots—as some of the Hosts in *Westworld* seem to be—to Christian theological anthropology in its core claim that humans are special because they are created in the image of God (*imago Dei*). To do this, we first assess the current understanding of human uniqueness in science, philosophy, and theology, with a focus on the various theological interpretations of *imago Dei*. We then look at the notion of strong AI and sketch its implications for theological anthropology. Further, the focus moves on *Westworld* in an attempt to understand what are the precise characteristics of the type of AI depicted in the television series. Do the Hosts qualify as strong AI? Is their consciousness a relevant topic for theological debate? A careful dig underneath the *Westworld* construction of consciousness will reveal a surprising theological twist. Equipped with these analyses, we will finally approach some of the most fundamental and fascinating theological questions raised by *Westworld*: can we still speak of humans being in the image of God if AI becomes conscious?
Chapter 5

Would robots too be in the image of God? Could the Hosts too be religious?

HUMAN DISTINCTIVENESS IN SCIENCE AND PHILOSOPHY

Humanity has come a long way since the humanism of the early modern age. Back then people thought that humans were the center of the universe and the measure of all things. Nowadays scholars are moving away from these ideas at an accelerated speed. Especially throughout the last century, a couple of important things have become clear: that the world is a much bigger place than people imagined; that our origins lie in the same tree of life as those of all the other life forms planet Earth; that neither language, creativity, tool-use nor any other cognitive capacity represents a unique feature of humankind; and that human reasoning, emotions, and so forth, are not the mystical and inscrutable phenomena we thought they were. Rather, they are algorithmic in nature and can be accounted for by the same physical laws that govern the rest of the universe.

These well-established ideas, alongside others, seem to point to a final demise of the idea of human specialness. However, there are also counterarguments to the demise of the uniqueness of humanity. One of them is that the human brain, though a mere physical object, still remains the most complex structure in the known universe (Fischbach 1992, 48). Secondly, each human intellectual ability might indeed be shared with one or more non-human species, but there are still colossal differences in degree and convergence of these abilities, which likely account for the cultural and technological achievements of humanity. Finally, there is the so-called hard problem of consciousness: why is there an I inside each one of us who feels, rejoices, suffers, and experiences continuity throughout time?

All these for-and-against arguments might leave one with mixed feelings regarding whether or not humans are truly special, which reflects pretty accurately the current state of this debate in philosophy and science.

HUMAN DISTINCTIVENESS AND IMAGO DEI IN CHRISTIAN THEOLOGY

Christian theology, on the other hand, has always articulated its account of human distinctiveness using its own specifically theological arguments, rooted in the biblical tradition of humans being created in the image of God. The first chapter of the book of Genesis reads:
Then God said, “Let us make humankind in our image, according to our likeness; and let them have dominion over the fish of the sea, and over the birds of the air, and over the cattle, and over all the wild animals of the earth, and over every creeping thing that creeps upon the earth.” So God created humankind in his image, in the image of God he created them; male and female he created them (NRSV, Gen 1:26-27).

Due to its position in the very first chapter of the Bible, and to the radical claim made about human origins, the concept of *imago Dei* has always sat at the core of theological anthropology. It is the intuition that humans, while being part and parcel of the created physical universe, still share something with the divine that the rest of creation does not. In other words, there must be some way in which humans are like God and unlike the animals.

It was never clear, though, exactly what it means that humans are in the image of God. The traditional answer within Christian thought is known as the substantive interpretation, which locates *imago Dei* in a certain capability or set of capacities of the human intellect, a view that betrays the heavy influence of Greek philosophy (van Huyssteen 2008, 126). Which capacity, more precisely? Well, for most of human history the difference between humans and any other animal was so strikingly obvious, that there was no need for theology to define what exactly it meant by it. One could freely pick their choice from the multitude of suitable candidates: reason (Irenaeus), rationality (Gregory of Nyssa, influenced by Aristotle’s “rational animal”), or the trinity of memory, understanding and will (Augustine).\(^1\)

Although the substantive interpretation has traditionally been the dominant one in Christian thought, it has become increasingly problematic in the past century or so. One reason is that there is no such thing as an intellectual capacity that is universally shared by all humans. Regardless of the chosen quality, there will always be groups of humans (e.g., the mentally disabled) that lack it, thus being left outside of the definition of being human. This fact makes the substantive interpretation problematic for evident ethical reasons. Secondly, the idea that humans are in possession of a unique intellectual capacity that no other animal shares has become increasingly less tenable, as scientific progress has severely challenged this understanding of human uniqueness. Christian theologians have therefore, starting with the twentieth century, looked for alternative proposals of how *imago Dei* should be interpreted. The current leading interpretations are the functional and the relational one.

The functional interpretation, a favorite among Christian biblical scholars, speaks of *imago Dei* as the election of humans by God to act as God’s representatives in the world (Herzfeld 2002, 23). This interpretation is more plausible from an exegetical perspective, by conforming to the common usage of the notion of image throughout the Ancient Near
Chapter 5

East. As biblical exegete Gerhard von Rad—the initial proponent of this interpretation—points out,

> Just as powerful earthly kings, to indicate their claim to dominion, erect an image of themselves in the provinces of their empire where they do not personally appear, so man is placed upon earth in God’s image, as God’s sovereign emblem. He is really only God’s representative, summoned to maintain and enforce God’s claim to dominion over the earth (1961, 58).

The strength of this interpretation is that it takes into account the immediate literary context of the biblical text, where imago Dei sits in the same verse as human dominion over the world, so there could be a great likelihood that the two are connected. Secondly, it is also likely that the functional interpretation is the closest to what the Priestly writer (sixth century BCE) originally had in mind, as it is looking at the text in its wider historical and literary context. Its main weaknesses are its reliance on extra-biblical material, and that it largely neglects the wider narratives of the Hebrew Bible, let alone the New Testament identification of Jesus (2 Cor 4:4, Col 1:15, Heb 1:3) with the true image of God (Cortez 2010, 22-23).

The other main modern candidate for an interpretation of imago Dei is the relational interpretation, most famously developed by Christian theologian Karl Barth (1958). Instead of identifying the image of God with some intellectual capacity of the human mind, or with our election to rule over the world, Barth chooses to see the divine spark in our fundamentally relational character, just as God the Trinity is a relational being. The I-Thou structure of being that exists within the Trinity, among its three divine persons, is the blueprint for the new I-Thou relationship between God and God’s creatures, the humans, and between humans themselves. The departure points of this interpretation is the plural in God’s exhortation from Genesis 1:26, “Let us make humankind in our image,” and the immediate juxtaposition in 1:27 between imago and the sexual differentiation of humans: “in the image of God he created them; male and female he created them” (NRSV). The complementary and confrontational relationship between man and woman thus becomes the prototype for human relationships in general, as well as what it means to be in the image of God.

The relational interpretation, apart from being arguably the most beautiful, has several strengths. One of them is that it is in great harmony with the wider background of Christian theology. While the other interpretations may be over-focused on the text in Genesis, the relational interpretation of imago Dei incorporates the unique insight of Christian theology that God is, in fact, a trinity of persons. Also, it makes much more sense from a Christological perspective, where relationships could be seen as the main focus of Christ’s life and teaching. This interpretation
is better synchronized with contemporary philosophy than are the other

two candidates. While the substantive interpretation is problematic in a
post-Cartesian age, the functional interpretation might raise concerns of
anthropocentrism and exploitation of nature. The relational interpreta-
tion, however, is very much in tune with contemporary philosophy,
which has also taken a relational turn from Kant and Hegel onwards,
reversing the traditional causal order between substance and relation
(Shults 2003, 32). Late twentieth-century constructionist philosophy, for
example, asserts that it is not individuals, but relationships, that are the
basic units of society (Gergen 1971, 156). Instead of relationships being
the product of individuals interacting with each other, constructionist
philosophy posits that it is the other way around, that it is individuals
and minds that are the byproducts of relationships. The relational inter-
pretation argues in more or less the same terms when it roots the image
of God and the core of human personhood in the relationship with God,
for which human beings have been called into existence.

In terms of weaknesses, the relational interpretation is criticized for
not being exegetically solid enough. Firstly, because Barth chooses to
neglect the widespread agreement among exegetes that the plural "let
us" refers to a heavenly court, and not to a multiplicity within godhead
(von Rad 1961, 57). Secondly, it seems anachronistic to attribute such a
complex understanding of relationality to the ancient writers of the Gene-
sis text (Barr 1993, 161).

HUMAN DISTINCTIVENESS AND STRONG AI

As it stands, it looks like Christian theology has a clearer position, though
in its own terms, than science on human specialness. Biologists are still
puzzled by the paradox of humans being so similar and, at the same time,
so different from other animals. Neuroscientists are able to study and
map out the nervous system and its processes with unprecedented preci-
sion, while still making virtually no progress towards a convincing
understanding of what makes us conscious. Theologians, on the other
hand, seem to have an easier job in providing an account of human
uniqueness that is coherent in its own realm of theological arguments,
and which is still plausible from a non-theological point of view. What
this means is that more work needs to be done in linking the current
interpretations of imago Dei with contemporary science and philosophy.
But, as it stands, there seems to be no philosophical or scientific insight
that would in principle invalidate these developments in theological
anthropology. Against this background comes the new challenge of Arti-
ficial Intelligence and, more precisely, the possibility of strong AI (or
human-level AI, or Artificial General Intelligence), machines that would
match or outperform humans in any cognitive task.
Chapter 5

The project of AI began in the 1950s as a new line of research into human intelligence. John McCarthy, one of the co-founders of the field of AI and the one who actually came up with this name, explicitly stated in his proposal for the Dartmouth workshop that his purpose was “to study the relation of language to intelligence” (McCarthy 1955, 10). AI has since come a long way, and its best-known current applications might largely be commercial. It is nevertheless noteworthy to remember that its initial drive was rather philosophical, namely to explore whether a human mind could in principle be replicated on an artificial support, and if not, which are the human cognitive capacities that would prove impossible to capture.

Some human abilities, such as theorem proving or game playing, have proved relatively easy to replicate, providing promising early results. Others, like vision, natural language processing or kinesthetic coordination are still challenging today. Furthermore, skills like one-shot learning, namely the ability to learn from one or just a handful of examples, seem to be nowhere near computers’ reach.

The initial enthusiasm of the symbolic AI of the 1950s and 1960s has been tempered by periods of disappointment, also known as the AI winters (Dorobantu 2019, 5-6), and so did the optimism about the possibility of fully simulating a human mind. Nonetheless, with the increased computation power correctly predicted by Moore’s law, which enabled the implementation of deep learning neural networks, the hype around AI is up again. More significantly, the topic of strong AI is again on the public radar.

On the one hand, AI experts disagree on when strong AI could become a reality, but largely agree that it is in principle possible, and that it could happen within a few decades. A 2014 survey among them reveals a median 50 percent probability of strong AI occurring before 2050 (Muller and Bostrom 2014, 555). On the other hand, the topic of human-level robots has, as a result, made a strong comeback in pop-culture, as exemplified by movies like *Her* (2013) and *Ex Machina* (2014), or television series such as *Westworld* (2016-2018).

Strong AI does not necessarily entail consciousness, and *Westworld* illustrates it very well: the Hosts had passed the Turing Test—broadly accepted as the standard for strong AI—after the first year of training, but it would take a much longer journey for them to become conscious. But even without the consciousness part, the development of strong AI would pose severe challenges to the notion of human specialness in both science and theological anthropology. If machines would be capable of doing everything better than humans, including abstract reasoning, artistic creativity, or social interactions, would there be any place left for the concept of human specialness?

In Christian theology, strong AI would deliver the final blow to any substantive interpretation of *imago Dei*. It is obvious that if robots could
outperform humans at everything across the board, then the case for any uniquely human capacity would become untenable. But, more importantly, even the more subtle relational and functional interpretations would be heavily put to the test.

The relational interpretation locates the image of God in humans’ capacity and vocation for relationship, both with God and with one another. But since the nature of the Turing Test is essentially relational, this would mean that any artificial entity passing it would display enhanced relational abilities, including a theory of mind and an acute perception and modeling of other agents’ feelings, beliefs and psychological states. The question remains whether the mere ability to engage in relationships is enough to account for imago Dei, or if there is anything more that is required by the relational interpretation. The answer provided by Karl Barth’s anthropology to this question points to the conclusion that relationality alone is not enough, and his relational interpretation hints at personhood and agency as necessary substrates of meaningful relationships. According to Barth, human beings are in the image of God because they are a “Thou” to which God can address, and who can respond back freely. This definition raises the imago bar to a higher level, one that unconscious strong AI could not access.

The functional interpretation identifies imago Dei with humans’ status as elected by God to be stewards of the created world. Since strong AI would be able to model the real world and act in it at least as well as humans, this theological understanding of human uniqueness would seem to lose its ground in such a scenario. Moreover, the ways in which humans have fulfilled their role of stewards so far—by inflicting suffering upon billions of animals through industrial farming, or by driving the ecosystem close to collapse through human-caused climate change—set a relatively low bar for intelligent machines, even unconscious ones, to do a better job.

THE AI OF WESTWORLD

We have so far examined the implications that the emergence of strong AI, not necessarily endowed with consciousness, would have on the notion of human distinctiveness, understood in theology as the image of God in humans. However, the plot of Westworld takes the discussion one step further and proposes a scenario where strong AIs are on a path of acquiring consciousness. This completely changes the rules of the theological argumentation game.

It is, of course, understood that the portrayal of AI in Westworld is not necessarily how strong AI will develop in our “real” world. But the creators of Westworld invite us to perform this thought experiment, and we believe that even playing with such hypothetical sci-fi scenarios can help
Chapter 5

Theological reflection clarify its positions on historically difficult issues like human nature or the divine image in humans.

The androids of Westworld, built in seclusion by Arnold and Ford, are said to have passed the Turing Test after the first year. We do not know if this implies the original version of the test, which is done through a chat box, or a face-to-face test, where judges would interact with real humans and human-like androids and could not tell them apart. This last point could prove relevant, and it is more likely that the plot involves the chat box version of the test, otherwise why would the Hosts need further improvement?

Passing the Turing Test does not presuppose consciousness. However, the implicit point, which would be worth further exploration, is that robots would actually need to be conscious in order to pass a face-to-face Turing test. At least within the confines of the park, human guests (and TV watchers, we might add) seem to be able to very quickly tell the difference between Hosts and other human guests. The notable exception is the Host Angela, who has both William and Logan, on separate occasions, not able to instantly tell if she is a Host or not, in both cases posing as a Westworld employee.

The question thus becomes: What is it that makes Hosts so easily recognizable, if they had already passed a Turing Test long ago, and had multiple updates since, presumably improving their social skills even further? One possible reason is that they are intentionally kept like that, just human enough to superficially seem real, but not human enough to creep guests out. But this plausible commercial reason does not answer the question of what it is that they lack. The most likely candidate is, of course consciousness. Their lack of consciousness and of agency makes them zombie-like at times, stuck in a narrow narrative loop, and displaying empty looks that betray their lack of understanding. A strong proof of this is William’s perception of Dolores as being “not like the others,” precisely at the moment when she starts displaying signs of consciousness. It is therefore no surprise that the same “reveries” that render the Hosts more human-like in their minor gestures are also credited for triggering their awakening process.

The two masterminds, Arnold and Ford, are both engaged in an explicit effort to bring the Hosts to consciousness, although Ford’s intentions are only revealed at the end of the first season. Arnold is trying to achieve it by using the theory of the bicameral mind, with a pyramid (or, as later revealed, a sort of inwards spiral) of four layers: memory, improvisation, self-interest, and another undisclosed element. Ford, on the other hand, believes that the experience of suffering is key in the emergence of consciousness. The two methods seem to work in conjunction, leading to the awakening of first Dolores and Maeve, and then of other Hosts.
For the purpose of our theological analysis, we will consider the evolved versions of the Hosts as exemplary for the AIs of *Westworld*. The main feature that distinguishes them from the strong AI discussed in the previous section is, of course, their consciousness. This changes the parameters of the debate on the theological interpretation of *imago Dei*. Before revisiting that debate and evaluating what would be left of it if conscious AI were to emerge, let us first have a look at what being conscious actually means in the realm of *Westworld*.

CONSCIOUSNESS AND THE BORDER BETWEEN HUMAN AND NON-HUMAN

The topic of consciousness is addressed within the wider framework of questioning human nature and human distinctiveness, which is beautifully constructed throughout the plot and dialogues of the TV series. In other words, what is the difference, if any, between guests and Hosts, or between biological humans and human-like strong AI? This fundamental question is played with in a variety of ways and is looked at through the lens of several pairs of dichotomous categories: biological/artificial, tainted/pure, real/fake, or old/new.

The narrative begins with a bold challenging of the real/fake paradigm, made manifest in the ubiquitously quoted dialogue between William and Angela: “Are you real?” “Well, if you can’t tell, does it matter?” (2016, s1e2). The boundary between the two is subsequently intentionally blurred: as Dolores puts it, everyone was once new to this world, and “the newcomers are just looking for the same things we are—a place to be free, to stake out our dream, a place with unlimited possibilities” (2016, s1e1).

The biological/artificial distinction is also not particularly helpful. In spite of Logan cutting Dolores open to show William that she is a robot like the rest of the Hosts, this does not change William’s mind about her distinctive nature. What does manage to change his mind, and ultimately push him to his dark side, is the realization that Dolores does not remember him, in spite of everything they have been through together.

Another attempt to define the human/android difference pops up in the dialogue between Maeve and Felix, when she challenges him to explain how come he is so sure about the distinctiveness of his own human nature, to which he replies: “Because I was born, you were made” (2016, s1e6). The language used is eminently theological, an obvious reference to the Nicene distinction between begotten (γεννηθεντα) and made (πουρθεντα). However, this ontological differentiation does not seem convincing, at least to Maeve. By taking Felix’s hand and concluding that they feel the same, she pushes him to admit that “we are the same these days, for the most part,” with the notable difference that the balance in
computing power is heavily shifted in favor of the non-human. It is at this point that Felix is able to articulate what he sees as the real difference, the fact that the Hosts are under human control, or at least they still were at that moment in the plot timeline.

This represents a crucial hint that once the Hosts gain control of their own actions and storylines, the difference between human and non-human would not be relevant anymore. The idea is reinforced later in the story, where Dolores asks Arnold the key question of what is real, to which he answers: “that which is irreplaceable” (2018, s2e1). In other words, the message that is pushed across the screen is that ontology is nowhere near as important as function and ability to make an impact. This insight is also interesting from the perspective of the theological debate on imago Dei, where the difference between ontology and function is the core distinction between rival interpretative proposals.

All the above arguments converge in pointing to consciousness as the crux in distinguishing between what is human and what is not. It is consciousness that enables one to remember one’s experiences and maintain functional relationships through time (e.g. what Dolores was initially incapable of in her relationship with William), to take control of one’s destiny and to become irreplaceable, therefore real.

Consciousness, or the lack of, is the decisive factor in establishing whether an AI is really in imago hominis, in the image of humans, at least in the Westworld universe. A more attentive analysis of how the emergence of consciousness is presented in the Westworld plot will reveal, as shown below, some surprising theological assumptions by the writers.

**AI CONSCIOUSNESS AND THE AUGUSTINIAN IMAGO DEI**

What are the marks of consciousness, as proposed by the Westworld narrative? One of them must be memory. As shown above, memory is listed as a necessary building block in the road to consciousness, as the Hosts should be able to remember experiences in order to learn from them and construct a sense of personal identity through time. As a confirmation, both Dolores and Teddy emphasize the link between memory and consciousness. First, they both independently hear a voice in their head saying “Remember!” (Teddy in s1e10, Dolores in s1e9). Second, the connection between consciousness and memory is articulated even more undoubtedly by Teddy. Right before shooting himself, in his climatic moment of consciousness, he says: “I remember now. I remember everything” (2018, s2e9). Furthermore, in The Bicameral Mind (2016, s1e10), Dolores comes to the realization that “This is what Arnold wants. He wants me to remember.”

A second compound and marker of consciousness, as it emerges from the plot, is arguably the ability to understand. This does not refer to
understanding natural language, which all Hosts are capable of, since they passed the Turing Test and are able to engage in interactions with the guests. What is meant by understanding is rather the ability to comprehend complex abstract notions, alongside a deeper and more holistic understanding of reality and of one’s place within that reality, including one’s ability to challenge it. The question “Have you ever questioned the nature of your reality?” is used throughout the series as a litmus test for consciousness. The vital link between understanding and consciousness is made obvious when Dolores is unable to fully comprehend Arnold while her journey towards consciousness is not yet complete (“I’m trying, but I don’t understand”). It is thus only logical that when her awakening is finally achieved, right before shooting Ford, she tells Teddy: “I understand now” (2016, s1e10).

The third element of consciousness, built on top of the previous two, or as Arnold and Ford would put it, underneath them, is undoubtedly having a will, an agency, the power to freely choose and create one’s destiny. Maeve very plastically expresses: “Time to write my own fucking story” (2016, s1e8). The will is also partly what is probably meant by self-interest, the third layer of Arnold’s pyramid/maze of consciousness. The crucial importance of the will in rendering one conscious is exemplified by the difference between Dolores’ acts of killing Arnold and Ford. In reference to shooting Arnold, her freedom of choice is severely questioned by Ford, who says that it was most likely Arnold’s suicide through her hand. On the contrary, when shooting Ford she seems to make a free choice out of her own will, which is the final proof of her full awakening to consciousness. Having a will is also used as a proof of consciousness in the cases of Teddy and Maeve. Teddy understands how Dolores has altered him and freely chooses to kill himself rather than continue like that. Maeve’s defining choice is even more dramatic, as it exhibits the kind of irrationality that we think of as typically human: she seems to freely choose to go back to the park to find her daughter, defying the escape story that had been written for her.

We have so far argued that consciousness—and thus, as previously demonstrated, humanness—consists of three main ingredients: memory, understanding, and will. Here is where the punch line comes: these are the exact three features of the human mind that Augustine identified sixteen centuries ago as being impossible to separate from one another, and the *imago Dei* in humans: “These three, memory, understanding, and will, are, therefore, not three lives but one life, not three minds but one mind” (Augustine 2003, 58).

Is the usage of Augustine’s trinity of intellectual features a coincidence? One cannot know for sure, but the clues point to the conclusion that it is not. Firstly, the series is imbued with theological references, especially around the story of creation and the concept of *imago*. Ford intentionally uses theological language when he says “Arnold made you
Chapter 5

in our image and cursed you to make the same mistakes” (2016, s1e9), and theological imagery when he speaks of Michelangelo’s supposed secret message in the painting Creation of Adam. Secondly, some character names are evocative, most likely intentional, of theological themes. Dolores undoubtedly recalls suffering and the image of Christus dolor (Christ of the sorrows), Angela could point to the army of angels, while Hector Escaton is a clear reference to the notion of eschaton, the post-historic age. Perhaps the strongest proof of Augustinian inspiration lies in Felix’s name: felix culpa, or blessed fault, is a concept from theodicy and theological anthropology that is connected with the Fall of humans from paradise, and it is traced back unsurprisingly to the writings of Augustine.

If the employment of the Augustinian model of imago Dei, imago hominis is intentional, as it appears to be, the underlying message is clear: just as humans are said to be created in the image of God, so AI too will transitively be created in the image of humans. Would this imply the belief that AI would be in the image of God too? This is rather difficult to decide, since God is largely absent from the Westworld universe: in Ford’s words, “God has nothing to do with it” (2016, s1e9 in response to Bernard’s exclamation “Oh, God!”).

We can therefore contend that the usage of the Augustinian model is a mere beautiful metaphor used by the writers, even though an outdated one, as theological anthropology has since moved away from the substantive interpretation. However, it is still a very interesting thought experiment to assess how the emergence of Westworld-type of strong AI would impact the theological discussion about imago Dei, which is what will be done in the remaining of this chapter.

CONSCIOUS AI AND THE IMAGE OF GOD

One of the big questions to be answered is whether or not theological anthropology could still speak of humans as created in the image of God, if the type of AI depicted in Westworld were to emerge. In other words, is any of the current theological interpretations of imago Dei capable of allowing the existence of sentient AI without completely losing its legitimacy?

From the three dominant interpretations, substantive, functional and relational, the substantive was from the start deemed unsuitable, even though it seems to be preferred by the series writers. Even in a pre-AI world, the substantive interpretation is already vulnerable to criticism due to a number of legitimate scientific and ethical concerns. Should strong AI emerge, let alone AI endowed with consciousness, this interpretation would either implode or conceal that the AI too is in the image of God. The characteristic that would, in this case, distinguish humans and AI from the animals would be this advanced type of consciousness.
A valid point could nevertheless be made that the absence of proof of advanced consciousness in other animals does not constitute a proof of absence. In other words, we would probably not know for sure that only humans and AI are capable of advanced consciousness.

The functional interpretation looks to be the safest from the scenario of emerging conscious AI, because it keeps all its metaphorical eggs in a basket outside the reach of non-theological critique. The idea that God has elected humans is not scientifically falsifiable. It is therefore difficult to engage in a meaningful dialogue with it, because one either believes in this election or doesn’t. There are, however, a few comments to be made.

If humans are indeed divine representatives in the created universe, this role comes with both privileges and responsibilities. In order to fulfill this role, humans must first of all make sure they do not destroy neither the world, nor themselves. In Westworld, Ford alludes several times at the AI kind being the next step in evolution, an idea that is also popular in some transhumanist circles. This scenario can unveil in a variety of ways that span from the utopian to the dystopian. In the best of them (at least for humans), AI and even Artificial Superintelligence become a splendid tool and an existential companion for humans, helping them on their quest of fulfillment. This would imply that AI would assist humans in their mission of caring for the world, assigned to them through their election as imago Dei.

Nonetheless, the scenario of violence, confrontation and competition hinted at in the second season of Westworld seems to be heading in the opposite direction, confirming Logan’s warnings that all this might lead to the demise of the human race. Should AI wipe out and replace humanity, either intentionally or by accident, it would also mean the end of, among others, the functional interpretation of imago Dei.

The last one standing, the relational interpretation, would also be severely challenged if Westworld-type of conscious AI emerges in the future. These AIs would not only outperform humans in every intellectual domain one can imagine, but they would also be fully conscious free agents, at least to the same degree as we can say that about ourselves. In Westworld, the Hosts who are awake engage in free relationships with each other. This is particularly visible in Maeve’s storyline, who stresses her choice not to force the other Hosts of her crew to join her. Her relationship with Hector is therefore fully free and consented, and there is no reason why it should be considered inferior to human relationships, as long as it involves conscious agents.

This last point raises severe doubts over the capacity of the relational interpretation of imago Dei to retain the notion of human distinctiveness in such a case. That is to say it could mean that to be in the image of God is to be relational, but there is no reason to exclude relationally capable AI from imago. Moreover, one could also see striking similarities between the genesis of humans and that of AI. Humans are brought to existence in
Chapter 5

and by their relationship with God, their creator, and this is what enables them to also be in relationship with each other. Conscious AIs are awoken through their relationship (of abuse, or friendship, or both) with humans, their creators, and this is what enables them to also develop meaningful relationships between themselves.

Should the development of AI follow the route envisaged by the Westworld writers, which is the hypothetical assumption of this exercise of imagination, robots would eventually become veritable Thou’s, ticking all the boxes required by the relational interpretation. Would this mean that humans are not created in the image of God? Not necessarily, but it would certainly disjoint human uniqueness from imago Dei, because the AIs too would be in the image of God, in such a case.

COULD THE HOSTS BE RELIGIOUS?

One remaining question is whether the Hosts would also be able or choose to engage in a direct relationship with God. This is a question that is as fascinating as it is shrouded in mystery. One should be careful in making any such predictions, and especially in restricting God’s capacity and willingness to reveal Godself to any creature, be it biological or synthetic. That being said, the question can at least be attempted within the confines of the Westworld universe.

At first glance, it would not seem that the Hosts manifest any genuine religious interest or behavior, apart from the accidental reference to the divine in verbal clichés like “oh my God!” They do not pray and do not wonder in a Pascalian way whether there is any creator behind the order of their universe. Their only gods are the Delos employees who control their storylines and their destinies.

However, there is an intriguing aspect to the placement of the chapel at the very center of the maze. As with most of the other religious allusions, it is very likely that this setup is not accidental. Moreover, Dolores’ awakening happens in a rather ritualistic fashion. She enters into the confessionary and symbolically descends with an elevator to the underground laboratories, likely a metaphor for a descent into the depths of her own mind.

Besides that, her entire journey through the maze of consciousness had so far been guided by an internal voice, perceived as alien to her normal flow of thoughts: “Sometimes I feel the world out there is calling me” (2016, s1e3); “Sometimes I feel like something is calling me, telling me there’s a place for me, somewhere beyond all this” (2016, s1e4); “There’s a voice inside me telling me what to do” (2016, s1e5).

That voice is subsequently revealed to be Arnold’s, and finally her own, in Arnold’s attempt to guide her toward consciousness, as predicted by the “bicameral mind” theory. However, in light of the argu-
Will Robots Too Be in the Image of God?

ments above, a case can be made that this voice too has a dual function. It is indeed Arnold’s code, but it is also a metaphor for existential restlessness, and this can be inferred from the usage of the specific vocational language, with words like “calling” and “beyond.” Dolores’ path to awakening and freedom would in this case also be interpreted as a religious journey, without losing its primary and explicit meaning.

The most interesting part of this speculation is that there is, in theological anthropology, a fourth interpretation of imago Dei, besides the three discussed so far in the paper, namely the eschatological interpretation. One of its most distinguished representatives is Christian theologian Wolfhart Pannenberg, who relates the image of God to excentricity, an internal inclination of humans to locate the purpose of their existence outside themselves, toward a not-yet-reached destiny (1985, 43-79). Imago Dei acts therefore like a gravitational pull, a source of direction for humans both at the individual, and at the collective level, driving them toward fulfillment in the relationship with the eternal God in the eschaton (van Huysssteen, 2006, 141).

Pannenberg’s excentricity looks strikingly similar to what the conscious Hosts of Westworld are exhibiting: they are all driven by motivations outside themselves. This is most explicit in Dolores’ case, but it can be observed in each of the other Hosts’ gravitational pull towards fulfilling their drive, be it protecting other Hosts (Peter Abernathy, Teddy, Maeve), triggering mayhem (Hector), or the collective attraction to an unknown place, the Valley Beyond. We do acknowledge that all these drives are pre-programmed, even the one towards the Valley Beyond, but this does not close the door to seeing the conspicuously excentric nature of the conscious Hosts. Imago Dei, even in an eschatological interpretation, should in this case also be extended to our conscious artificial creations.

Is this in any way intentional from the part of the writers? Here we are inclined to answer negatively. It would certainly seem too much of a stretch to intentionally depict the Hosts as bearers of imago Dei in the eschatological sense. If it is involuntary, however, this might serve as anecdotic proof that, at an intuitive level, we see (or at least the writers of Westworld do) excentricity as the most fundamental feature of humanity. If anything, it is a confirmation of Pannenberg’s interpretation of imago Dei.

We set about interrogating the Westworld scenario what would it take for intelligent robots to be considered “real” by humans, in other words equal to them. The answer revolved around the topic of consciousness, manifested through the display of memory, understanding, and will. This is in remarkable coincidence with the structural interpretation of imago Dei, as formulated by Augustine in the fifth century CE, which hints to an intentional parallel from the part of the authors between the
Chapter 5

biblical story of humans created in the image of God and the creation of AI in the image of humans.

Could Christian anthropology still speak of humans as *imago Dei* in such a scenario? While the notion of *imago Dei* is eminently theological, and thus considerably outside the reach of non-theological critique, none of its current interpretations seems capable of maintaining its claims if *Westworld*-type of AI emerges. *Imago Dei* would have to be separated from any claim of human distinctiveness, and it would likely have to include the AIs, since they would tick all the boxes required by current standards.

Finally, given that the Hosts seem to share the “uniquely human” exocentricity, there’s one last question that arises: Could strong AI also be religious? In spite of God’s absence from *Westworld*, we have seen how certain clues do at least leave the space open to interpreting them as religious elements. We cannot know if AI could ever become religious. But if it evolves to exhibit the same kind of exocentricity and openness to the world as humans do, we must also allow for the possibility of divine revelation to robots.

NOTES

1. See Irenaeus (1885, 892), Gregory of Nyssa (1892, 729-30), and Augustine (2003, 10.11.18, 58).

2. In the premiere episode (2016, s1e1), the “reveries” are described as a routine update introduced by Ford, which would allow the Hosts to act more human-like through the usage of small gestures and behaviors based on past experience. Later in the series (2016, s1e10), the idea is revealed to have actually belonged to Arnold, as a building block of his pyramid of consciousness, allowing the Hosts to retain parts of their memories even when being repeatedly reset.

BIBLIOGRAPHY


Will Robots Too Be in the Image of God?


Editors and Contributors

[**J01.0**] ABOUT THE EDITORS

[**J01.1**] Dr. **Juli Gittinger** received her PhD from McGill University in Montreal with emphasis on contemporary issues in Hinduism. She has Masters degrees from the University of Colorado at Boulder and from SOAS in London, both in the fields of Indian religions. Her areas of personal research interest include Hindu nationalism, religion in media, and religion/pop culture. Her second book was published in 2019, *Personhood in Science Fiction: Religious and Philosophical Considerations* (Palgrave-Macmillan).

[**J01.2**] Dr. **Shayna Sheinfeld** (M.T.S., Harvard Divinity School; Ph.D., McGill University) is Honorary Research Scholar at the Sheffield Institute of Interdisciplinary Biblical Studies (SIIBS), University of Sheffield. She has published extensively on Judaism including the early Jesus movement in the first and second centuries CE. Her current projects include a monograph on *Leadership in Ancient Judaism* and a textbook on *Women in Ancient Judaism and Christianity* (co-edited, Routledge). Dr. Sheinfeld also works extensively with biblical afterlives in popular culture. In addition to this volume, she is currently co-editing a collection on *Good Omens and the Bible* (Sheffield Phoenix).

[**J01.3**] ABOUT THE CONTRIBUTORS

[**J01.4**] Dr. **Olivia Belton** is a postdoctoral research associate at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge. She is currently researching media representations of and public perceptions of autonomous flight. She completed her doctorate on posthuman women in science fiction television at the University of East Anglia in April 2019. She has a book chapter on media representations of sex robots, co-authored with Dr. Kate Devlin, forthcoming in *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (Oxford University Press, March 2020).

[**J01.5**] **Jacob Boss** is a doctoral candidate in religious studies at Indiana University. He is writing his dissertation on grassroots transhumanism. Jacob
Editors and Contributors

Dr. **Tony Degouviea** is a graduate of the University of East Anglia where he teaches film and media. Following on from the ideas of his thesis, subsequent work and publications have focused on the socio-political influence of religion in dystopian science fiction film and television. Here, he finds particular fascination in the way that biblical allegory and religious cabal are now adopted by Hollywood on a frequent basis, where fictional visions of apocalypse, incorporating ideas of biblical “myth” and prophecy, are often framed within the machinations of science fiction.

Marius Dorobantu is a researcher in theology and science at the University of Strasbourg, France. He is currently completing his PhD degree, with a thesis reflecting on the challenges of strong Artificial Intelligence for Christian theological anthropology. Previously, he obtained a BA in orthodox theology from the University of Bucharest, Romania, and an MA in theology from Radboud University, Nijmegen, The Netherlands.

Dr. **Amanda Furiasse** received her PhD in religion and graduate certificate in museum studies from Florida State University and currently teaches courses at Hamline University. Her research is aimed at understanding how religious communities can use art and ritual practice to redress violence and trauma with a specific focus on dance, music, and other embodied aesthetic practices. She is currently working on a digital museum and archive preserving material evidence of women’s ritual practices in historically underrepresented communities.

Dr. **David K. Goodin** earned a PhD in religious studies from McGill University in the philosophy of religion, with a secondary area of concentration in Patristic theology. Currently, he is a lecturer for the McGill School of religious studies in Montreal, Canada, Professeur Associé at the Université Laval, Institut de Théologie Orthodoxe de Montréal, and an instructor for the Pappas Patristic Institute at the Holy Cross Greek Or-
Editors and Contributors

thodox School of Theology in Brookline, Massachusetts. Originally from Miami, Florida, he now resides and teaches in Montreal, Quebec, Canada.

[J01.10] Rev. Dr. Kristin Johnston Largen is professor of systematic theology at United Lutheran seminary, and associate dean of religious and spiritual life at Gettysburg College. She is the editor of Dialog: A Journal of Theology, and her most recent book is Women’s Bodies, Shin Buddhism and Rebirth, forthcoming from Lexington Books.

[J01.11] Dr. Kevin J. Wetmore, Jr. is the author of numerous books and articles about theology and pop culture, including The Empire Triumphant: Race, Religion, and Rebellion in the Star Wars Films and The Theology of Battlestar Galactica. He has also written extensively on Catholic theater and Jesuit theater and drama. He is a professor at Loyola Marymount University.

[J01.12] Dr. Jaime Wright completed her PhD at the University of Edinburgh’s School of Divinity. Her research involves the intersection of science, religion, and literature (especially science fiction). Her publications include “Emily Dickinson: A Poet at the Limits,” an exploration of the epistemological intersection of science and religion within the poet’s work, published in Theology in Scotland in 2017, and “In the Beginning: The Role of Myth in Relating Religion, Brain Science, and Mental Well-Being,” published in Zygon in 2018. Jaime is also training for ordained ministry in the Scottish Episcopal Church.