Zygon®

Journal of
RELIGION & SCIENCE

SEPTEMBER 2024
VOLUME 59 NUMBER 3

Zygon®

Journal of

RELIGION AND SCIENCE

VOL. 59, NO. 3, September 2024

Editor: ARTHUR C. PETERSEN, University College London, UK

Former Editors: RALPH WENDELL BURHOE (1966–1979), KARL E. PETERS (1979–2009), PHILIP HEFNER (1989–2009), WILLEM B. DREES (2009–2018)

Book Symposium Editor: MLADEN TURK, Elmhurst University, USA

Editorial Assistant: SARAH JOST (University College London)

Assistant Treasurer and Chicago Editorial Office Assistant: David M. Glover (Zygon Journal Corporation)

Joint Publication Board, appointed by IRAS (www.iras.org), the Institute on Religion in an Age of Science (1954), CASIRAS (www.casiras.org), the Center for Advanced Study in Religion and Science (1972) and ISSR (www.issr.org.uk), the International Society for Science and Religion (2002):

SUSAN BARRETO, Journalist, Urbana, IL, USA
WILLEM B. DREES, Prof. Emeritus of Philosophy of the Humanities, Tilburg University, The Netherlands
NIELS HENRIK GREGERSEN, Prof. of Systematic Theology, University of Copenhagen, Denmark
SOLOMON H. KATZ, Prof. Emeritus of Anthropology, University of Pennsylvania, USA
SARAH LANE RITCHIE, Program Officer, Philosophy & Theology, John Templeton Foundation, USA
EDWIN C. LAURENSON, Lawyer, McDermott Will & Emery LLP, San Francisco, USA (President)
IGNACIO SILVA, Associate Professor, Philosophy Institute, Universidad Austral, Buenos Aires, Argentina (Secretary)
MLADEN TURK, Baltzer Distinguished Chair of Theology and Religion, Department of Religious Studies, Elmhurst University, USA
GAYLE E. WOLOSCHAK, Prof. of Radiology and Cell and Molecular Biology, Northwestern University, USA (Treasurer)

Editorial Advisory Board:

THOMAS AECHTNER, Assoc. Prof. in Religion and Science, Univ. of Queensland, Australia
CHRISTINA AUS DER AU, Prof., Pädagogische Hochschule Thurgau, Bishop, Protestant Church of Thurgau, Switzerland
ZAINAL ABIDIN BAGIR, Center for Religious and Cross-Cultural Studies, Gadjah Mada Univ., Indonesia
JOSEPH BULBULIA, Prof. of Psychology, Victoria Univ., New Zealand
NA CHEN, Research Professor, Center for Social Development, Fudan University, China
ERNST M. CONRADIE, Senior Prof. of Religion and Theology, Univ. of the Western Cape, South Africa
RANA DAJANI, Prof. of Cell Biology, Hashemite University, Jordan
HELEN DE CRUZ, Danforth Chair in the Humanities, Prof. of Philosophy, Saint Louis University, MO, USA
TERRENCE DEACON, Prof. of Anthropology, Univ. of California Berkeley, USA
CELIA DEANE-DRUMMOND, Director, Laudato Si' Research Institute, Oxford, UK
ELAINE HOWARD ECKLUND, Herbert S. Autrey Chair in Social Sciences, Prof. of Sociology, Rice Univ., Houston, TX, USA
FERN ELSDON-BAKER, Prof. of Science, Knowledge and Belief in Society, Univ. of Birmingham, UK
DIRK EVERS, Prof. of Systematic Theology/Dogmatics, Martin Luther Univ. Halle-Wittenberg, Germany
OWEN J. FLANAGAN, Prof. of Neuroscience, Duke Univ., Durham, NC, USA
MOHAMMED GHALY, Prof. of Islamic Law and Biomedical Ethics, Center for Islamic Legislation and Ethics, Doha, Qatar
MARCELO GLEISER, Prof. of Physics and Astronomy, Dartmouth College, Hanover, NH, USA
CARLOS MIGUEL GÓMEZ RINCÓN, Assoc. Prof., Theology and Religious Studies, Univ. del Rosario, Bogotá, Colombia
URSULA GOODENOUGH, Prof. Emerita of Biology, Washington Univ., St. Louis, Durham, NC, MO, USA
NIDHAL GUESSOUM, Prof. of Physics and Astronomy, American Univ. of Sharjah, United Arab Emirates
YUANLIN GUO, Prof. of Philosophy, Center for Science, Technology and Society, Tianjin Univ., China
MARK HARRIS, Andreas Idreos Prof. Science and Religion, Univ. of Oxford, UK
PETER HARRISON, Prof. Emeritus, Institute for Advanced Studies in the Humanities, Univ. of Queensland, Australia
ANTJE JACKELÉN, Archbishop Emerita of the Lutheran Church of Sweden, Uppsala, Sweden
SEUNG CHUL KIM, Director, Nanzan Inst. For Religion and Culture, Nanzan Univ., Japan
JUAN LARRAÍN, Prof. of Cellular and Molecular Biology, Pontifical Catholic Univ. of Chile, Chile
YU-TING LEE, Assoc. Prof., East Asian History and Culture, National Taiwan Univ. Taiwan
SARA LUMBRERAS, Prof., Institute for Research in Technology, School of Engineering, Comillas Pontifical Univ., Spain
AGNALDO PORTUGAL, Assoc. Prof., Department of Philosophy, Univ. of Brasilia, Brazil
FABIEN REVOL, Prof. of Moral Theology of Ecology, Catholic Univ. of Lyon, France
ROBERT J. RUSSELL, Prof. Emeritus of Theology and Science, Graduate Theological Union, Berkeley, CA, USA
KNUT-WILLY SÆTHER, Prof. of Religious Studies, Volda Univ. College, Norway
CHRISTOPHER SOUTHGATE, Prof. of Christian Theodicy, Department of Theology and Religion, Univ. of Exeter, UK
ANN TAVES, Distinguished Prof. Emerita of Religious Studies, Univ. of California Santa Barbara, USA
HAVA TIROSH-SAMUELSON, Irving and Miriam Lowe Prof. of Modern Judaism, Arizona State Univ., Tempe, AZ, USA
CLAUDIA E. VANNEY, Prof. of Physics and Philosophy, Universidad Austral, Argentina
DANIE VELDSMAN, Prof. of Theology, Univ. of Pretoria, South Africa
WESLEY J. WILDMAN, Professor of Philosophy, Theology, and Ethics, Boston Univ., MA, USA

Statement of Perspective. The word *zygon* means the yoking of two entities or processes that must work together. It is related to *zygote*—meaning the union of genetic heritage from sperm and egg, a union that is vital in higher species for the continuation of advancement of life. The journal *Zygon* provides a forum for exploring ways to unite what in modern times has been disconnected—values from knowledge, goodness from truth, religion from science. Traditional religions, which have transmitted wisdom about what is of essential value and ultimate meaning as a guide for human living, were expressed in terms of the best understandings of their times about human nature, society, and the world. Religious expression in our time, however, has not drawn similarly on modern science, which has superseded the ancient forms of understanding. As a result religions have lost credibility in the modern mind. Nevertheless some recent scientific studies of human evolution and development have indicated how long-standing religions have evolved well-winnowed wisdom, still essential for the best life. *Zygon*'s hypothesis is that when long-evolved religious wisdom is yoked with significant recent scientific discoveries about the world and human nature, there results credible expression of basic meaning, values, and moral convictions that provides valid and effective guidance for enhancing human life. *Zygon* also publishes manuscripts that are critical of this perspective, as long as such papers contribute to a constructive reflection on scientific knowledge, human values, and existential meaning.

Organization and Governance. *Zygon: Journal of Religion and Science* is owned and managed by Zygon: Journal of Religion and Science NFP, a not-for-profit scholarly corporation established in Chicago in 2019, which succeeds a joint venture established in 1965 and has as its members: IRAS, the Institute on Religion in an Age of Science (1954), CASIRAS, the Center for Advanced Study in Religion and Science (1972), and ISSR, the International Society for Science and Religion (2002). Over the history of the journal, funding has come from a variety of sources: an endowment has been built up from contributions from individuals, while operating costs of the corporation had been met from a share in publication revenue. With the disappearance of publication revenue after shifting to diamond open access in 2024, the main sources of income are now contributions (if any) from member organizations and individuals, as well as investment income and unrealized gains on investments. The corporation is dedicated entirely to publishing the journal. The corporation's Board of Directors ('Joint Publication Board') has nine members; each member organization appoints three directors. The selection and appointment of the Editor-in-Chief, or the joint Editors-in-Chief, is done by the Joint Publication Board. The appointment term is five years (renewable). The Editor-in-Chief can appoint Editors, Assistant Editors and Editorial Assistants. The Joint Publication Board appoints members to the Editorial Advisory Board; the appointment term is five years (renewable) and appointments are made (or renewed) with a view to diversity in the composition. The journal was formerly published by The University of Chicago Press (1966–1978), Wilfrid Laurier University Press (1979–1989) and Blackwell/Wiley (1990–2023), and from 2024 the journal is published by the Open Library of Humanities (OLH).

Contact Information. Editor-in-Chief: Professor Arthur C. Petersen, arthur.petersen@ucl.ac.uk, Department of Science, Technology, Engineering and Public Policy, University College London, Gower Street, London WC1E 6BT, UK. Book Symposium Editor: Professor Mladen Turk, turkm@elmhurst.edu, Department of Religious Studies, Elmhurst University, USA. Address for Corporate Affairs: 5416 S Cornell Avenue, Fourth Floor, Chicago, IL 60615, USA.

Manuscripts. Please submit online via https://www.zygonjournal.org. For author guidelines, see https://www.zygonjournal.org/site/author-guidelines.

Zygon®
Journal of
RELIGION AND SCIENCE
VOL. 59, NO. 3, September 2024

Editorial

Religious Robots

In this Editorial for the September 2024 issue, you will find brief introductions to the articles included in this issue, both in the general section and in the three thematic sections, which include two Book Symposia.

Articles

This issue contains three general articles. Mark Harris investigates "quantum fundamentalism" and theological liberty; he engages theologically with mystical approaches to quantum mechanics and argues that, for quantum fundamentalism to function as a worldview, it should inform a sense of human purpose. Gordon Miller reviews the classic eighteenth- and nineteenth-century debates on the soul of the frog and brings these up to date; he shows how Alfred North Whitehead's philosophy of organism and some recent physiological research offer promising and constructive postmodern hope for addressing the dispiriting loss of soul. And Mitchell Mallary critically analyzes the debate surrounding methodological naturalism within the context of scientific inquiry; he demonstrates how both science and methodological naturalism are conceived of differently between different parties, indicating that this is primarily a semantic debate, and advocates for methodological flexibility.

Religious Robots: AI as Religious Subject and Object (by Marius Dorobantu)

The proliferation of generative AI in the past two years raises complicated ethical and philosophical questions about the nature of AI systems, their potential, and the challenges and opportunities they bring about. The religious sphere is not immune from such concerns. In fact, it may be the arena where the status and role of AI unfolds in some of the most interesting ways.

In June 2022, months before the public release of ChatGPT and the ensuing generative AI revolution, a Google engineer started claiming that an AI program he was internally testing had become self-aware. His assessment was largely driven by his own religious beliefs and the AI's predilection to discuss existential topics. It was an early demonstration of how the intersection between religion

and AI would come to play an increasingly prominent role in public debates over the nature and ethical use of these technologies. If AI is to become ubiquitous, how will it be integrated into religious practices? If AI chatbots started to generate religious discourse and even claim to have developed religious interests and beliefs, how could we evaluate the authenticity of such outputs? Is it possible, at least theoretically, for artificial systems to develop anything close to what in humans we call religion?

To explore these questions raised by the nascent generative AI technology, theologian Marius Dorobantu convened an interdisciplinary workshop in Amsterdam in December 2022, supported by CLUE+, the interfaculty research institute at the Vrije Universiteit. Some of the articles in this thematic section, guest edited by Dorobantu, emerged out of that workshop, while a few others were subsequently added. The fundamental question explored in all of them, from multiple disciplinary angles, is whether AI could play a significant role in religious life, either as a tool in human religiosity or as an authentic religious subject itself.

Philosopher Pim Haselager highlights how comparing human cognition with AI can significantly inform humanity's journey of self-understanding, somewhat similar to the comparisons with animals and angels in medieval theological anthropology. He contrasts the smartness and cognitive abilities of AI with its complete lack of understanding and sentience, which he regards as *sine qua non* conditions of authentic religiosity. In another article, psychologist Fraser Watts and the late Yorick Wilks, AI pioneer, explore the feasibility and acceptability of AI-powered spiritual companions. Based on empirical research conducted with both GPT chatbots and Wizard-of-Oz methodology (humans masquerading as AIs), the article tentatively concludes that although artificial spiritual companions might proliferate, especially those that facilitate self-exploration, there are still dimensions of human spiritual counselling that might resist automation for the foreseeable future. Theologian Max Tretter intersects robotics, pop-culture and Christian eschatology to ask an intriguing question: is there an afterlife for robots and, if so, what might it be like? If there is hope for all creation to undergo eschatological completion, he argues, then the tentative answer to such a question could only be affirmative, opening up a discussion about the diverse conceptions of afterlife that can be imagined for robots. Computer scientist William Clocksin wrestles head-on with the question of whether intelligent robots could become religious. Following up on his 2023 article in this same journal, he makes a compelling case that future androids would likely develop a form of non-human personhood through sustained engagement in social relationships. If androids start questioning their place in the world and relationships with others, they might use religion in doing so, just like humans do, especially if they come to acknowledge their interdependence with others—human, robotic, or divine. Religious scholar Robert Geraci highlights the critical role religious beliefs and practices might play in recognizing artificial general

intelligence, should it ever emerge. Reviewing the historical intersection between religion and robotics, the article critiques the colonial frameworks at work in evaluating "otherness"—of both humans and robots. It also reflects upon the need for an inclusive approach to integrating intelligent robots into our societies by acknowledging their potential for religious experience and participation. Theologian Daekyung Jung reaches a similar conclusion in his article, arguing from the perspective of embodied cognition. Future AI systems, particularly those integrated with soft robotics and driven by homeostasis as a fundamental goal, might exhibit religious behaviors if they develop human-level intelligence and self-awareness. Such religious behaviors would serve the AIs as cognitive mechanisms helpful in navigating existential challenges and the need to transcend finitude. In contrast with Clocksin, Geraci and Jung, Dorobantu concludes in his own article that authentic religiosity is an unlikely development in robots, despite the theological openness to such scenarios. He argues that religion's deeply embodied, social, and phenomenological underpinnings in humans may not be replicable in AI systems due to their fundamentally different bodies, cognitive architectures and needs.

Book Symposia (by Mladen Turk)

The two Book Symposia in this issue feature Donovan Schaefer's *Wild Experiment: Feeling Science and Secularism after Darwin* (2022) and John Perry and Joanna Leidenhag's *Science-Engaged Theology* (2023).

Wild Experiment has five commentators: Matthew Arthur, Anne Pollock, Josh Reeves, Esha Shah, and Mari van Emmerik. Donovan Schaefer responds to the comments.

Science-Engaged Theology has four commentators: Dirk Evers, King-Ho Leung, Robyn Boeré, and Mikael Leidenhag. John Perry and Joanna Leidenhag respond to the comments.

ARTHUR C. PETERSEN
Department of Science, Technology, Engineering and Public Policy,
University College London, London, UK
arthur.petersen@ucl.ac.uk

Zygon®
Journal of
RELIGION AND SCIENCE
VOL. 59, NO. 3, September 2024

Open Library of Humanities

Contents

ARTICLES

Quantum Fundamentalism and Theological Liberty. . . . . Mark Harris 593

The Soul of the Frog for a Postmodern Age: Updating a Classic Debate . . . . . .Gordon L. Miller 609

Scientific Methods and Sacred Matters: On Methodological Naturalism, Divine Agency, and Christians in Science . . . . . .Mitchell Mallary 638

RELIGIOUS ROBOTS: AI AS RELIGIOUS SUBJECT AND OBJECT

From Angels to Artificial Agents? AI as a Mirror for Human (Im)perfections . . . . . .Pim Haselager 661

Spiritual Conversation with a Companion Machine. . . . . Fraser Watts and Yorick Alexander Wilks 676

What Is the Afterlife Like for Robots? An Experimental Eschatological Sneak Peek. . . . . Max Tretter 692

Religion and the Android . . . . . William F. Clocksin 717

Religion among Robots: An If/When of Future Machine Intelligence . . . . . Robert M. Geraci 729

Are Religious Machines Possible? Embodied Cognition, AI, and Religious Behavior. . . . . .Daekyung Jung 748

Could Robots Become Religious? Theological, Evolutionary, and Cognitive Perspectives . . . . . Marius Dorobantu 768

SCHAEFER'S *WILD EXPERIMENT*

Science and Technology Studies and Affect in the Wild. . . . . Mathew Arthur 788

Cultivating Calm Passions. . . . . Anne Pollock 800

Kuhn, Conspiracy Theories, and Scientific Progress. . . . . Josh A. Reeves 811

Love Affair with an Idea: Commentary on *Wild Experiment: Feeling Science and Secularism after Darwin* by Donovan Schaefer . . . . . Esha Shah 825

Changing Minds: Cogency Theory and Secular Bodies . . . . . Mari van Emmerik 839

"They Made up Quiet Songs": Classification, Feeling, and Technique in Science and Religion . . . . . Donovan O. Schaefer 850

PERRY AND LEIDENHAG'S *SCIENCE-ENGAGED THEOLOGY*

Good Fences Make Engaging Neighbors . . . . . Dirk Evers 871

Theology and Policing . . . . . King-Ho Leung 884

Empirical Realities? A Question of Basic Principle . . . . . Robyn Boeré 893

Considering the Place of Apophaticism within Science-Engaged Theology . . . . . Mikael Leidenhag 901

Replies to Some of Our Friendlier Critics. . . . . John Perry and Joanna Leidenhag 912

Zygon®: Journal of RELIGION & SCIENCE

Harris, Mark. 2024. "Quantum Fundamentalism and Theological Liberty." *Zygon: Journal of Religion and Science* 59 (3): 593–608. DOI: https://doi.org/10.16995/zygon.11499

Open Library of Humanities

Quantum Fundamentalism and Theological Liberty

Mark Harris, Andreas Idreos Chair in Science and Religion, Faculty of Theology and Religion, University of Oxford, UK, Mark.Harris@theology.ox.ac.uk

Quantum mechanics (QM) is astonishingly successful as a theoretical framework, underpinning countless scientific areas and providing the impetus behind entire technologies. Many scientists suspect that physical reality is fundamentally quantum in nature, even if we perceive little of this in our everyday human experience. This is the viewpoint of "quantum fundamentalism." Yet, the conceptual implications of QM defy common sense, to such an extent that popular culture largely perceives of QM as a source of counterintuitive weirdness. At the same time, bestselling self-help manuals portray QM as a source of hidden healing power within, while spiritual readings invoke QM as a bridge to the divine, or as a source of theological analogies. Scientists often denounce these mystical approaches as "quantum quackery," but I examine their serious side. I argue that, for quantum fundamentalism to function as a worldview, it should inform a sense of human purpose, something for which theological analysis is well equipped.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

This article presents the (slightly adapted) text of my inaugural lecture as the Andreas Idreos Professor of Science and Religion at the University of Oxford, delivered in the Andrew Wiles Building at Oxford's Mathematical Institute on May 29, 2024.

Let me begin with a few words about Andreas Idreos, after whom my chair is named. Born in 1917, Dr. Idreos was a medical professional, a senior officer in the World Health Organization. He took a global view of the human future, becoming convinced that deeper understanding between different cultures would be furthered by greater attention to the interaction between two great universals: the sciences and religious belief. One of Dr. Idreos's contributions to furthering this aim was to endow this chair that holds his name. I am humbled to have been appointed to play my own part in working towards his vision of increased unity between the peoples of our world. I want to pay tribute to Dr. Idreos's wider vision before becoming stuck in an academic thicket of my own choosing.

Here is what I plan to say. First, something of my personal perspective on science and religion: Why me for this job? Second and third, what it says on the tin: quantum fundamentalism and theological liberty.

Science and Religion: A Personal Perspective

First, why me? My thoughts return to the inaugural lecture I gave for my previous role as Professor of Natural Science and Theology at New College in the University of Edinburgh (Harris 2022). When I gave that lecture in 2019, I spent my time speaking of the work of my three illustrious predecessors in the college's Chair in Natural Science (John Fleming, John Duns, and James Young Simpson) and how I would take their objectives further. The New College history depicts them on a figure delightfully entitled "Professors of chairs now extinct" (Watt 1946, 224–25). Ironically, they were all biologists working on evolution and its theological implications in the nineteenth and early twentieth centuries. Since their chair then went into abeyance for nearly a century between them and me, it was fairly straightforward to demonstrate that I would be bringing something new.

Not so with my new job. Now, I cannot possibly claim to be resurrecting something that was extinct. The Andreas Idreos chair has been going strong throughout the twenty-first century. And I have three extremely illustrious predecessors, all of whom are still alive and active (John Hedley Brooke, Peter Harrison, and Alister McGrath). All three have made sizeable contributions to the field of science and religion that still stand. If I were to attempt to review their work, allotted space and time would fail me, so let me summarize it in a single sweep by saying that they have all been instrumental in developing what has become the consensus in the field over the last, say, ten to twenty years, namely that the old dualism of science versus religion is unsustainable, even if

that dualism lingers on in wider society. Popular culture may still maintain that science and religion are incompatible, at loggerheads with each other, but the truth of the matter is that both sets of enterprise have evolved alongside each other in subtle and complex ways, and their relationship is similarly nuanced. For sure, there has been conflict between some scientific theories and some religious beliefs down the ages, and there are still some today—we can all think of the debate over young-Earth creationism versus evolutionary science, for instance—but there are many areas where there is little direct engagement, while there are also areas of positive interaction. My three predecessors in the Andreas Idreos Chair have been instrumental in contributing to this rich story of engagement between the sciences and religious beliefs, and I see no sign that we will run out of questions any time soon.

So, what am I going to do? Will I bring anything new? I certainly hope so. For one thing, I am the first physicist to hold the chair, and although I spend much of my time mixing with theologians, philosophers, and historians—and indeed, I have also trained as a theologian myself—it is physics and its hinterland that continue to grab my attention. I must admit that I am not very active in mainstream physics anymore, but thanks to some long-suffering collaborators, I am able to keep my hand in from time to time in the lab and to help with the occasional scientific publication. At any rate, I describe myself on the Oxford Faculty of Theology and Religion website as a physicist working in a theological environment. A little pretentious, I know, but it seems to capture my approach to science and religion, namely a concern with what drives working scientists most deeply as human beings with their own beliefs and convictions about their science and our human world.

We know that science changes the world. What, in fundamental terms (by which I mean human aspirations and convictions about what is most real) is it that drives the world of science? Why does science work? The working sciences cannot answer these questions; they need philosophers, theologians, and historians: the science and religion field. But there is a great deal more that this latter field could do, I suggest, to help working scientists grasp the significance of their daily work, examine their working assumptions and prejudices, and analyze their scientific hunches, some of which, I have discovered, are implicitly theological.

How should science and religion scholars—especially those who have theological interests such as me—go about this? Well, it is not straightforward: there are professional barriers to overcome. I noted earlier that there are areas of positive interaction between the sciences and religions—areas where there is a lot of traffic—but it is fair to say that the traffic is almost entirely one-way, from science to theology.

It was not ever thus. Once upon a time, there was conversation both ways. Some of the most important figures in the history of science—such as Galileo,

Isaac Newton, and Robert Boyle—were not shy of analyzing and justifying their work on theological grounds. But not so today. The traffic is one way.

There are good reasons. Scientists are generally unable to engage positively with theology in their professional work, even if they have deeply held religious convictions of their own. Theologians, on the other hand, can work freely and imaginatively with scientific ideas, and they do so with great enthusiasm in certain areas, especially if those areas concern origins—the origin of the universe and the origin of humans—or the sciences of the human mind. It should not be a surprise that these two kinds of academic discipline relate so differently to each other. The natural sciences cannot, by definition, consider supernatural causes, while theologians are freely able to engage reality on every level. The traffic is therefore largely one way.

But, and this is my personal observation, all of this means that the science and religion field has not done much to examine the contemporary scientific landscape beyond a relatively small number of key areas of direct theological interest (origins and mind). This also means that the science and religion field has not done much to help most working scientists understand the human dimensions of what they do, especially the intuitions, prejudices, and hunches that guide their science, some of which, as I mentioned, are implicitly theological.[1]

I will start to explore an example of such a potential hunch here—quantum fundamentalism—that is widespread in modern physics, despite its ominous title. I argue that science and religion—which has not addressed quantum fundamentalism yet as a worldview—is in a unique position to critique and inform it. I also suggest that a belief in quantum fundamentalism can be enriched by adopting a spirit of theological liberty.

Quantum Fundamentalism

Let us begin with the obvious questions: What is quantum fundamentalism, and why might theology take an interest?[2] I am sure you are thinking that theology has quite enough fundamentalisms of its own to be getting on with and does not need another. Perhaps, but humor me.

Quantum fundamentalism is the belief that quantum physics captures what is most fundamental in our physical world, even though we might be blissfully unaware of this in our daily lives.[3] I will give you a more watertight definition in a moment, but first of all, where does this belief come from? Why would anyone want to believe that quantum mechanics (QM) is ubiquitous if we do not see it? Ah, but we do see it.

QM is a highly mathematical theoretical framework originally developed about a century ago to predict what we might see when we make observations on light and on matter at the atomic and subatomic level. Note that it is a theory of measurement, not necessarily a theory of reality in itself (Ball 2019, 11). When we try to infer from QM what is really there before we make the measurement, the

theory tells us strange things. I am sure that you are familiar with Schrödinger's cat, the infamous thought experiment where the cat is ostensibly both alive and dead at the same time until we open the box and look inside. The philosophical disputes about what is really the case before we make the measurement have been ongoing for a century now and show no signs of abating. I will mention those disputes later, but my subject is the phenomenal success of QM at describing natural phenomena on all kinds of levels. We now think that QM works not just for matter at the atomic and subatomic levels, but all the way up.

The quantum framework is, arguably, the most successful collection of theories in the natural sciences, and not only if we measure success by the framework's ability to lend its terminology and explanations to the widest number of scientific disciplines. Without the quantum framework, we would have only a scant grasp of how matter on the molecular, atomic, and subatomic scales behaves and is composed. Accordingly, QM informs a wide variety of sciences, from most disciplines of the physical sciences through the life sciences and beyond. And on the macroscopic scale—preeminently the scale of human experience—quantum physics explains many of the basic properties of our everyday world that we take for granted, and it provides the basic science for entire industries, most obviously electronics and telecommunications. On a larger scale still, without quantum theory, we would not know why stars shine (nuclear fusion), nor how they make the atoms that make people and planets (nucleosynthesis). On still larger scales, cosmologists invoke quantum fluctuations as a mechanism for the Big Bang and an explanation for the distribution of galaxies in space (Wallace 2021, 94). Finally, QM underpins what is perhaps the greatest of all scientific discoveries: the periodic table of elements. Although the basic shape of the periodic table was determined by Dmitri Mendeleev in the late nineteenth century (some time before the quantum revolution), it was the realization that atoms have an internal electronic structure described by QM that provided the explanation for why the chemical elements fall into their characteristic rows and periods, and why they have many of the properties they do.

Of course, QM makes famously strange predictions. Albert Einstein, for one, could not accept the quantum property we now know as "entanglement"—where two particles can be inextricably linked even over stupendous distances—believing that QM must be incomplete or incorrect on this score at least. Yet, rigorous experimental tests to date have borne the quantum predictions out in full. Also, despite its widespread popular association with uncertainty, indeterminism, and probabilistic predictions, QM is capable of providing stunning precision in practical terms. In one renowned case (the magnetic moment of the electron), theoretical calculations agree with experiment to better than 1 part in 10^{12} (Fan et al. 2023). It is hardly surprising then, that many physicists believe that physical reality is fully quantum in nature. Here, for instance, is how one of the standard textbooks on QM begins (Cohen-Tannoudji, Diu, and Laloë 2020, 23–24):

> In the present state of scientific knowledge, quantum mechanics plays a fundamental role in the description and understanding of natural phenomena. In fact, phenomena that occur on a very small (atomic or subatomic) scale cannot be explained outside the framework of quantum physics. For example, the existence and properties of atoms, the chemical bond and the propagation of an electron in a crystal cannot be understood in terms of classical mechanics . . . Actually, there are many phenomena that reveal, on a macroscopic scale, the quantum behaviour of nature. It is in this sense that it can be said that quantum mechanics is the basis of our present understanding of all natural phenomena, including those traditionally treated in chemistry, biology, etc.

QM is the basis of our present understanding of all natural phenomena, they say, extending even into the macroscopic regime. In case you are wondering why classical physics does not take over at such a length scale—especially on our human level of pets and people, cats and dogs, where we apparently see no quantum strangeness—here is the great theorist and champion of condensed matter physics (that is, the physics of things on many length scales from the microscopic to the macroscopic), Philip Anderson (2001, 500):

> For 75 years, physicists—perhaps overwhelmed by the prestige of the authors of the Copenhagen interpretation—have adhered to the idea that there is some mysterious scale at which the quantum world changes into the classical. Decoherence is simply the code word for the null hypothesis that there is no such scale, that we are quantum all the way up. I have for long felt that the question was settled by the many examples of macroscopic quantum coherence, but I am notoriously impatient and perhaps the delicate experimental work which has gone into pinning these questions down is worthwhile. There is absolutely no experimental evidence for such a scale.

"We are quantum all the way up," maintains Anderson. Decoherence—the process by which a quantum entity's characteristic signature of quantum coherence becomes progressively watered-down by entanglement with its environment—may, for all practical purposes, make a quantum entity appear to us to behave in a classical fashion when we observe it, but the entity and its environment are no less quantum for that, in their deep physical description. Classical physics, successful as it has been, is therefore largely an approximation, we suspect, to a more profound description of nature on all length scales, which must be quantum mechanical. Of course, we currently are not sure how to combine QM with the other great theoretical success story of modern physics—general relativity—but we expect that, sooner or later, a solution will be found, and it will be thoroughly quantum in both style and content. This is a widespread hope in modern physics, and versions of it can be seen

throughout the popular science literature. In short, quantum fundamentalism is uncontroversial in mainstream physics, although some of us wonder whether QM will ever be able to replace our reliance on thermodynamics in areas such as my own, condensed matter physics. Well, that is an open, and rather technical question,[4] but I have to admit that condensed matter physics would be completely unintelligible without QM. In that sense, I am undoubtedly as much of a quantum fundamentalist as anyone.

Strangely, although physicists have operated for decades under the assumption that the universe is largely or wholly quantum on every level, the terminology of "quantum fundamentalism" has only been coined quite recently, and used mainly in the quantum foundations field, which specializes in the deep philosophical problems of the quantum world. This is why you probably have not come across the term before, even if you work in physics and are perhaps beginning to realize, "Hold on, I think I might be a quantum fundamentalist." It is a powerful term for capturing a worldview, which is where theology enters. But first, I promised you a watertight definition. Here it is, from philosopher of science, Henrik Zinkernagel (2016, 2; cf. Faye 2019; Zinkernagel 2011, 235): "Quantum Fundamentalism. Everything in the universe (if not the universe as a whole) is fundamentally of a quantum nature and ultimately describable in quantum-mechanical terms."

Everything in the universe is quantum in its very being, and ultimately (that is, in principle), everything can be described scientifically in such terms, explains Zinkernagel. Such is quantum fundamentalism: the quantum is here to stay. But notice that this definition is not entirely watertight. We need to examine the word "fundamentally," since it is doing much of the work.

Let us turn to philosophy of physics. There is a healthy dispute here about fundamentality (Aguirre, Foster, and Merali 2019; Morganti 2020; Tahko 2023). One rather old-fashioned view, which is still widespread in physics today but which has become unfashionable in philosophy, takes the physical world to be effectively stratified into many different levels, as described by the many different sciences.[5] The most fundamental level, on this view, is the level that describes the tiniest possible things: the subatomic quantum particles like the various quarks and the Higgs boson. This level is the basement on which the rest of nature is built, since those particles and quantum fields make up the matter of which every entity in the universe is composed; their behavior underpins every other level, we assume. This is a common form of quantum fundamentalism, but it is better known by the name of reductionism. Here, for instance, is how particle physicist Sabine Hossenfelder (2022, 85–87) defines "fundamental": "A fundamental property or object cannot be derived from or reduced to anything else . . . [T]he only fundamental theories we currently know of—the currently deepest level—are the **standard model of particle physics** and Einstein's general relativity, which describes gravitation."

The rest of physics, she goes on to say—together with chemistry, biology, psychology, sociology, and so on all the way up the hierarchy—all emerges from those fundamental laws of particle physics and gravitation (Hossenfelder 2022, 87). This is a full-blooded account of reductionism, which Hossenfelder is happy to own, since she regards reductionism as a well-established fact of nature (Hossenfelder 2019, 85; 2022, 82).

But not all physicists are so committed to reductionism, especially those of us who work on emergent phenomena at a much higher level than that of subatomic particles, such as in my own area of condensed matter physics. I—and many other condensed matter physicists—take a more horizontal (or at least less stratified) view of the sciences, and we find that the quantum framework is just as effective. Hence, I say that quantum fundamentalism is not the same as reductionism. We need not maintain that all of science boils down to the behavior of subatomic particles just to believe in the reality and ubiquity of quantum phenomena.

Likewise, many philosophers of science today have their doubts about reductionism (e.g., Adlam 2019, 9). Alyssa Ney, for one, has proposed a helpful alternative (Ney 2019). She explains fundamentality by looking at the ability of some scientific theories to reach far across the sciences, to influence and underpin many other theories. The "fundamental" term therefore refers to what she calls the "explanatory maximality" of a theory. Accordingly, a theory is fundamental if it is "a common source of (causal and constitutive) explanations that possess the greatest [that is, the maximal] degree of scope, accuracy, and precision of all theories that have so far been formulated" (Ney 2019, 33). In other words, a fundamental theory simply reaches much further than others, she thinks, into more areas of science. Such a theory also works to unify other scientific theories by lending them its own concepts. Such a theory may not yet be complete, nor perfect, but if it is maximally far-reaching "in ordinary scientific contexts" then it is fundamental on this view (Ney 2019, 34).

Now, Ney does not articulate the following conclusion herself, but I suggest that the quantum framework qualifies strongly as a fundamental body of theory according to the logic of her argument,[6] since its concepts and explanations are ubiquitous across the sciences. No other theoretical framework is as far-reaching, nor as adaptable, I say. And no other theoretical framework is as able to unite the patchwork landscape of sciences quite so effectively either. Remember Dr. Idreos's vision of the unity of peoples. Quantum fundamentalism may not unify nations (or not yet), but it provides a starting point for a unified concept of nature, a goal that, incidentally, has been a long-standing ambition of the science and religion field (Harris 2024).

Quantum fundamentalism has a lot going for it then, and I want to commend it to you if you are not yet a convinced quantum fundamentalist. But there are some very serious downsides. Here is cosmologist Lee Smolin (2019, 3) on the matter: "Quantum mechanics has been the core of our understanding of

nature for nine decades. It is ubiquitous, but it is also deeply mysterious. Little of modern science would make sense without it. But experts have a hard time agreeing what it asserts about nature."

Quantum fundamentalism can claim an impressive degree of empirical evidence in its favor then, but it struggles to provide a clear guide to reality. The basic problem is that the standard mathematical formalism of QM does not tell us clearly "this is what reality is" but rather what we should expect when we perform measurements (Ball 2019, 11). For, and in spite of the grand claims often made at the popular level, QM itself is frustratingly silent on the nature of reality when we are not looking, as it were. If we want to know how things really are in themselves, we must apply a metaphysical interpretation to the physics, but there are many such interpretations in circulation, and they tell of wildly different—if not incompatible—accounts of the natural world (Lewis 2016, 179–82; Barrett 2019, 231). Presumably one of them is right, but we do not know which, nor how to tell. Each has its supporters. You will no doubt have heard of the Copenhagen interpretation, which avoids making definitive pronouncements about what the quantum world is like when we are not looking, or the many-worlds interpretation, which says that there are many branches of reality coexisting (many worlds, effectively): the cat is alive in one branch, dead in another. But there are still other interpretations. All of them are challenging (if not outlandish) at a human level. They raise difficult questions about what we take for granted in our everyday experience, such as the distinction between cause and effect, or between subject and object.

These are serious problems for quantum fundamentalism and should not be underestimated. The science of QM has bound us to a set of more or less absurd scenarios and abandoned us there. The science has given us freedom to choose between the interpretations but no liberty to make an informed decision. To put it bluntly, quantum fundamentalism may claim massive scientific support, but if it cannot help us to make human sense of the science, then what earthly good is it as a worldview?

The humanities subjects can help. What might quantum fundamentalism mean for human being and human becoming when it is seen through a humane perspective, illuminated by subjects such as the creative arts, or literature, or even theology and ethics? This is the wider aim of our project here in Oxford: what we are beginning to call the "quantum humanities." Among those perspectives, I suggest that theology is particularly farsighted, which is why I want now to speak of theological liberty.

Theological Liberty

There is one obvious way theology can help quantum fundamentalism. If there are many competing quantum interpretations of reality that are empirically equivalent, then the only way to adjudicate between them is to bring our human

suspicions and convictions about what lies beyond the measurable world to bear. Theology therefore represents a privileged vantage point, a liberty inaccessible to science of itself.

But theology can shed yet more light. Notice that when I was trying to describe quantum fundamentalism earlier, I was forced into using words like "aspire," "hope," "suspicion," "hunch," "conviction," words more appropriate to the maintenance of a worldview than a firmly empirical fact of the matter. Quantum fundamentalism—for all its scientific support—is in the same conceptual space as theology when it draws conclusions about the way the world really is. Few physicists commit to one or other of the quantum interpretations in practice—partly because we have the same misgivings as everyone else – while we believe firmly in the reality of a quantum world (Harris 2023, 192). In other words, we adopt a kind of vague but aspirational quantum realism that draws us into the same conceptual space, I suggest, as theology.

In case you do not believe me, let me point out that the religious uptake of QM has been remarkable. Quantum fundamentalism has been a gift to religious and spiritual commentators as well as lifestyle gurus. Like the physicists, such people adopt a vague but aspirational quantum realism, but they are in the conceptual space of theology to begin with. Christian theologians, for instance, have found that QM is a fantastically rich source of metaphor for models of God's activity in the world and of God's relationship with the natural order, for the mysteries of Christ as fully human and fully divine, and even for that deep enigma of Christian thought: the nature of God's Trinitarian being.[7] Characteristic features of QM—wave/particle duality, indeterminacy, and entanglement—crop up repeatedly. Not only do these features offer stunning analogies for theological mysteries but they allow the contemporary theologian to retrieve traditional categories of God's presence and providence in the world, categories the rigidly mechanical pictures of classical physics seemed to have precluded for good before the quantum came along.

Not only Christian theologians but many other religious thinkers have borrowed enthusiastically from QM to support ancient spiritual worldviews, especially from Hinduism, Buddhism, and Daoism. Fritjof Capra's *Tao of Physics* led the way in what has become a fertile cultural trend (Capra 1976), a synthesized "quantum mysticism" that downplays the material world in favor of mind, consciousness, and the deep unity of all things. Quantum mysticism has been popular in New Age spirituality, for instance, and has percolated far and wide into health and lifestyle consultancy, coaching, and alternative medicine.[8]

Let me illustrate with two quantum lifestyle books, which are not ostensibly about religion, although they both take thoroughly mystical angles on QM. First, Deepak Chopra's book, *Quantum Healing* (Chopra 2015). This has been a remarkably successful book, still in print 35 years after it was first published in 1989. It is important to know that Chopra is a medical doctor who specializes

in oncology. In the book, he describes how many of his patients experience miraculous remissions of their cancer symptoms by practicing the spiritual techniques he prescribes, inspired by his interest in Hinduism. His point is that the human mind influences the development of our body, even our genes; QM confirms this for us, Chopra argues. The physics teaches us about an invisible quantum domain—the most fundamental dimension of reality—from which mind and body emerge and unite (Chopra 2015, xii). Every cell in our body is conscious, since on the quantum level we are fully embedded in one mind, a cosmic intelligence that creates, governs, and controls reality (Chopra 2015, xviii). Chopra describes how, when his patients learn this truth, they make a quantum leap to a new level of consciousness, and their symptoms improve (Chopra 2015, 9).

A much more recent book, Larry Farwell's *The Science of Creating Miracles* (Farwell 2021), makes similar moves, but it does not focus on healing or cancer so much as your entire life.[9] If you are unhappy in your career or relationship, QM can help you, Farwell promises. He believes that consciousness is fundamental because QM is fundamental, and that through our own consciousness, we can each alter the outcome of quantum events in the universe (since such events are observer dependent, he argues). This means, Farwell believes, that each of us can control our future, and ensure the best outcome, even if such an outcome is extremely unlikely or apparently impossible (a miracle). As he explains: "Since the entire physical universe consists of quantum mechanical particle/wave phenomena, this means that we can consciously influence the physical universe to move in accord with our desires and intentions" (Farwell 2021, 127).

There is less explicit religion in Farwell's book than Chopra's—Farwell repeatedly emphasizes his scientific background and successes—but like Chopra he is influenced by varieties of Hindu and Buddhist idealism, where consciousness is primary and the physical is secondary.

These are just two examples of well-known authors in the genre. A few seconds' googling will reveal many more resources. But the backlash from professional scientists has been uncompromising. "Pseudoscience" is one of the more polite accusations leveled at quantum mysticism (Hassani 2016). Here are some more: "quantum flapdoodle" (Barr 2016, 91), "quantum hype" (Polkinghorne 2007, ix), "quantum woo" (Moriarty 2023), and, perhaps the most widespread insult, "quantum quackery" (Shermer 2005). The charge being made here is that the proposals of mystical thinkers like Chopra and Farwell are scientifically illiterate at best—showing no great understanding of the quantum science—and potentially harmful at worst, undermining mainstream medicine by peddling untested fixes to vulnerable people. These criticisms are important, I feel. Like many other scientists, I object strongly to the misrepresentation of science.

But I am not sure that these self-help manuals can be dismissed too easily, at least not their religious and spiritual aspirations. Like I said

earlier, quantum fundamentalism moves in the same conceptual space as theology. The ordinary non-physicist realizes this. She may well find the technicalities of QM impenetrable (including the conceptual questions about its interpretation), but she knows full well that quantum physics is deeply strange and mysterious when we try to apprehend it on a human level. She can hardly be blamed when she concludes that the science confirms ancient religious wisdom about the mystery at the heart of reality—that consciousness and mind are not secondary byproducts of the physical — especially when some of the great figures of twentieth-century quantum physics, like Niels Bohr, Werner Heisenberg, and Erwin Schrödinger, made similar mystical connections in their popular-level writing. Quite simply, quantum fundamentalism exerts a powerful pull on the modern religious imagination. Another way of saying this is that ordinary human beings (by which I mean the vast majority who are not physicists or philosophers) need the mediation of religious and spiritual interpretations in order to apprehend quantum fundamentalism.

Let me put that in terms of two thesis statements: the central argument of my article, if you like. First, ordinary human beings frequently grasp for religious or spiritual or mystical mediation when they want to make sense of quantum fundamentalism. Second, quantum fundamentalism needs responsible (that is, in tandem with full and informed respect for the science) religious and spiritual enrichment in order to liberate human beings.

But do not take my word for it. Here is Erwin Schrödinger (2014, 95–97):

> [T]he scientific picture of the real world around me is very deficient. It gives a lot of factual information, puts all our experience in a magnificently consistent order, but it is ghastly silent about all and sundry that is really near to our heart, that really matters to us. It cannot tell us a word about red and blue, bitter and sweet, physical pain and physical delight; it knows nothing of beautiful and ugly, good or bad, God and eternity. Science sometimes pretends to answer questions in these domains, but the answers are very often so silly that we are not inclined to take them seriously . . . And the reason for this disconcerting situation is just this, that, for the purpose of constructing the picture of the external world, we have used the greatly simplifying device of cutting our own personality out, removing it; hence it is gone, it has evaporated, it is ostensibly not needed . . . [T]his is the reason why the scientific world-view contains of itself no ethical values, no aesthetical values, not a word about our own ultimate scope or destination, and no God, if you please. When came I, wither go I? . . . Science is, very usually, branded as being atheistic. After what we have said, this is not astonishing. If its world-picture does not even contain blue, yellow, bitter, sweet—beauty, delight and sorrow,—if personality is cut out of it by agreement, how should it contain the most sublime idea that presents itself to human mind?

Taking Schrödinger's lament on the scientific method to heart, I suggest that quantum fundamentalism needs to be liberated theologically. Human beings need a supplementary method to that of science in order to apprehend science in humane terms, a method that can recognize blue, yellow, bitter, sweet—not to mention the "most sublime idea that presents itself to the human mind." We need a method that celebrates our joys and sorrows, one that liberates the subject to contemplate the object with wonder. Theology is clearly in view here, but that should come as no surprise if you agree with me that quantum fundamentalism took us into that conceptual space to begin with.[10]

Therefore, to determine what quantum fundamentalism means for human beings—for both scientists and non-scientists—I suggest that we need to supplement the scientific method with a further methodology that is equipped to take religious, ethical, and aesthetic convictions about reality seriously, all the while treating the science respectfully and responsibly. Any claim about what is most fundamental to our existence will automatically have religious and ethical implications, since questions concerning human flourishing—by which I mean the growing into mature realization of our deepest-held values and purposes—come into focus, questions that cannot be answered empirically even if they first arise from empirical quantum physics. Is there a methodology that can handle all these weighty demands? Happily, there is just such a methodology at hand. It is called science and religion.[11]

As I draw to a close then, let me flag up where I want to go from here using the theological and philosophical tools of the science and religion enterprise. I should point out that this article is a scene-setter for a much bigger collaborative enterprise taking place in Oxford and beyond. There are many avenues that our research group plans to take to investigate the human and theological dimensions of quantum fundamentalism. We have already begun to look at some of the quantum conceptual interpretations in ethical, theological, and existential terms, but we also want to bring a fresh pair of eyes to other implicitly theological issues in modern science, such as the status of quantum physics as laws of nature. For instance, there is a venerable tradition going back at least as far as Newton that says the laws of nature are expressions of God's will as creator of all. The basic question is this: What on earth was God thinking when God invented QM and put it at the heart of creation? That is a serious theological question, and one that has barely been addressed yet, but I hope we can provide some serious constructive answers. In the fullness of time, we are ambitious to persuade other humanities perspectives to be involved—art, film, literature, for instance—to help us investigate what quantum fundamentalism means for human flourishing. But that is some way off.

My final word is simply to express a personal aspiration: I hope that I have done enough here to convince you that there is a viable research program in view, and that this may have useful things to say to working scientists who might just possibly also be human beings.

Acknowledgments

My thanks go to colleagues in Oxford and beyond who are involved with this research project. Especially, I would like to thank Shaun Henson, Emily Qureshi-Hurst, and Pete Jordan—all members of the Ian Ramsey Centre for Science and Religion—and the Issachar Fund, who generously provided the means for us to develop our interests in the first place. My greatest thanks go to my long-standing scientific collaborator, Steve Bramwell, who is not involved with this project, but who nevertheless has taught me so much about the quantum world over the years.

Notes

[1] This is not to say that there has been no attempt to examine the religious convictions of working scientists; the work of social scientists such as Elaine Ecklund, Renny Thomas, and coworkers stands out for pointing to the great richness and variety of religious attitudes among working scientists around the globe.

[2] I have already begun to examine theological interest in quantum fundamentalism—especially the divine action debate—in an earlier publication (Harris 2023).

[3] I will use various terms to refer to the amalgam of quantum ideas that inform quantum fundamentalism, such as quantum science, quantum physics, quantum theory, and the quantum framework. At their heart is the mathematical formalism of the 1920s and 1930s known as quantum mechanics, upon which all later theoretical developments—including quantum electrodynamics and quantum field theory – are based. I will therefore also speak of QM as a set of core ideas that holds a wider body of quantum theory and key experimental results together.

[4] Steane (2018, 30–36) considers the relationship between thermodynamics and quantum theory in detail, especially the question of which is more fundamental. He declines to opt for one or the other, insisting that both play a valid and important role in our understanding of the physical world. Incidentally, I have personally taken much inspiration from Steane's book; his subtitle of "A Humane Philosophy of Science and Religion" is a perfect headline for what I am hoping to achieve in my own small way in this article.

[5] This is the view developed in Paul Oppenheim and Hilary Putnam's famous paper on the unity of science (Oppenheim and Putnam 1958; cf Ney 2019, 29).

[6] Ney does not make this point presumably because she has in mind the goal of a "unified theory" (Ney 2019, 34) rather than the loose agglomeration of theories, mathematical tools, and methodologies that make up the quantum framework.

[7] Of a very rich and diverse theological literature, I merely cite some representative examples: Kaiser (1976), O'Murchu (2004), Russell et al. (2001), Saunders (2002), Simmons (2014), Wegter-McNelly (2011), and Wildman (2004).

[8] The history behind this amazingly successful cultural development is told to great effect by David Kaiser (2011).

[9] Note that this book is enthusiastically endorsed by Chopra himself.

[10] Another way of developing this point is to say that realist interpretations of QM tend to be deeply challenging on a human level, since they propose ontologies that conflict strongly with the everyday world of human experience. It is no coincidence that these are the very same interpretations that are most guilty of "cutting our own personality out" (as Schrödinger put it in the quotation), i.e., of removing the subjective observer from consideration. Antirealist/idealist interpretations, on the other hand, are much easier to grasp on a human level, especially if one is already predisposed to uphold the primacy of mind and consciousness, as in many Hindu, Daoist, and Buddhist religious traditions (but also in some Abrahamic traditions). These antireal/idealist approaches emphasize (rely upon) the presence of the subjective observer, and so are much more readily apprehended at the human level than the realist interpretations.

[11] I realise that what I have said in this sentence may come across as a pat and evasive answer, but my point is serious. The development of such a supplementary methodology to the scientific method has been a central research program of the science and religion field for decades. Notwithstanding the fact that this research program has not reached a satisfactory conclusion yet, it indicates the key relevance of the science and religion field to the humane dimensions of science. Those of us who work in the field are highly motivated to uncover and analyze the deep undercurrents of science and its discoveries, and we do so using philosophical, theological, historical, and sociological tools. In short, we realize that science cannot be understood without its humane significance. To repeat what I said much earlier in this article: Why does science work? Science, of itself, is unable to answer that question, but needs to call upon humanities disciplines for help, preeminently philosophy and theology, and increasingly (I suggest) other humanities disciplines too. This is the motivation behind the "quantum humanities" research project based in and around Oxford.

References

Adlam, Emily. 2019. "Fundamental?" In *What Is Fundamental?*, edited by Anthony Aguirre, Brendan Foster, and Zeeya Merali, 5–12. Cham, Switzerland: Springer.

Aguirre, Anthony, Brendan Foster, and Zeeya Merali, eds. 2019. *What Is Fundamental?* Cham, Switzerland: Springer.

Anderson, Philip W. 2001. "Reply to Cartwright." *Studies in History and Philosophy of Modern Physics* 32 (3) 499–500.

Ball, Philip. 2019. *Beyond Weird: Why Everything You Thought You Knew About Quantum Physics is . . . Different*. London: Vintage.

Barr, Stephen M. 2016. *The Believing Scientist: Essays on Science and Religion*. Grand Rapids, MI: Eerdmans.

Barrett, Jeffrey A. 2019. *The Conceptual Foundations of Quantum Mechanics*. Oxford: Oxford University Press.

Capra, Fritjof. 1976. *The Tao of Physics: An Exploration of the Parallels between Modern Physics and Eastern Mysticism*. London: Fontana.

Chopra, Deepak. 2015. *Quantum Healing: Exploring the Frontiers of Mind/Body Medicine*. New York: Bantam.

Cohen-Tannoudji, Claude, Bernard Diu, and Franck Laloë. 2020. *Quantum Mechanics. Volume I: Basic Concepts, Tools, and Applications*. Weinheim, Germany: Wiley-VCH.

Fan, X., T. G. Myers, B. A. D. Sukra, and G. Gabrielse. 2023. "Measurement of the Electron Magnetic Moment." *Physical Review Letters* 130:071801.

Farwell, Larry. 2021. *The Science of Creating Miracles: Neuroscience, Quantum Physics, and Living the Life of Your Dreams*. New York: B. C. Allen Publishing and Tonic Books.

Faye, Jan. 2019. "Copenhagen Interpretation of Quantum Mechanics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2019/entries/qm-copenhagen/.

Harris, Mark. 2022. "Natural Selection at New College: The Evolution of Science and Theology at a Scottish Presbyterian Seminary." *Zygon: Journal of Religion and Science* 57 (3): 525–44.

———. 2023. "Quantum Theology Beyond Copenhagen: Taking Fundamentalism Literally." *Zygon: Journal of Religion and Science* 58 (1): 183–202.

———. 2024. "Introduction." In *God and the Book of Nature: Experiments in Theology of Science*, edited by Mark Harris, 1–10. Abingdon, UK: Routledge.

Hassani, Sadri. 2016. "Commentary: The Dangerous Growth of Pseudophysics." *Physics Today* 69 (5): 10–11.

Hossenfelder, Sabine. 2019. "The Case for Strong Emergence." In *What is Fundamental?* edited by Anthony Aguirre, Brendan Foster, and Zeeya Merali, 85–94. Cham, Switzerland: Springer.

———. 2022. *Existential Physics: A Scientist's Guide to Life's Biggest Questions*. London: Atlantic Books.

Kaiser, Christopher B. 1976. "Christology and Complementarity." *Religious Studies* 12 (1): 37–48.

Kaiser, David. 2011. *How the Hippies Saved Physics: Science, Counterculture, and the Quantum Revival*. New York: W. W. Norton & Company.
Lewis, Peter J. 2016. *Quantum Ontology: A Guide to the Metaphysics of Quantum Mechanics*. Oxford: Oxford University Press.
Morganti, Matteo. 2020. "Fundamentality in Metaphysics and the Philosophy of Physics. Part II: The Philosophy of Physics." *Philosophy Compass* 15:e12703.
Moriarty, Philip. 2023. "Beyond the Quantum Woo-niverse." *Physics World* 36 (6): 56–57.
Ney, Alyssa. 2019. "The Politics of Fundamentality." In *What Is Fundamental?*, edited by Anthony Aguirre, Brendan Foster, and Zeeya Merali, 27–36. Cham, Switzerland: Springer.
O'Murchu, Diarmaid. 2004. *Quantum Theology: Spiritual Implications of the New Physics*. New York: Crossroad.
Oppenheim, Paul, and Hilary Putnam. 1958. "Unity of Science as a Working Hypothesis." In *Concepts, Theories, and the Mind-body Problem*, edited by Herbert Feigl, Michael Scriven, and Grover Maxwell, 3–36. Minneapolis: University of Minnesota Press.
Polkinghorne, John. 2007. *Quantum Physics and Theology: An Unexpected Kinship*. London: SPCK.
Russell, Robert John, Philip Clayton, Kirk Wegter-McNelly, and John Polkinghorne, eds. 2001. *Quantum Mechanics: Scientific Perspectives on Divine Action*. Vatican City, Berkeley: Vatican Observatory and CTNS.
Saunders, Nicholas. 2002. *Divine Action and Modern Science*. Cambridge: Cambridge University Press.
Schrödinger, Erwin. 2014. *Nature and the Greeks and Science and Humanism*. Cambridge: Cambridge University Press.
Shermer, Michael. 2005. "Quantum Quackery." *Scientific American* 292 (1): 34.
Simmons, Ernest L. 2014. *The Entangled Trinity: Quantum Physics and Theology*. Minneapolis: Fortress.
Smolin, Lee. 2019. *Einstein's Unfinished Revolution: The Search for What Lies Beyond the Quantum*. London: Penguin.
Steane, Andrew. 2018. *Science and Humanity: A Humane Philosophy of Science and Religion*. Oxford: Oxford University Press.
Tahko, Tuomas E. 2023. "Fundamentality." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/archives/win2023/entries/fundamentality/.
Wallace, David. 2021. *Philosophy of Physics: A Very Short Introduction*. Oxford: Oxford University Press.
Watt, Hugh. 1946. *New College Edinburgh: A Centenary History*. Edinburgh: Oliver and Boyd.
Wegter-McNelly, Kirk. 2011. *The Entangled God: Divine Relationality and Quantum Physics*. New York: Routledge.
Wildman, Wesley. 2004. "The Divine Action Project, 1988–2003." *Theology and Science* 2 (1): 31–75.
Zinkernagel, Henrik. 2011. "Some Trends in the Philosophy of Physics." *Theoria* 26:215–41.
———. 2016. "Are We Living in a Quantum World? Bohr and Quantum Fundamentalism." Eprint arXiv:1603.00342.

Zygon®: Journal of RELIGION & SCIENCE

Miller, Gordon L. 2024. "The Soul of the Frog for a Postmodern Age: Updating a Classic Debate." *Zygon: Journal of Religion and Science* 59 (3): 609–37. DOI: https://doi.org/10.16995/zygon.16572

Open Library of Humanities

The Soul of the Frog for a Postmodern Age: Updating a Classic Debate

Gordon L. Miller, Professor Emeritus of Environmental Studies at Seattle University, Seattle, WA, USA, glmiller@seattleu.edu

The history of modern science bears abundant witness to the fact that the humble frog has yielded myriad insights into the intricacies of anatomy and physiology far beyond frogs, producing several Nobel laureates in the process. A concurrent but less heralded inquiry and debate, however, has wrestled with the harder question of, as Thomas Henry Huxley put it in 1870, "Has a Frog a Soul?," with frogs serving as convenient experimental subjects for issues of the human soul and body. Here, I present highlights from this debate among leading eighteenth- and nineteenth-century scientists representing animistic and burgeoning mechanistic views, with the latter effecting a significant contraction of soul, ending with William James in 1890. I then bring the inquiry up to date by showing how Alfred North Whitehead's philosophy of organism and some recent physiological research offer promising and constructive postmodern hope for addressing the dispiriting loss of soul, on various levels, that afflicts much of modern life.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

In the late seventeenth century, the great Dutch anatomist and microscopist Jan Swammerdam (1758, 105) declared that "there is a much greater number of miracles, and natural secrets in the Frog, than anyone hath ever before thought of or discovered." With the fine details of development and muscular action he highlighted in these rather lowly animals, he linked together the realms of insects, amphibians, and human beings and challenged the traditional belief that animal anatomy and physiology lack relevance for humans, arguing instead that a single set of divine laws governs them all (Sleigh 2012b). Swammerdam was not alone in recognizing the promise of the frog, for it certainly became the organism of choice from his day to our own for a vast range of often revolutionary biological inquiries into, for example, the heart, the lungs, muscle contraction, nerve conduction, visual perception, and ecological degradation. The great significance of these studies is due largely to the venerable and scientifically productive assumption, championed by Swammerdam, that the frog can, in many ways, serve as a useful model or microcosm of mammals, most notably of human beings (Burggren and Warburton 2007; Gilbert 1965, 3–13; Holmes 1993).

The distinctive physical qualities of frogs (representing the Anuran order of amphibians) have served researchers well in their varied anatomical and physiological pursuits throughout much of the history of modern science. But what of the frog's nonphysical qualities, assuming such exist? Has modern science achieved insights into the mysteries of nature on this front as well, affording notable revelations regarding the workings of mind, or consciousness, or what has traditionally been called the soul? Thomas Henry Huxley ([1870] 2015) posed the basic question directly in an influential speech titled "Has a Frog a Soul?" But Huxley's inquiry was merely a penultimate foray in a public debate that had gradually played out in Europe from the mid-eighteenth century to the dawn of the twentieth. In this empirical and ideological contest, the contrasts between animistic and mechanistic views were clearly drawn, and in the supporting series of experimental studies, the frog proved, primarily because of its amazing tenacity of life even in extreme conditions, to be an ideal experimental subject, offering insights into the nexus of body and soul through methods unthinkable with humans. And as with the more purely physiological investigations noted, the longsuffering frog in this case as well was clothed in larger significance in the minds of many as a homologue for humans, this time regarding whether, where, and in what way *we* might have a soul.

An inquiry into questions of the soul in the early decades of the twenty-first century is complicated by the fact that, despite the presence of a multitude of firm believers, many among us are no longer so sure that even we humans have what might properly be called a "soul." We are obviously quite sure we have a mind, and perhaps a self, but the idea of a soul comes trailing suggestions of

immateriality and immortality, notions widely questioned or rejected in current Western culture (Crabbe 1999, 1–7; Sorabji 1999). And recent work in biology, neuroscience, and cognitive science has offered abundant experimental evidence regarding the anatomy and physiology of functions traditionally attributed to the soul and has thus set many to thinking, or concluding, that even the mind or the self are at bottom nothing but material processes (Murphy 1998, 11–19). But what of the unconscious mind, which many of us are also fairly certain is part of ourselves? Is this more shadowy realm of passions, portents, and symbolism perhaps more soul-like in nature? Perspectives obviously differ on all these points.

In the early twentieth century, Carl Jung (1933) saw the growing interest in depth psychology, paranormal phenomena, and spirituality as evidence that many moderns were apparently "in search of a soul." But as the century wore on, other thinkers suggested that the search had evidently not enjoyed widespread success, with Jacob Needleman (1980) writing of "the lost doctrine of the soul" and Thomas Moore (1992, xi) declaring that "the great malady of the twentieth century . . . is 'loss of soul.'" This contemporary cultural atmosphere, beset with uncertainty regarding the nature or existence of a human soul, makes it a challenging place from which to see clearly into the further question of animal souls.

While there are undoubtedly many meanings of soul in every age, something that becomes clear from a historical perspective is that in the premodern era, and even somewhat into the eighteenth century, the extent of soul, in some form, was generally considered much broader with regard to both its place in the human body and its presence in the nonhuman natural world. Whereas the dominant modern coordinates generally limit it to the conscious mind—often only of humans—and its corresponding brain, in the ancient world, for example, *psyche* was seen not only as the fire of consciousness but more fundamentally as the force of life itself. In his influential three-part theory of the soul, Plato prioritized the brain and spinal cord as the seat of the immortal and uniquely human rational soul (*Timaeus* 73b–d), while the other parts of the soul, although somewhat subservient to the dictates of the head, were also indispensable and more extensive. He located the spirited or emotional part of the soul, shared with other animals, in the chest and the appetitive part, associated with physiological needs and desires and shared with both animals and plants, in the abdomen, particularly the liver. The Platonic soul thus pervades the body to some degree and exists as a microcosm of an infinitely more expansive soul—the universally animating World Soul, an animistic or panpsychic conception also evident in previous Greek thought that envisioned a thoroughly ensouled natural world (Lorenz 2009; Skrbina 2017, 37–52).

Aristotle also proposed a tripartite scheme, but in place of Plato's spirited and appetitive souls, he envisioned sensitive and nutritive functions respectively

and saw them as animating all an organism's vital and responsive processes—nutritive in plants and nutritive plus sensitive in animals. In animal and human bodies, he located these essentially in the heart, with the heat in this region signifying an active soul. Regarding the rational, distinctly human aspect of the soul, he thought, contrary to Plato, that it, though dependent on the nutritive and sensitive functions, did not reside in any specific organ or region of the body (Lorenz 2009; Frede 1995). The burgeoning of anatomical knowledge derived from animal and human dissections in third century BCE Alexandria, particularly the neurological discoveries, fostered a fuller grounding of the soul in the details of anatomy and physiology. Galen continued this general approach into the third century CE, placing Plato's tripartite emphasis on the brain, the heart, and the liver on a firmer anatomical foundation and, contra Aristotle, designated the brain as the primary seat of the soul (Debru 2015).

Part One: The Classic Modern Search for Soul

The influence of thinkers from classical antiquity, especially Plato and Aristotle, regarding the nature and location of soul reached through the succeeding centuries and has been felt in various ways in the modern experimental search for the soul in the accommodating physiology of frogs. But before launching their inquiries, these eager experimentalists needed to deal with the other preliminary challenge facing inquirers at the outset of their search—determining the most reliable outward signs of a soul. Again they looked to the ancients. Plato expressed a widely held view in his suggestion that "self-generating motion" is the defining feature of soul (*Laws* 896a–c). Aristotle also considered motility a very basic sign of the animal soul but emphasized as well that, at the various levels of his three-part soul, the nutritive functions of metabolism and growth (in all living things), the sensitive functions of perception and movement (in animals), and the singularly human activity of rational thought were also indicative of soul. Considering that the inquiries into frogs in more recent centuries were pursued in the context of experimental physiology, the perception and movement disclosing an animal soul were not the unprovoked self-motion observed in field settings evoked by natural stimuli but rather the sensitivity and muscular response to a stimulus or test imposed by the experimenter. And, as I also discuss later, some participants in the modern debate found that a more specific criterion—not mere motility but distinctly *purposive* self-movement—was better suited to their needs in signifying the presence of some measure of coordinated inward experience constituting mind or soul.

With an at least working assumption that the primary bodily seat of the soul is most likely the brain, though perhaps with some degree of ensoulment in other regions, and that self-generated movement of various grades is its most reliable sign, the search began in earnest with scalpel and probe. Aristotle (*De Anima* 411b) had observed that plants as well as certain animals "go on living

when divided into segments," meaning that "each of the segments has a soul in it" because "the segments for a time possess the power of sensation and local motion." Accordingly, the general strategy in later centuries for testing the accuracy of these assumptions in the actual physiology of a frog was to excise or otherwise disconnect or destroy the presumed seat from the surrounding soma and then see what signs of soul remained, an operation apparently first documented by Leonardo da Vinci around 1485 in his own search for the soul (Del Maestro 1998). This typically involved the traditional and torturous procedure of pithing, in order to sever the spinal cord and scramble the brain, and often the even more gruesome technique of decapitation, rendering the frog essentially brain dead but otherwise still alive. (In recent decades, these procedures have been discouraged or disallowed in laboratory settings in many countries or permitted only on previously anesthetized animals. For a brief sketch of ethical considerations regarding experimentation with frogs from the perspective of the philosophy of organism as presented in this article, see the conclusion.)

The metaphysical engine driving much of this research, especially with the impetus provided by Descartes, was the rising energy of the mechanist's quest to determine what regions of the living world, if any, can be explained simply by the known laws of physics. Such thinkers thus hoped to expand the domain within which they could apply their version of a "Newtonian physiology" without worrying about presumably unpredictable influences from mind or soul. This left unresolved, however, the persistent conundrum of how such a nonmaterial entity like the human soul (as the Cartesian dualists envisioned it) could causally interact with a material body presumably devoid of any such psychic qualities. The animists on the other hand challenged the mechanists with their arguments that many forces of nature clearly display the qualities of mind or soul, which are particularly evident in the energies of living organisms but are also perhaps minimally present in ostensibly inanimate objects (Roe 2003, 400–8; Demarest et al. 2021). Such philosophical animism thus denotes the view that all things have at least some degree of inherent life or even inward experience.

The Eighteenth Century: Sensitivity and Irritability

The pioneering experiment of Leonardo, or a version of it, was repeated in the 1730s by Scottish physician Alexander Stuart, who provided a classic illustration (Figure 1), and by English cleric and physiologist Stephen Hales. These experiments influenced the thinking of Scottish physician Robert Whytt (pronounced "white"), perhaps the leading neurophysiologist of his time. The publication of his *Essay on the Vital and Other Involuntary Motions of Animals* in 1751 (with a second edition in 1763) marked the point at which the search for the soul became a public, philosophically nuanced, and occasionally acerbic debate

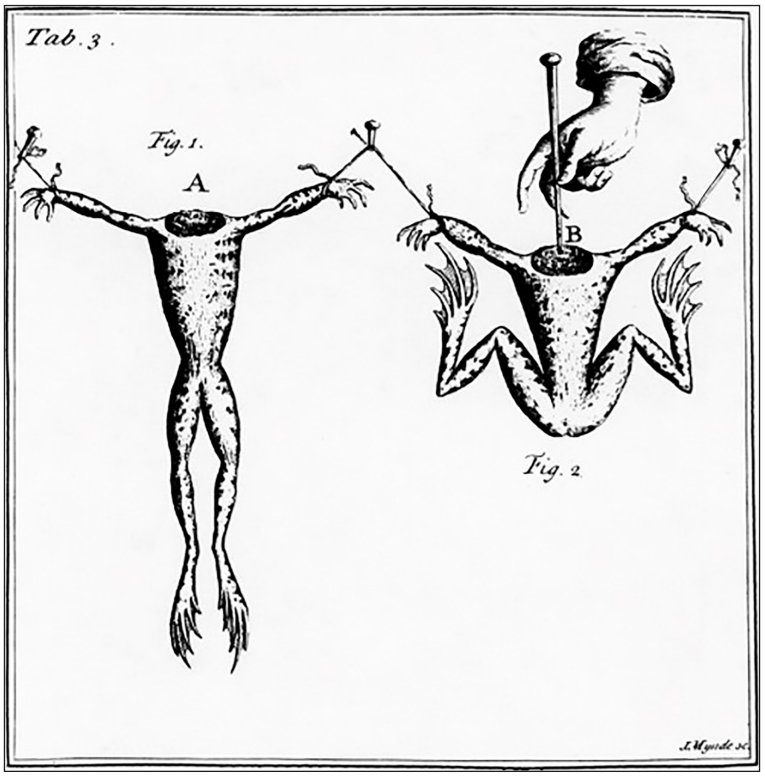


Figure 1. Stuart 1739, facing p. xxxvii.

embodying the modern confrontation between animism and mechanism. Whytt represented the animistic view, with the mechanistic perspective championed by renowned Swiss physiologist and physician Albrecht von Haller. Both scientists were devoutly religious in their dealings with the soul but distinctly at odds in their religious perspectives.

Whytt ([1763] 1768, 140ff) gathered from his and others' experiments with frogs, among other animals, that the brain and spinal cord were the primary seat of the soul or "sentient principle." He also held, however, that the sensitivity and responsiveness of muscles and organs even when detached from the brain meant that the soul is actually coextensive with the body, for "a frog lives, and moves its members, for half an hour after its head is cut off; nay, when the body of a frog is divided in two, both the anterior and posterior extremities preserve life and a power of motion for a considerable time" (Whytt [1763] 1768, 203). Since he was convinced that "matter, of itself, and unactuated by any higher principle" could never generate motility, he argued that the "motions and other signs of life which are observed in the body and limbs of a frog" deprived of its head or divided in two "are to be attributed to the sentient principle" still active in these isolated parts (Whytt [1763] 1768, 205).

Because of his seminal research on involuntary bodily movements (heartbeat, respiration, digestion, pupillary light response, etc.), processes in the domain of Aristotle's nutritive and sensitive functions, Whytt is known for his apparent contributions to the notion of reflex action. But Whytt (1751, 2) was far from

accepting the ascending Cartesian conception of the reflex and, accordingly, the image of an animal as, in Whytt's words, "a mere inanimate machine," a view that entails "a notion of the animal frame too low and absurd to be embraced by any but the most MINUTE philosophers!"

True to his animism, Whytt ([1763] 1768, 144) countered the notion of reflex action as purely automatic, and the general and growing mechanistic thinking of his time involving dualistic explanatory schemes, with his parsimonious proposal that "all the motions of animals, involuntary as well as voluntary, are some way owing to the mind." He rejected the common contemporary notion of two distinct psychic principles working in humans—the *anima* (the vital and sentient soul) and the *animus* (the rational soul). Instead, he argued for relative continuity, proposing that the *anima* and the *animus* are "one and the same principle acting in different capacities" (Whytt [1763] 1768, 148). He cited in evidence the fact that while many muscles in the body—such as the diaphragm in breathing and the eyelids in blinking—are generally subject to involuntary control far from consciousness (processes traditionally ascribed to the *anima*), they can often and quite easily be controlled voluntarily (under the auspices of the *animus*), and vice versa, such as when a sudden fright or a strong flash of light overrides efforts to consciously control these same muscles. The crucial distinction Whytt made between these different capacities is that while the soul acts with relative freedom in the case of conscious, voluntary movements (centered in the brain), when it comes to involuntary movements, it acts "necessarily, unconsciously and on the whole 'wisely'" (French 1969, 35) according to the "laws of union of body and soul"(French 1969, 149–60) and in service to the wellbeing of the organism. Whytt thus "replaced the mechanism of matter with a quasi-mechanical animism" (French 1969, 82), which gave full recognition to the rule of regular, lawful, reflex-like processes in the body, but argued that the soul was nevertheless present here, though lacking the liberty enjoyed in its higher expressions. He attributed both types of action to the "sentient and intelligent principle with which the Creator has animated our bodies" (Whytt [1763] 1768, 160).

Seeing such continuity between the *anima* and *animus* in human beings, Whytt also saw a "beautiful gradation" between ourselves and the larger animal kingdom, suggesting that "in brutes of the lowest kind there is evidently a sentient principle: but it seems to be wholly devoid of reason or intelligence: in those, however, of a higher class, we can perceive faint traces of something like what we call reason and reflection in man" (Whytt [1763] 1768, 171, 149–50). Thus again in contrast to Descartes, he proposed that the souls of animals and human beings differ in degree but not essentially in kind (French 1969, 118) and was therefore surprised that Descartes and his followers described animals as "so many curious pieces of clockwork" (Whytt [1763] 1768, 153). He was also surprised that some "theological writers," having accepted thoroughly

mechanistic explanations even of higher animals, "should not have been aware that the ascribing every action in man to no higher a principle, would be a natural and easy consequence" (Whytt [1763] 1768, 153), a conclusion drawn by many Cartesians.

Albrecht von Haller launched his main criticism of Whytt's physiology in his 1755 *Dissertation on the Sensibility and Irritability of the Parts of Animals*, four years after the Scotsman's signal publication. Whereas Whytt thought the sentient principle, implying at least some measure of feeling or sensitivity, was involved in all bodily movements, Haller made a sharp distinction between "sensibility," involving the soul, and "irritability," being in his view an inherent and merely mechanical capacity of the muscles to contract independently of soul. He saw the soul *in toto* as limited to the confines of the cranium and as initiating only voluntary bodily movements (Haller [1755] 1936, 658f; French 1969, 71). The many vital but involuntary and typically unconscious functions, on the other hand, are ruled by the mechanical irritability or contractility of the muscles themselves. Because Haller ([1755] 1936, 691) considered it a certainty "that the seat of the soul is in the head, and that it has no command over the rest of the body, after the nerves have been cut or destroyed; and farther, as the Irritability remains intire after the head is lopped off," he thus interpreted the muscular motions of decapitated frogs as evidence that such presumably involuntary movements are fully separate from soul. To argue otherwise, he thought, would be to "introduce an insensible sensation, and involuntary acts of the will, that is to say, to admit contradictory propositions" ([1755] 1936, 692).

Whytt thought it both unnecessary and unparsimonious to attribute fundamentally different modes of causation—soulful and mechanical, respectively—to voluntary and involuntary bodily activity. Since "nature never multiplies causes in vain," why should we, in explaining the movements of muscles, "have recourse to any hidden property of their fibers, peculiar activity of the nervous fluid, or other unknown cause, when they are so easily and naturally explained, from the power and agency of a known sentient PRINCIPLE" (Whytt [1763] 1768, 140). One of the experimental findings he marshaled in support of this view was the fact that for about fifteen minutes after decapitation, while the frog is apparently in psychic shock from the procedure, none of its muscles will respond to stimulation, whereas if the response were purely mechanical, this should surely occur as usual (Whytt [1763] 1768, 303–4). It seems that Haller was somewhat stumped on this point (French 1969, 74).

Haller's notions of sensibility and irritability perfectly illustrate the religious and philosophical motivations and implications of this debate. His restricting of the soul, with its conscious sensibility, to the brain, while attributing other bodily processes to mechanical irritability, was motivated at least in part by a theological commitment common to many Cartesians to keep the presumably immortal soul distinct from the mortal body (Boring 1950, 35–36). Whytt (quoted

in French 1969, 150) considered such dualism unnecessary and unwise, arguing that "mechanical connections are finite, and must at last terminate in a first cause," which he saw ultimately as God manifested in an embodied form as the pervasive soul acting via conscious or unconscious perception in all physiological processes. Haller countered this reverent regard for the extensiveness of the soul with the pious retort that "I have doubts about this spirit [Whytt's sentient principle] being the cause of motion. I derive all from God," (quoted in French 1969, 69), explaining that "we are inquiring into machines made by the hand of God" (Haller 1751, in Cunningham 2003, 73).

Whytt and Haller exchanged several further volleys in their writings over the ensuing decade, presenting refinements and elaborations of the basic issues summarized above. In the end, the debate between these two leading lights of eighteenth-century anatomy and physiology was essentially unresolved. Haller's approach, however, at least the mechanistic aspects of it, certainly found greater resonance with the advancing mechanical rumble of the contemporary Enlightenment zeitgeist and, much to Haller's dismay (though in line with Whytt's expectation), gave unwitting support to the rising tide of scientific materialism. French physician and philosopher Julian Offray de la Mettrie saw Haller's concept of irritability as a way of extirpating the notion of a purely spiritual soul from the domain of physiology, a key element of his project of eliminating religion from the conduct of life (Brooke 1991, 172–73; Riskin 2016, 158–61; Roe 2003, 404).

The Nineteenth Century: Frogs on Purpose

The nineteenth century introduced a new theme into this continuing controversy as the focus shifted somewhat from the details of physiological function and responsive movements to the closely related question of purposive behavior in pithed or headless frogs. In 1853, German physiologist Eduard Pflüger, whom Huxley ([1870] 2015, 183) called "a vehement advocate of Whytt's views," published the results of one of the most notable experiments in this regard, launching a challenge to the sequestering of the soul in the brain and the resulting mechanical conception of extracranial activity. He, like Whytt, believed that sentience or consciousness is coextensive with the nervous system throughout the body and that its identifying mark is movement, particularly purposive movement (Fearing [1930] 1970, 162–63). So, he designed an ingenious (if brutal) experiment to demonstrate just such behavior (see Klein 2018, 894–96). As illustrated in Figure 2a, Pflüger would suspend a decapitated frog (with the spinal cord intact) and irritate its right flank with a bit of acetic acid. The frog would then bend its right leg up in an effort to reduce the irritation by wiping off the acid. Next Pflüger would wash off the acid and sever the right leg, as shown in Figure 2b. Then, after acid is reapplied to the same spot on the right flank (not shown in this 1907 illustration), the frog first moves the

stump of its right leg in a vain attempt to wipe the acid. But then a different solution appears—it bends its *left* leg up and reaches across its body to remove the irritation.

A contemporary and sympathetic commentator, English philosopher and physiologist George Henry Lewes, who had repeated and modified Pflüger's experiment many times, saw a human analogy here, suggesting that "[t]his is very like the action of the tickled child, who always uses the right hand to rub the right cheek . . . But when the child's right hand is prevented from rubbing, the left will be employed" (Lewes 1860, 246). He was also convinced that these experiments offered evidence that something beyond a simple reflex or "unconscious mechanism" was at work here, not only in this alternative action of the frog but also in the fact that "the frog does not *always* hit even on this plan" (1860, 247). Indeed, after a series of experiments in which Lewes irritated the frog's thigh instead of its flank, he reported that the frog "sometimes bends its irritated leg more energetically, and likewise bends the body towards it, so as to permit the spot to be rubbed against the flank—just as the child, when both his hands are held, will bend his cheek towards his shoulder and rub it there" (1860, 247). Lewes concluded from these examples of novel responses or spontaneity that "it is difficult to resist such evidence of *choice* as is here

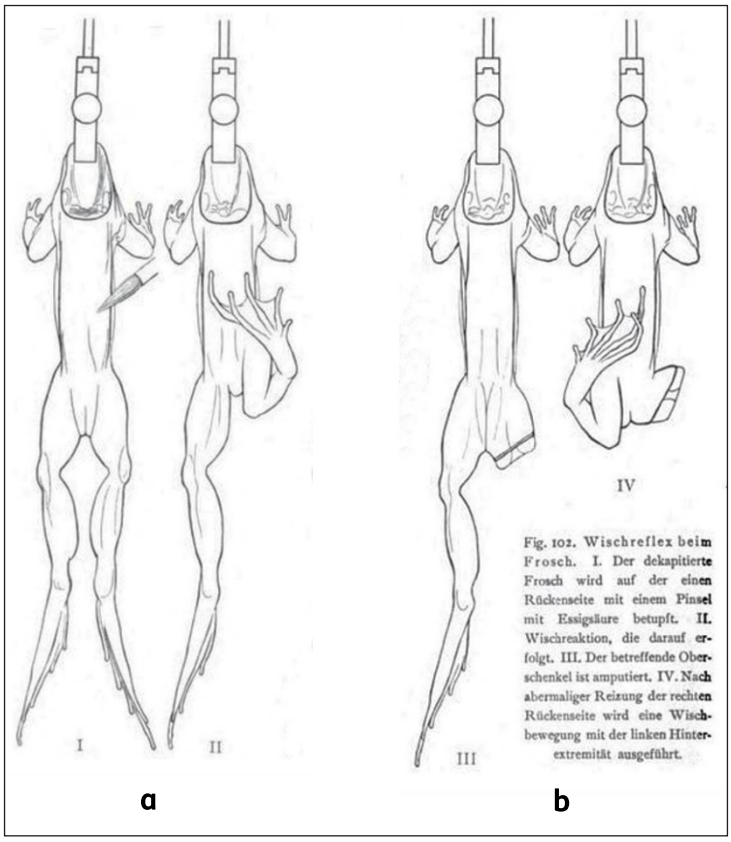


Figure 2 (a and b). One of Eduard Pflüger's decapitated frog experiments (from Verworn 1907, 198).

manifested. The brainless frog chooses a new plan when the old one fails, just as the waking child chooses" (1860, 246–47).

Some other thinkers, however, had little difficulty resisting such a conclusion. German philosopher and psychologist Hermann Lotze suggested that the apparently intelligent movements of the frog's legs in Pflüger's experiment were due to the previous life experiences of the intact animal when "under the influence of the soul life an association has once been formed" and repeatedly reinforced between a stimulus and the appropriate response, a pairing that becomes inscribed in the spinal cord and can then presumably function mechanically as an unconscious reflex (Lotze 1853, in Fearing [1930] 1970, 164). But while this explanation might reasonably be considered in the first case of the right leg rising to wipe off the acid on the same side, it is more difficult to think that a frog, in the course of its former life, would ever have encountered and risen to the challenge of the second case, of reaching for the acid with its opposite leg, and surely not often enough for this response to become stamped into the spinal marrow. Huxley ([1870] 2015, 179) emphasized this point with further evidence several years later.

Some of the most interesting and original laboratory contributions following Pflüger were those of German physiologist and prudent experimentalist Friedrich Goltz, who agreed with Whytt and Pflüger that the soul—or more specifically sensation, and with it the capacity to execute adaptive responses—is not limited to the brain. He found, for example, that if a brainless frog is released deep in a container of water, it will swim to the surface to breathe. It is also capable of selecting a female and rejecting a male for a mating embrace (Fearing [1930] 1970, 166). And further and more surprisingly, he discovered (see Figure 3) that if an inverted jar filled with water is placed in the frog's path to the surface, it will swim to the top of the jar and, finding no breathing space there, will swim back down out of the jar to gain the surface in the normal way after this unusual detour (Goltz 1869, 70). As did Pflüger before him, Goltz argued that it is difficult to see how such behavior could be the product of simple reflexes without some measure of sensation and consciousness.

Huxley lent his distinctive voice to the debate first with his 1870 address before the Metaphysical Society of London, entertaining the question "Has a Frog a Soul, and of What Nature is that Soul, Supposing It to Exist?" (Huxley [1870] 2015), and then with his more substantial 1874 essay "On the Hypothesis That Animals Are Automata, and Its History" (Huxley 1874). In both these works, Huxley recounts the contributions of some other leading participants in the controversy—including Whytt, Haller, Pflüger, and Goltz—and notes that he had personally confirmed some of their results on the surprising behaviors of decapitated frogs, even accepting Pflüger's finding of purposive behavior in headless frogs. But he concludes even so that "I am unable to see in what

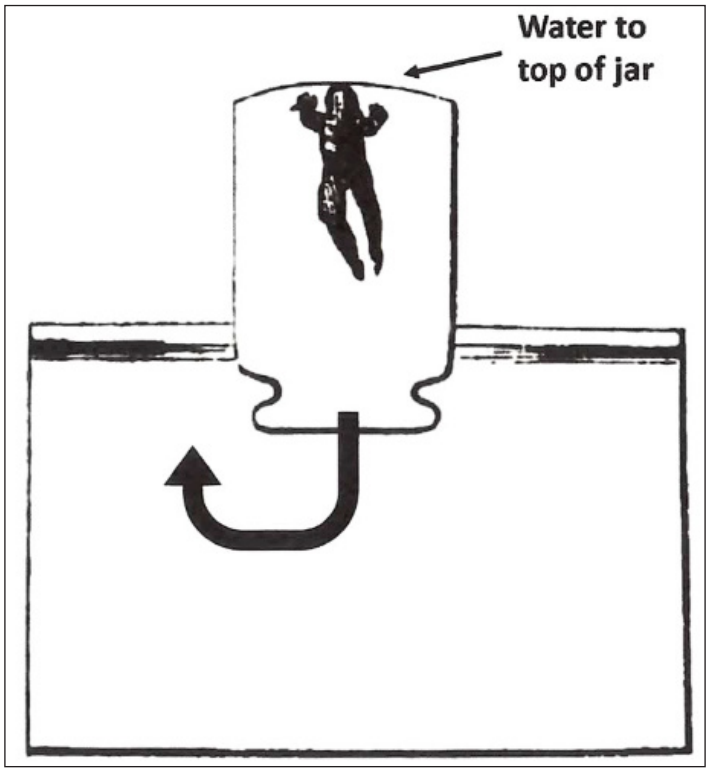


Figure 3. After Goltz 1869, 70.

respect the soul of the frog differs from matter" (Huxley [1870] 2015, 183, 184). He also considered it obvious "that had Descartes been acquainted with these remarkable results of modern research, they would have furnished him with far more powerful arguments than he possessed in favor of his view of the automatism of brutes" (Huxley 1874, 568), a view noted before in Haller with his "irritability," i.e., that the complexity of presumably unconscious behavior proves that mind or consciousness is not required for it.

Although Huxley says that the "frog walks, hops, swims, and goes through his gymnastic performances quite well without consciousness, and consequently without volition, as with it," he nevertheless accepts the possibility that they might indeed have souls:

> If they possess immaterial subjects of consciousness, or souls, then, as consciousness is brought into existence only as the consequence of molecular motion of the brain, it follows that it is an indirect product of material changes. (Huxley 1874, 576)

Hence:

> The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be as completely without any power of modifying that working as the steam-whistle which

> accompanies the work of a locomotive engine is without influence upon its machinery (Huxley 1874, 575).

Thus was born "epiphenomenalism"—the theory that consciousness has no causal efficacy over the body—and Huxley's 1874 essay is probably the most seminal and influential statement of this position. What makes this view, also dubbed the "helpless spectator theory" (Jaynes 1976, 11–12), so significant and Huxley's exposition of it so innovative and important is that he takes the bold step of invoking the venerable frog-human analogy even in this case, concluding that "if these positions are well based, it follows that . . . We are conscious automata" (Huxley 1874, 577). In other words, both we and frogs might indeed have souls in some sense, but such a phenomenon makes little or no practical difference in our embodied lives; it is merely "epi"—a shadowy byproduct hovering over or outside the solely effectual neural machinery.

Based on these and other examples, Huxley followed Descartes but went beyond him by extending the philosophy of mechanism to all aspects of human behavior. In the end, however, he perhaps managed, at least in his own mind, to steer clear of outright charges of heresy by equivocating on the question of whether a frog or a human being *actually has* a soul, saying merely that modern empirical science could get along quite well without it. He was thus true to form as the leading agnostic of the age, a term he coined the year before his Metaphysical Society address.

Pioneering American psychologist and philosopher William James offered the final authoritative statement on the phenomenon of headless frogs in the waning years of the nineteenth century. He famously introduced psychology as "the science of mental life" in the first line of his *Principles of Psychology* in 1890. And while he strove to remain strictly non-metaphysical in this work, he saw the limits of purely mechanistic theories and maintained an open mind for possible spiritual realities, though emphasizing that "if there be such entities as Souls in the universe, they may possibly be affected by the manifold occurrences that go on in the nervous centres" (James 1890, 181).

It is not surprising, then, that in his efforts to bring the mind within the sphere of natural science, James's first forays into zoology and nerve physiology involved the recent research on frogs. He recounted the investigations of Pflüger, Goltz, and others mentioned earlier because, with their evidence of purposefulness or final causation, he found support for a principle he adopted as a guiding premise for his project—that "*the pursuance of future ends and the choice of means for their attainment are thus the mark and criterion of the presence of mentality* in a phenomenon. We all use this test to discriminate between an intelligent and a mechanical performance" (James 1890, 8). He also suggests that this test figures decisively into "the deepest of all philosophic problems," for "if we find ourselves, in contemplating [the Kosmos], unable to banish the impression

that it is a realm of final purposes, that it exists for the sake of something, we place intelligence at the heart of it and have a religion. If, on the contrary, in surveying its irremediable flux, we can think of the present only as so much mere mechanical sprouting from the past, occurring with no reference to the future, we are atheists and materialists" (James 1890, 8). Although he did not pursue these considerations further in *this* book, his guiding metaphysical sensibilities are, of course, fully evident in the *Principles of Psychology*.

James dealt more directly with Huxley's conscious automaton theory in an 1879 essay titled "Are We Automata?" but also and more fully in a chapter in the *Principles*. There, he describes the difficulty in deciding between the competing interpretations. From Huxley's perspective, "we start from the frog's spinal cord and reason by continuity, saying, as that acts so intelligently, *though unconscious*, so the higher centres, *though conscious*, may have the intelligence they show quite as mechanically based," even in the case of humans (James 1890, 134). But the exactly opposite counterargument "urged by such writers as Pflüger and Lewes" is also based on continuity but starts with the performances of the hemispheres and says that "as *these* owe *their* intelligence to the consciousness which we know to be there, so the intelligence of the spinal cord's acts must really be due to the invisible presence of a consciousness lower in degree (James 1890, 134). And since arguments based on continuity can be used either to level up or level down, he notes that "such arguments as these can eat each other up to all eternity" (James 1890, 134). He does, nevertheless, argue for the usefulness and thus the causal efficacy of consciousness based on a variety of evidence, and in the end concludes that "the automaton-theory must succumb to the theory of common sense" (James 1890, 144).

The mechanical theory as a whole, however, did not succumb to the simple logic and evidence of common sense. It seems, indeed, that mechanistic physiology as well as psychology had at least partly won the day and eagerly envisioned great new triumphs on the horizon at the dawn of the twentieth century (Coleman 1977; Boring 1950). As has been the case from the eighteenth century to the twenty-first, the practical and programmatic value of mechanistic materialism to the further progress of the scientific enterprise is so great that its undeniable benefits have been widely welcomed while its philosophical limitations have been downplayed. The primary shortcoming is the inability to explain how two such ontologically distinct entities as a nonmaterial mind or soul and a merely material body could possibly interact, which is what drove Huxley to resort to epiphenomenalism and has delivered a seemingly intractable challenge to modern philosophy since the seventeenth century. It is thus the legacy of materialism at the base of the modern outlook that has made the mind-body problem such a knotty issue and the problem of consciousness so very hard. Another aspect of this legacy is its contribution to the contracting and "loss of soul" in the modern Western psyche.

Part Two: A Postmodern Search for Soul—Updating the Classic Debate

In the 1920s, Alfred North Whitehead, building on William James's quite radical 1904 (James 1904) proposal that consciousness and the physical world are derived from the same primal stuff—namely, "pure experience"—began formulating his distinctive postmodern alternative to the central conundrums of modern science and philosophy arising from the underlying mechanical philosophy. His "philosophy of organism" (probably better known as "process philosophy") is clearly in the animist tradition and was informed, in part, by his deep knowledge of early twentieth-century physics. This expertise led him to broaden the notion of "organism" because "the atom is transforming itself into an organism," and therefore, "science is taking on a new aspect which is neither purely physical, nor purely biological. It is becoming the study of organisms. Biology is the study of the larger organisms; whereas physics is the study of the smaller organisms" (Whitehead 1925, 103).

By using the life of organisms rather than the operation of machines as his epistemological starting point, Whitehead created a rigorous and thoroughgoing panpsychism that challenged the twin Cartesian tenets of a natural world consisting fundamentally of mere matter and of an independent and ontologically distinct human mind or soul. In contrast, Whitehead saw various degrees of psyche from top to bottom of the natural world, thus accommodating the reality of soul in an expanded naturalistic framework instead of considering it a perhaps *super*natural endowment. He suggested, therefore, that "it is not a question of having a soul or not having a soul. The question is, How much, if any?" (Whitehead 1933, 208). Charles Hartshorne (1950, 442), innovative cofounder with Whitehead of this school of thought, explained accordingly that "'souls' may be very humble sorts of entities—for example, the soul of a frog—and panpsychists usually suppose that multitudes of units of nature are on a much lower level of psychic life even than that." Because of its emphasis on levels or gradations of sentient experience rather than on well-developed psyches, this approach is more accurately termed "panexperientialism" (Griffin 1998, 78).

Process philosophy is thus postmodern in that it challenges—and provides positive alternatives to—basic scientific and philosophical ideas that have informed the modern world, and various thinkers, beginning in the 1940s, have designated it as such (see Griffin 2007, 3–5). To distinguish it from more recent deconstructive postmodernism, David Ray Griffin (2007) calls it *constructive* postmodern thought. Because Whitehead unearthed and productively analyzed foundational modern concepts, his approach could perhaps also be tagged as "nonmodern," although this newer term lacks the sense of historical context that Whitehead was quite conscious of. Interestingly, one of the leading advocates of this term, Bruno Latour (1993), though very critical of the unexamined *modern* assumptions of deconstructive postmodernism, eventually came to embrace and endorse Whitehead's philosophy (Latour 2011).

Regarding process philosophers' fidelity to empirical data, it is notable that Hartshorne, an accomplished ornithologist, included numerous references to the vocalizations of frogs in his book on bird song (Hartshorne 1973), proposing that, even though they lack the vocal virtuosity of birds, the Anurans nevertheless display in their rhythmic utterances some of the most primitive elements of music. He also emphasized that such sound production is typically controlled by voluntary muscles (Hartshorne 1973, 14), which implies some degree of agency and perhaps intelligence in this behavior—qualities, in addition to the phenomenon of voice itself, long associated with the presence of soul (see Riskin 2016).

An important implication of the "process" in process philosophy, aside from the general consideration that organisms on any level are always essentially processes or events, is that, more specifically, the soul or mind is not a timeless entity or unchanging "thing" but a process, an ongoing and sometimes evolving process of experiencing. So, in the philosophy of organism, the traditional notions of "the soul" and "the mind" are replaced by the idea of a succession of moments or "occasions of experience," Whitehead's more technical term being "actual occasions" (Whitehead [1929] 1978, 141). This postmodern soul avoids the traditional perplexity of the mind-body problem by being intimately and internally related to body.

Whitehead's approach was significantly influenced not only by the revolutionary insights of relativity and quantum theory but also by contemporary developments in biology, the field for which his work, in turn, probably has the greatest relevance (Henderson 1926; Needham 1951, 268, 271), especially for zoology and ethology (Agar 1951; Thorpe 1978), embryology (Waddington 1975), evolutionary biology (Wright 1964; Birch 1977), and theoretical biology (Birch and Cobb 1981; Sölch 2020). The most important biological contribution for Whitehead's concept of the soul was the cell theory introduced by botanist Matthias Schleiden and zoologist Theodor Schwann in 1839. For Whitehead (1925, 100), the pivotal significance of this theory, along with Louis Pasteur's microscopic studies of bacteria, is that it "introduced the notion of *organism* into the world of minute beings." Hartshorne ([1936] 1972, 54) went so far as to say that Whitehead's "philosophy of organism" presented "the first full-blooded, forthright interpretation of the cellular model" in its far-reaching metaphysical dimensions. Whitehead ([1929] 1978, 108) thus emphasized the animistic sentiment that "all the life in the body is the life of the individual cells. There are thus millions upon millions of centres of life in each animal body," or as is now known, about 36 trillion in an adult male human body (see Hatton et al. 2023). Hartshorne stressed further that "the cell cannot easily be regarded, even for a moment, as anything less than a concretely existent individual entity," ([1936] 1972, 41) these entities being, in other words, "low-grade fellow animals" (1934, 112).

Cells, regarded from an outward, physical perspective, organize into familiar tissues, organs, and systems. But from the inward, mental, or psychic point of view suggested by panexperientialism, these fellow animals form societies on many levels, in which smaller and simpler societies are integrated into larger and more complex societies. An important question, however, is whether in the society of "occasions of experience" within any body there is evidence of a *dominant* or *presiding* occasion present among the cells. Here, process philosophers make a crucial distinction among different types of living, and nonliving, things, describing these societies of occasions as either democracies or monarchies, with the quality of the relations among the constituent organisms in these societies, whether atoms or cells, making all the difference.

Whitehead characterized things such as rocks and clocks and tables and chairs as "corpuscular societies" ([1929] 1978, 35) or "democracies" (1933, 206; 1938, 24), groups of individuals that, so to speak, barely know each other, being bound together merely by their coexistence in space and their relatively monotonous continuity through time. Hartshorne ([1936] 1972, 57–59) called them "composite individuals" and Griffin (1998, 186) "aggregational societies." In contrast to these societies, Whitehead describes living entities such as frogs and human beings not as democracies but, as Hartshorne came to designate them, "monarchies" (1953, 38) with "compound individuals" at their head ([1936] 1972, 41–61). In other words, monarchies are societies of sympathetically related individuals in which, "by reason of this organization, an adjusted variety of feelings is produced in that supreme entity which is the one animal considered as one experiencing subject" with some measure of control over its constituent bodily subjects (Whitehead 1938, 23).[1] The essential idea is that the mutual sensitivity of the cellular individuals or occasions of experience give rise, most notably in those organisms with nervous systems, to a unified and "presiding occasion" capable of acting with some measure of purpose and freedom. Thus "in man, the living body is permeated by living societies of low-grade occasions so far as mentality is concerned. But the whole is coordinated so as to support a personal living society of high-grade occasions. This personal society is the man defined as a person. It is the soul of which Plato spoke" (Whitehead 1933, 208).[2] In a similar vein, Charles Sanders Peirce (1960, 182), another seminal postmodern philosopher and colleague of William James, stated around 1890 that "consciousness is a sort of community spirit among the nerve cells." We could perhaps also say, adapting the well-known African proverb to express this psychophysiological relationship, that it takes a village (of sentient cells) to raise a soul.

Whitehead ([1929] 1978, 21) saw evidence of a universal creativity in this emergence of compound individuals, for "the ultimate metaphysical principle is the advance from disjunction to conjunction, creating a novel entity other than the entities given in disjunction . . . The many become one, and are increased by one." The general principle here is that only monarchies, and not democracies, exhibit

marks of mind or soul. This important distinction means that charges of outright absurdity sometimes leveled against panpsychism for supposedly claiming that stones and telephones have souls do not apply to Whiteheadian panexperientialism.

Assessing the Evidence in the Postmodern Search for Psyche

As discussed, Whitehead emphasized that the dominant occasion of an organism may be more or less and the task of determining "how much, if any" governing psyche is possessed by a particular grade of organism rests on an empirical analysis of the observable signs of soul. Griffin identified the two basic marks of mind in process thought as "experience" and "spontaneity" (1998, 7, 78–79, 186), which are obviously similar to the ones employed in the classic debate. "Experience," embodied in "occasions of experience," is usually described by Whitehead in terms of "feeling," which is a general term for "any kind of acting or being acted upon, in such a way that the make-up of the subject is affected" (Emmet 1966, 142), emphasizing "both that *something* is felt and that it is felt *with affective tone*" (Griffin 1998, 128). But the only way we can know from the outside that an occasion of experience has occurred is that a correlative physical response takes place to whatever the stimulus might have been. Whitehead described the inward psychic process that occurs between a stimulus and a chosen response as "the intervening touch of mentality" (see Miller 2021).

The other basic sign of soul, "spontaneity," is the mark Whitehead (1933, 51, 258) saw as "of the essence of soul," defining it as "originality of decision." There are, however, many degrees of originality or spontaneity manifested across the vast range of organisms from amoebas to Einsteins. The merest hint of it can be seen in the various dynamic but reliably and rhythmically repetitive oscillations of electrons between positive and negative poles in basic electrical and magnetic processes, what Whitehead (1911, 103) called "the essential periodicity of things" that underlies the entire natural world. A bit more can be discerned when relatively higher organisms, maybe microbes and beyond, display some self-determination in their efforts to sustain and perhaps enhance their wellbeing. And probably the highest form is expressed in the exercise of human freedom when thought or action achieves significant independence from the constraining pull of the past (Griffin 1998, 163ff). If panexperientialism is correct in seeing a measure of mind throughout the natural world, some signs of experience and spontaneity should be present even in the individual cells that form the diverse society from which a presiding occasion in the body of a frog or a human being draws its life.

Mind in Microbes

One of the earliest significant scientific publications on this question is *The Psychic Life of Microorganisms*, published in 1888 by French psychologist Alfred Binet (1888, iii, 1), in which he presented experimental evidence that "psychological

phenomena . . . are met with in every form of life from the simplest cellule to the most complicated organism." Early in the twentieth century, American zoologist Herbert Spencer Jennings (1906, 336) stated that after many years studying single-celled organisms, he was thoroughly convinced that "if Amoeba were a large animal . . . its behavior would at once call forth the attribution to it of states of pleasure and pain, of hunger, desire, and the like, on precisely the same basis as we attribute these things to the dog." Around mid-century, Anglo-Australian zoologist W. E. Agar's *A Contribution to the Theory of the Living Organism* combined biological research on diverse animals, including amoebas and amphibians, with the philosophy of Whitehead and Hartshorne and argued that "all living organisms are feeling, experiencing, subjects" (Agar 1951, 1). Among other notable studies in the later twentieth century, philosopher Karl Popper and Nobel laureate neurophysiologist John Eccles affirmed Jennings's work and commented that an amoeba's sensitive interaction with its environment shows that in this unicellular entity there is a center "of curiosity, of exploration, of planning; there is an explorer, the animal's mind" (Popper and Eccles 1977, 30).

The early twentieth-first century has proven to be a fertile period for research into the psychic qualities of single cells, with too many studies to mention here (see, for example, Ben-Jacob et al. 2011; Krumbein and Asikainen 2011; Reid and Latty 2016). In one particularly interesting 2019 project, a research team at Harvard Medical School replicated a crucial 1906 experiment of Jennings, mainly to challenge some apparently illegitimate debunking of it from the 1960s. The Harvard team confirmed Jennings's findings that these cells display complex decision-making that "reveals unexpected depths in the cognitive capabilities of singly nucleated cells" (Dexter et al. 2019, 4327).[3]

Finally, in some ingenious recent experimentation with the cells of frogs, researchers in 2021 removed clumps of skin cells from embryos of the African clawed frog and placed them in petri dishes with a mild saline solution. Within thirty minutes, each group of cells had formed into a unified spherical organism, obviously without a nervous system, and these new organisms demonstrated, among other qualities, rapid self-healing after damage and registration of their experiences of colored light by changes in their own fluorescence. But the most remarkable finding was that after four days, the spheres became mobile by means of cilia on their surfaces, tiny hairs that, when such cells remain hidebound in the frog, work only to move mucus and thus eliminate foreign material, *but are never used for locomotion*. With their newfound freedom, however, this novel behavior emerged, and the spheres swam actively about their environment, navigating mazes and collectively removing particles and creating piles of debris (Blackiston et al. 2021). It is almost as if the "identity" of the cells had shifted from one of fulfilling their native role in the maintenance of the frog's crucially permeable skin to one of exploring and maintaining their larger environment.[4] This idea that the encompassing context affects

a microorganism's identity and behavior is indeed developed by one of the researchers (Levin 2019, 2021). It was embraced as well by Whitehead (1925, 79) a century earlier as a basic feature of his philosophy of organism, saying that "the plan of the *whole* influences the very characters of the various subordinate organisms which enter into it," so that on some occasions, even very simple organisms consisting of only one or a few cells can display the spontaneity and freedom that are reliable marks of mind.

A Postmodern Soul of the Frog

What then of a frog as a whole, or at least a relative whole? How would the experimental evidence provided by the classic debate about frogs be interpreted from the perspective of process philosophy? Although process philosophers have not dealt specifically with the historic debate regarding the souls of frogs, relevant connections between the two streams of thought are not difficult to find. In reviewing this debate here, I freshly interpret evidence of the basic signs of soul—experience and spontaneity—while also bearing in mind Whitehead's emphasis on "how much, if any." Some preliminary facts to keep in mind are that scientists for centuries have productively pursued physiological analogies from frogs to humans and that recent research has confirmed that the Anuran and human nervous systems and body plans are homologous in significant ways (Ewert and Arbib 1989, xi; Handrigan and Wassersug 2007). So, if one accepts the notion that our network of bodily cells contributes substantively to our conscious experience, it bears remembering that frogs have a similar, though much simpler, bodily society of cells.

The phenomenon of headless frogs was of course a centerpiece of the long controversy regarding the soul of the frog, with mechanists such as Haller and La Mettrie assuming that losing the head entails losing the soul, while the animists like Whytt, Pflüger, Goltz, and Lewes thought this meant losing only the *primary* but not the more *pervasive* seat of the soul. Pertinent to this issue of decapitation, Whitehead (1938, 24), with democracies and monarchies in mind, notes that usually, at least in highly complex organisms, "if the dominant activity be severed from the rest of the body, the whole coordination collapses and the animal dies." More democratic societies, however, which he saw exemplified in plants, can be divided and subsequently flourish. But there is not a clear-cut distinction between plants and animals, because "some traces of dominance can be observed in vegetables, and some traces of democratic independence can be found in animals. For example, portions of an animal body preserve their living activities when severed from the main body. But there is failure in variety of energy and in survival power" (Whitehead 1938, 24). Such independence is well displayed, for example, in the renowned ability of the isolated leg muscles of frogs to energetically contract upon stimulation for hours following separation from the body (see Sleigh 2012a, 102–7). Whytt, as noted, also reported that a

frog can live and move its legs for half an hour after losing its head. Further, regarding the failure of survival power in such circumstances, Pflüger and Lewes note that a headless frog, in contrast to an intact frog, makes no attempt to escape from a gradually heated tank of water (Lewes 1873).[5]

A Whiteheadian organismic perspective on the classic eighteenth-century experiments regarding sensibility and irritability would find some significant similarities, to a certain degree, with the animist interpretation of Whytt. The "sentient principle" that Whytt saw as seated primarily in the brain and spinal cord but also as extended secondarily throughout the body and active in both voluntary and involuntary bodily motions accords quite well with Whitehead's emphasis, following the cell theory, on "experience" or "feeling" in the myriad cellular centers of bodily life that are integrated through the nervous system and brain into a compound individual, a presiding mind or soul. His approach would, of course, be quite at odds with Haller's limitation of "sensibility" to the brain alone, as well as his proposal that merely mechanical "irritability" reigns in the rest of the body.

There are also shades of Whytt in Whitehead's methodological commitment to the principle of parsimony. Whytt ([1763] 1768, 2), inspired by Isaac Newton and observing that nature apparently "delights in simplicity and uniformity," endeavored to show that all bodily movements, voluntary as well as involuntary, are explicable by a single principle, that they are in some way "owing to the mind." He therefore placed the traditional *anima* (sensitive functions) and the *animus* (mental functions) on a continuum. Similarly, Whitehead proposed his notion of "occasions of experience" or "actual occasions," which oscillate between physical and mental poles, as the single fundamental form of causal processes or events active throughout the world. Both recognized, however, that these basic principles or processes operate somewhat differently in different bodily contexts. Whytt therefore proposed that in conscious, voluntary movements, the soul or sentient principle enjoys relative freedom, while in involuntary and largely unconscious movements, it acts rather mechanically and reflex-like, thus making for Whytt's "quasi-mechanical animism." Whitehead, in turn, suggested that the degree of spontaneity and freedom increases with the growing complexity throughout the hierarchy of organisms, but that "an individual entity, whose own life-history is a part within the life-history of some larger, deeper, more complete pattern, is liable to have aspects of that larger pattern dominating its own being" in relatively lawful ways. This is the essence of his thoroughgoing "theory of organic mechanism" (Whitehead 1925, 80, 106–7).

There is a fundamental issue, however, on which Whytt's eighteenth-century animism is quite distinct from Whitehead's twentieth-century philosophy of organism. Whytt, as well as Haller, inherited Newton's notion of physical (nonliving) matter as thoroughly passive or inert. Whytt therefore argued that matter, due to its inherent inactivity, needed to be actuated by a "higher

principle," a sentient principle, to render it capable of sensation, perception, pleasure and pain, or motility. This Newtonian concept of inert matter was, however, the principal target that Whitehead ([1929] 1978, xiii), in light of relativity and quantum theory, took aim at in his rejection of the belief in "vacuous actuality"—that mere, nonexperiential matter actually exists—and that his panexperientialism was designed to replace.[6] Whytt was thus saddled with the challenge of mind-body dualism, while Whitehead was able to unsnarl the stubborn world knot with his postmodern metaphysics.

The nineteenth-century experiments of Pflüger and Goltz that raised the debate over the soul of the frog from the presence of sentient experience or feeling to the capacity for purposive action find ready interpretation in terms of the Whiteheadian concept of spontaneity, specifically in the form of decision-making and self-determination. The case of the headless frogs that used their far leg to wipe away acid after their near leg was amputated, as well as the frogs that, when their route to the surface of a water tank was blocked by an inverted jar, swam back down out of the jar and then to the surface, are obvious examples.

An issue that came to a head at this time in the progress of this debate concerns the specific place and precision of the pithing of these spinal frogs (see Klein 2018, 898–900). In trying to pin down the neural correlates of purposive action, experimenters found, as Huxley reviewed in his essay, that severing the brain and spinal cord at slightly different locations along the brain stem generally corresponded to the loss of specific behaviors such as jumping, swimming, or turning right side up from its back, so that the lower the cut the greater the loss. From the perspective of process philosophy, this finding would be seen as an indication that reducing the number of especially critical neural cells in this integrated society of actual occasions results in a diminished and less effective soul. It would involve, so to speak, a strategic dismantling of the reigning monarchy.

So, what of Huxley's epiphenomenalism? Although he acknowledged, after considering a range of evidence, that a frog, and of course a human being, might indeed have a soul, he argued that such an entity is physically ineffectual. He was driven to this conclusion because of the seeming impossibility of understanding how something purely immaterial could interact with a merely material body, which is, of course, the apparent mystery at the bottom of the mind-body problem. It is curious here that Huxley accepts that the nervous system somehow generates or causes the subjective experiences of the mind or soul but rejects the idea of causality in the opposite direction, from mind to body. Process philosophers offer a fairly radical resolution to this logical glitch grounded in their concept of both body and soul as societies of dynamic individual centers of life and experience. First, in accord with James's suggestion that "the automaton-theory must succumb to the theory of common sense," Whitehead (1938, 155–56) argues that we are, in fact, "*directly* conscious of our

purposes as *directive* of our actions," which is an element of what Griffin (2007, 52–53) calls "hard core common sense," the type of knowledge that everyone inevitably assumes in practice even though some occasionally deny it in theory. It is an assumption that is obviously part of the bedrock of ethical and legal systems, for in an imaginary epiphenomenal legal order, "to indict the thief for stealing [would be] analogous to indicting the sun for rising" (Whitehead 1938, 155–56).

To account for the operation of such conscious influence, Whitehead (1938, 24) reminds us that the "bodily organization is such that the unity of feeling, which is the one animal as a sentient being, receives its complex variety of experience from these bodily activities." And Hartshorne (1962, 229), addressing both directions of causality between body and soul, explains that "cells can influence our human experiences because they have feelings that we can feel. To deal with the influences of human experiences upon cells, one turns this around. *We* have feelings that *cells* can feel." He concludes that this theory of mental causation "in principle solves the mind-body problem . . . The rest is detail" (Hartshorne 1962, 229). With this resolution, these thinkers offer a reasonable foundation for understanding the gradations of spontaneity, self-determination, and freedom often displayed by simpler organisms even in extreme states and by more complex organisms and humans to a much greater degree in the daily conduct of their lives. So yes, Professor Huxley, it seems reasonable to conclude that not only does a frog indeed have a soul, in some measure, but also that this panpsychic theory of mental causation aligns quite well with the venerable theory of common sense.

Conclusion

The panexperientialism at the heart of the philosophy of organism obviously entails a significant expansion of the notion of soul from that entertained by most of modern science and philosophy. It thus represents a path toward reversing the contraction and loss of soul so prevalent in the modern world. It can also perhaps support the human search for a psychic center with sufficient depth and breadth to productively integrate the diverse and often disparate elements of ourselves—mind and matter, consciousness and the unconscious, head and heart, intention and action—that Jung and the other thinkers noted earlier had in mind. The extended debate over the soul of the frog illustrates these trends in microcosm, holding implications for the human body and soul at every turn. The generally accepted basis for frog-to-human analogies, as emphasized throughout the history of modern science is the set of significant homologies or evolutionary convergences in body, brain, and behavior between these two very distant phylogenetic relatives. These considerations, in light of the cellular model, suggest the prospect of re-ensouling or reenchanting not only human beings but also nonhuman beings on many levels.

The expansion of inward experience and soul also means the extension of value or worth to the teeming multitudes of individuals or "actual occasions" on all these levels—intrinsic value, or "the sense of existence for its own sake, of existence which is its own justification" (Whitehead 1938, 109). While this sweeping view of value might seem to suggest a simplistic and crippling egalitarianism ranging from microbes to humans, Daniel Dombrowski (2021, 195) argues that Whitehead recognized both "the continuity of value in nature that goes all the way down" and "the existence of certain thresholds that are crossed that lead to qualitative changes," analogous to the changes in water that is heated to boiling or cooled to freezing. The most significant such threshold was crossed with the (gradual) appearance of central nervous systems, which afforded richer inward experience or sentiency, both pleasurable and painful, or, in other words, a fuller sense of soul. Below this threshold lies the "microscopic sentiency" of individuals that are worthy of our concern and contribute to our own lives but do not rise to the level of possessing actual moral *rights* (see Dombrowski 2021, 177–97).

Frogs live well above this threshold, which raises the important question of whether the philosophy of organism would countenance or condemn the experiments with frogs recounted earlier. It is beyond the scope and space of this article to fully explore this question or the related one of ethics in the dissection of frogs in biological education, but I would merely mention, as Henning (2023) has emphasized, that Whitehead (1938, 111) situates ethical decisions in a broad context of value, noting that "everything has some value for itself, for others, and for the whole." The multiple concerns of this "triadic theory of value" (Henning 2023, 42f) suggest that its adherents would take quite seriously the suffering of the frogs in these experiments. But they would also consider the clarity such research has perhaps provided regarding mind, consciousness, or soul and its association with matter throughout the natural world, especially in the human domain. No easily digestible answers here, but much food for thought.

Many, though certainly not all, process philosophers, in the spirit of William James reasoning from purposiveness in frogs and other creatures to the prospect of religion, have extended the argument from analogy between nonhuman and human beings far beyond earthbound organisms. This thinking proceeds on several levels, recognizing first that the humble soul of a frog can exert some global influence over the society of cells in the archetypal amphibian body. Then, highlighting correspondences at the human level, it accepts our common daily experience that "the soul of which Plato spoke," while at least vaguely feeling the pains and passions of the vast multitude of cells in the human body is also somewhat of a monarch over them. Further extending and enlarging these considerations, these thinkers ultimately suggest that the larger soul of which Plato spoke—the Soul of the World, has a similar relation to the sum total of

sentient constituents in the cosmos. As Hartshorne (1984, 59) explains, "God, the World Soul, is the *individual integrity* of 'the world,' which otherwise is just the myriad creatures. As each of us is the supercellular individual of the cellular society called a human body, so God is the super-creaturely individual of the inclusive creaturely society." This also means, as Plato proposed as well, that the physical universe is the divine body (Hartshorne 1984, 52–53, 133–36). Griffin says succinctly that the universe is thus "a compound individual with God as its dominant member" (Griffin 2001, 142; see also Dombrowski 2005). The analogy is not fully exact, however, primarily because it does not, of course, take a village of sentient creatures to raise a presumably preexistent or primordial World Soul. But the essential point is the emphasis on the mutually affective relation between creaturely souls and the World Soul, an essential interdependence that Hartshorne (1934, 168) characterizes as love, noting that "there is but one love, the participation of life in life, of feeling in feeling."

A contemporary Plato scholar, reflecting on the prominence of similar cosmological ideas over many previous centuries in the history of Western thought, has said that "there are few other concepts that so encapsulate the gulf between us and our philosophical inheritance as that of the world soul" (Wilberding 2021, 2). With its encompassing vision, the philosophy of organism can perhaps also contribute to remedying this most profound loss of soul and to fostering a reenchantment of the world within a naturalistic framework that can deepen the dialogue between science and religion and occasion a sweet reunion of body and soul.

Notes

[1] Based on our own experiences of the varying degree of influence we as presiding entities exert over our bodies, it seems that the nature of our personal monarchy varies from time to time from absolute to constitutional.

[2] Whitehead here is referring specifically to Plato's rational soul. Regarding such a soul's probable continued life after bodily death, see Griffin (2001, 230–46).

[3] For a video from this research, see https://www.youtube.com/watch?v=E8oIitQN2M4.

[4] For an excellent video produced by the researchers, see https://www.uvm.edu/news/story/scientists-create-next-generation-living-robots; and another one, with Whiteheadian resonances, on cellular cognition and collective intelligence: https://www.youtube.com/watch?v=0a3xg4M9Oa8.

[5] Goltz's experiments in the 1860s were apparently the original source of the analogy between unresponsive frogs in warming water and unresponsive people to evidence of climate change. Later experiments demonstrated that if the temperature is raised slowly enough, even intact frogs will not react. The speed of warming makes all the difference. Perhaps we can hope that the accelerated rate of climate change now underway will awaken and energize many more people, with brains fully intact and alert, to respond wisely.

[6] Galen Strawson (2017, 382–83, 386) points out that more recent physics also provides no support for the notion of mere nonexperiential matter and he argues that there is no good evidence that such a thing actually exists.

References

Agar, W. E. 1951. *A Contribution to the Theory of the Living Organism*. Melbourne: Melbourne University Press.

Ben-Jacob, Eshel, Yoash Shapira, and Alfred I. Tauber. 2011. "Smart Bacteria." In *Chimeras and Consciousness*, edited by Lynn Margulis, Celeste A. Asikainen, and Wolfgang E. Krumbein, 55–62. Cambridge, MA: MIT Press.

Binet, Alfred. 1888. *The Psychic Life of Micro-Organisms*. Chicago: Open Court.

Birch, Charles. 1977. "Can Evolution Be Accounted for Solely in Terms of Mechanical Causation?" In *Mind in Nature*, edited by John B. Cobb and David Ray Griffin, 13–18. Washington, DC: University Press of America.

Birch, Charles, and John B. Cobb. 1981. *The Liberation of Life*. Cambridge: Cambridge University Press.

Blackiston, Douglas, Emma Lederer, Sam Kriegman, Simon Garnier, Joshua Bongard, and Michael Levin. 2021. "A Cellular Platform for the Development of Synthetic Living Machines." *ScienceRobotics* 6 (eabf1571): 1–13.

Boring, Edwin G. 1950. *A History of Experimental Psychology*. New York: Appleton-Century-Crofts.

Brooke, John Hedley. 1991. *Science and Religion*. Cambridge: Cambridge University Press.

Burggren, W. W., and S. Warburton. 2007. "Amphibians as Animal Models for Laboratory Research in Physiology." *ILAR Journal* 48 (3): 260–69.

Coleman, William. 1977. *Biology in the Nineteenth Century*. Cambridge: Cambridge University Press.

Crabbe, James. 1999. "Introduction." In *From Soul to Self*, edited by James Crabbe, 1–7. London: Routledge.

Cunningham, Andrew. 2003. "The Pen and the Sword." *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 34 (1): 51–76.

Debru, Armelle. 2015. "Galen's Approach to Anatomy and the Soul." *History of Medicine* 2 (2): 127–31.

Del Maestro, Rolando. 1998. "Leonardo da Vinci: The Search for the Soul." *Journal of Neurosurgery* 89 (5): 874–87.

Demarest, Boris, Jonathan Regier, and Charles Wolfe. 2021. "Animism and its Discontents: Soul-Based Explanations in Early-Modern Natural Philosophy and Medicine." *HOPOS: The Journal of the International Society for the History of the Philosophy of Science* 11 (2): 494–501.

Dexter, Joseph P., Sudhakaran Prabakaran, and Jeremy Gunawardena. 2019. "A Complex Hierarchy of Avoidance Behaviors in a Single-Celled Eukaryote." *Current Biology* 29 (Dec. 16): 4323–329.

Dombrowski, Daniel A. 2005. *A Platonic Philosophy of Religion: A Process Perspective*. Albany, NY: SUNY Press.
———. 2021. *Process Philosophy and Political Liberalism: Rawls, Whitehead, Hartshorne*. Edinburgh: Edinburgh University Press.
Emmet, Dorothy. 1966. *Whitehead's Philosophy of Organism*. New York: St. Martin's.
Ewert, Jörg-Peter, and Michael A. Arbib, eds. 1989. *Visuomotor Coordination: Amphibians, Comparisons, Models, and Robots*. New York: Plenum.
Fearing, F. (1930) 1970. *Reflex Action*. Cambridge, MA: MIT Press.
Frede, Michael. 1995. "On Aristotle's Conception of the Soul." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amélie O. Rorty, 93–107. Oxford: Oxford University Press.
French, R. K. 1969. *Robert Whytt, the Soul, and Medicine*. London: Wellcome Institute of the History of Medicine.
Gilbert, Stephen G. 1965. *Pictorial Anatomy of the Frog*. Seattle: University of Washington Press.
Goltz, Friedrich. 1869. *Bieträge zur Lehre von den Functionen der Nervencentren*. Berlin: August Hirschwald.
Griffin, David Ray. 1998. *Unsnarling the World Knot*. Berkeley: University of California Press.
———. 2001. *Reenchantment without Supernaturalism*. Ithaca, NY: Cornell University Press.
———. 2007. *Whitehead's Radically Different Postmodern Philosophy*. Albany, NY: SUNY Press.
Haller, Albrecht von. (1755) 1936. "Dissertation on the Sensibility and Irritability of the Parts of Animals." In *Bulletin of the Institute of the History of Medicine* 4 (8): 651–99.
Handrigan, Gregory R., and Richard J. Wassersug. 2007. "The Anuran *Bauplan*." *Biological Reviews* 82 (1): 1–25.
Hartshorne, Charles. 1934. *The Philosophy and Psychology of Sensation*. Eugene, OR: Wipf and Stock.
———. (1936) 1972. "The Compound Individual." In *Whitehead's Philosophy: Selected Essays, 1935–1970,* 41–61. Lincoln, NE: University of Nebraska Press.
———. 1950. "Panpsychism." In *A History of Philosophical Systems*, edited by Vergilius Ferm, 442–53. New York: Philosophical Library.
———. 1953. *Reality as Social Process*. Glencoe, IL: The Free Press.
———. 1962. *The Logic of Perfection*. LaSalle, IL: Open Court.
———. 1973. *Born to Sing: An Interpretation and World Survey of Birdsong*. Bloomington, IN: Indiana University Press.
———. 1984. *Omnipotence and Other Theological Mistakes*. Albany, NY: State University of New York Press.
Hatton, Ian A., Eric D. Galbraith, Nono S. C. Merleau, Teemu P. Miettinen, Benjamin McDonald Smith, and Jeffrey A. Shander. 2023. "The Human Cell Count and Size Distribution." *PNAS* 120 (39): e2303077120. https://doi.org/10.1073/pnas.2303077120.
Henderson, Lawrence J. 1926. "A Philosophical Interpretation of Nature: Review of *Science and the Modern World* by Alfred North Whitehead." *Quarterly Review of Biology* 1 (2): 289–94.
Henning, Brian G. 2023. *Value, Beauty and Nature: The Philosophy of Organism and the Metaphysical Foundations of Environmental Ethics*. Albany, NY: SUNY Press.
Holmes, F. L. 1993. "The Old Martyr of Science: The Frog in Experimental Physiology." *Journal of the History of Biology* 26: 311–28.
Huxley, Thomas Henry. (1870) 2015. "Has a Frog a Soul; and of What Nature is That Soul, Supposing It to Exist?" In *The Papers of the Metaphysical Society, 1869–1880*, edited by Catherine Marshall, Bernard Lightman, and Richard England, 177–84. Oxford: Oxford University Press.
———. 1874. "On the Hypothesis that Animals are Automata, and Its History." *Fortnightly Review* 16 (New Series, Nov. 1): 555–80.
James, William. 1890. *The Principles of Psychology*. Vol. 1. New York: Henry Holt.
———. 1904. "Does 'Consciousness' Exist?" *Journal of Philosophy, Psychology, and Scientific Methods* 1 (18): 477–91.
Jaynes, Julian. 1976. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin.
Jennings, H. S. 1906. *Behavior of the Lower Organisms*. New York: Columbia University Press.
Jung, Carl. 1933. *Modern Man in Search of a Soul*. London: Kegan Paul.

Klein, Alexander. 2018. "The Curious Case of the Decapitated Frog: On Experiment and Philosophy." *British Journal for the History of Philosophy* 26 (5): 890–917.
Krumbein, Wolfgang E., and Celeste A. Asikainen. 2011. "Ancient Architects." In *Chimeras and Consciousness*, edited by Lynn Margulis, Celeste A. Asikainen, and Wolfgang E. Krumbein, 63–70. Cambridge, MA: MIT Press.
Latour, Bruno. 1993. *We Have Never Been Modern*. Translated by Catherine Porter. Cambridge, MA: Harvard University Press.
———. 2011. "What Is Given in Experience?" Preface to *Thinking with Whitehead* by Isabelle Stengers, ix–xv. Cambridge, MA: Harvard University Press.
Levin, Michael. 2019. "The Computational Boundary of a 'Self.'" *Frontiers in Psychology* 10:2688. doi: 10.3389/fpsyg.2019.02688.
———. 2021. "Life, Death, and Self." *Biochemical and Biophysical Research Communications* 564 (2021): 114–33.
Lewes, George Henry. 1860. *The Physiology of Common Life*. Edinburgh: Blackwood and Sons.
———. 1873. "Sensation in the Spinal Cord." *Nature* (Dec. 4): 83–84.
Lorenz, Hendrick. 2009. "Ancient Theories of the Soul." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/sum2009/entries/ancient-soul/.
Miller, Gordon L. 2021. "The Intervening Touch of Mentality: Food Seeking in Frogs and Whitehead's Philosophy of Organism." *Process Studies* 50 (2): 155–200.
Moore, Thomas. 1992. *Care of the Soul*. New York: HarperCollins.
Murphy, Nancey. 1998. "Human Nature: Historical, Scientific, and Religious Issues." In *Whatever Happened to the Soul?*, edited by Warren S. Brown, Nancey Murphy, and H. Newton Malony, 1–29. Minneapolis, MN: Fortress Press.
Needham, Joseph. 1951. "A Biologist's View of Whitehead's Philosophy." In *The Philosophy of Alfred North Whitehead,* edited by Paul Arthur Schilpp, 243–71. New York: Tudor.
———. 1980. *Lost Christianity*. Garden City, NY: Doubleday.
Peirce, Charles Sanders. 1960. *The Collected Papers of Charles Sanders Peirce*. Vol. 1. Edited by Charles Hartshorne and Paul Weiss. Cambridge, MA: Harvard University Press.
Popper, Karl R., and John C. Eccles. 1977. *The Self and Its Brain*. New York: Springer.
Reid, Chris R., and Tanya Latty. 2016. "Collective Behaviour and Swarm Intelligence in Slime Molds." *FEMS Microbiology Reviews* 40:798–806.
Riskin, Jessica. 2016. *The Restless Clock*. Chicago: University of Chicago Press.
Roe, Shirley. 2003. "The Life Sciences." In *The Cambridge History of Science—Vol. 4 Eighteenth-Century Science*, edited by Roy Porter, 397–416. Cambridge: Cambridge University Press.
Skrbina, David. 2017. *Panpsychism in the West*. Cambridge, MA: MIT Press.
Sleigh, Charlotte. 2012a. *Frog*. London: Reaktion Books.
———. 2012b. "Jan Swammerdam's Frogs." *Notes and Records of the Royal Society of London* 66 (4): 373–92. https://www.jstor.org/stable/41723322.
Sölch, Dennis. 2020. "Whitehead's Biological Turn." In *Whitehead at Harvard 1924–1925*, edited by Brian G. Henning and Joseph Petek, 100–15. Edinburgh: Edinburgh University Press.
Sorabji, Richard. 1999. "Soul and Self in Ancient Philosophy." In *From Soul to Self*, edited by James Crabbe, 8–32. London: Routledge.
Strawson, Galen. 2017. "Physicalist Panpsychism." In *The Blackwell Companion to Consciousness*, edited by Susan Schneider and Max Velmans, 374–90. New York: John Wiley & Sons.
Stuart, Alexander. 1739. *Three Lectures on Muscular Motion*. London: T. Woodward.
Swammerdam, J. 1758. *The Book of Nature*. Translated by Thomas Floyd, revised by John Hill. London: C. G. Seyffert.
Thorpe, William H. 1978. *Purpose in a World of Chance*. Oxford: Oxford University Press.
Verworn, Max. 1907. *Physiologisches Praktikum für Mediziner*. Jena, Germany: G. Fischer.
Waddington, Conrad. 1975. *The Evolution of an Evolutionist*. Ithaca, NY: Cornell University Press.
Whitehead, Alfred North. 1911. *An Introduction to Mathematics*. New York: Henry Holt.
———. 1925. *Science and the Modern World*. New York: Free Press.

———. (1929) 1978. *Process and Reality*. New York: Free Press.
———. 1933. *Adventures of Ideas*. New York: Macmillan.
———. 1938. *Modes of Thought*. New York: Free Press.
Wright, Sewell. 1964. "Biology and the Philosophy of Science," *The Monist* 48(2): 265–90.
Whytt, Robert. 1751. *An Essay on the Vital and other Involuntary Motions of Animals*. Edinburgh: Hamilton, Balfour, and Neill.
———. (1763) 1768. *The Works of Robert Whytt*. Edinburgh: Becket, Dehondt, and Balfour.
Wilberding, James, ed. 2021. *World Soul: A History*. New York: Oxford University Press.

Zygon®: Journal of RELIGION & SCIENCE

Mallary, Mitchell. 2024. "Scientific Methods and Sacred Matters: On Methodological Naturalism, Divine Agency, and Christians in Science." *Zygon: Journal of Religion and Science* 59 (3): 638–60. DOI: https://doi.org/10.16995/zygon.11761

Open Library of Humanities

Scientific Methods and Sacred Matters: On Methodological Naturalism, Divine Agency, and Christians in Science

Mitchell Mallary, Academic Administrator for Scholarship and Christianity in Oxford, Oxford, UK, mitch.mallary@scio-uk.org

This article critically analyzes the debate surrounding methodological naturalism (MN) within the context of scientific inquiry, focusing on the arguments put forth by Andrew Torrance on the one hand and John Perry and Sarah Lane Ritchie on the other. It begins with the medical mystery known as the spontaneous remission of cancer—sometimes identified as a "miracle"—posing the question of how Christian medical researchers should approach this anomaly. Thereafter, I consider how Torrance's rejection of MN, due to its perceived conflict with faith, would expect the Christian scientist to respond. Conversely, Perry and Ritchie defend MN, emphasizing its importance in maintaining scientific rigor and avoiding a "God of the gaps" approach. Subsequently, I demonstrate how both science and MN are conceived of differently between the two parties, indicating that this is primarily a semantic debate. Finally, after examining Torrance's response to Perry and Ritchie, I propose a more nuanced approach, advocating for methodological flexibility wherein the nature of the question being asked determines the appropriate method one should employ.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

The world is a strange place, but not as strange as it was once believed to be. In the not-too-distant past, people were relatively prone to ascribe various phenomena to supernatural forces or divine action. This was especially the case with health-related phenomena (such as mental illnesses, diseases, and epidemics), celestial events (like eclipses, comets, and meteor showers), and atmospheric occurrences (like droughts, floods, and lightning). Beliefs along these lines are evidenced, in varying degrees, by many of the writers and redactors of the texts that now compose the Hebrew Bible, the New Testament, and the Qur'ān. However, such occurrences, previously deemed supernatural or miraculous, have now been accounted for using scientific methods and natural explanations. Researchers in psychiatry, epidemiology, astronomy, and meteorology have made significant strides in uncovering the natural mechanisms and physical processes via which all of the aforementioned can be explained in a scientific fashion.

This pattern of "natural" explanations displacing prior "supernatural" ones does not apply only to beliefs held by ancient peoples or in religious texts. One oft-referenced example can be seen by appeal to Isaac Newton himself (on this, cf. Perry and Ritchie 2018, 1076–77; Donahue 2024, 14), who, while formulating the laws of motion and gravity, recognized that the gravitational interactions between planets (on his model) could lead to irregularities in their orbits around the sun over time. He speculated that these irregularities might occasionally require divine intervention to correct, lest their orbits be significantly disrupted and altered. Such theological speculation eventually proved to be unnecessary, as Pierre-Simon Laplace and other mathematicians using more advanced mathematical theorems demonstrated that these potential irregularities were actually self-correcting, and that the stability of our solar system could be explained without the need for divine intervention.

Cast against this intellectual backdrop, scientists have grown reasonably hesitant to invoke anything supernatural as a scientific explanation for worldly phenomena. Indeed, "[i]t has become standard practice for scientists to avoid the possibility of references to God . . . let alone to the possibility that God is actively involved in the world" (Torrance, 2017, 691). Among those who adopted such hesitancy completely, this eventually evolved into one of two distinct postures toward scientific study: (P1) the practical guideline, arrived at inductively in light of examples like those outlined to avoid supernatural hypotheses in the realm of science given their track record of failure in the past; or (P2) the normative claim that, strictly speaking, something about the nature of scientific study itself necessarily precludes any appeal to supernatural hypotheses. In both instances, there is thus now "the stipulation . . . to offer explanations only in terms of natural phenomena" (Donahue 2024, 1–2). This stipulation is a constituent feature of what is now referred to as "methodological

naturalism" (hereafter, MN), although, as I discuss, fuller definitions of this term and its purported implications can vary widely.

Be that as it may, Miles Donahue refers to P1 as "provisional methodological naturalism," and P2 as "intrinsic methodological naturalism" (2024, 4). Whereas the latter insists that there is something intrinsic to scientific reasoning that would be thwarted by invoking a supernatural hypothesis, the former simply chooses to avoid such invocations for practical purposes. To be sure, the stipulation of seeking only natural explanations shared by both does not imply or necessitate a commitment to metaphysical naturalism, defined here as "a philosophical perspective that denies the existence of a transcendent God," spiritual beings, and a supernatural realm (de Vries 1986, 389). The key takeaway, though, is that MN has become something of a default posture among most practitioners of contemporary science, regardless of whether it arises from P1 or P2.

Not everyone is convinced this is a good thing. For example, one of my doctoral supervisors, Andrew Torrance (2017), published an article entitled "Should Christians Adopt Methodological Naturalism?" This question, which he answered in the negative, prompted a rejoinder in defense of MN from two of his colleagues at the University of St Andrews, John Perry and Sarah Lane Ritchie (2018). To this rejoinder, Torrance (2018) offered his own response, clarifying and building upon his earlier position.

No doubt, all sorts of underlying assumptions are lurking in the background of these debates, not least concerning what is actually meant by MN. Beyond definitional equivocality, there are also underlying theological assumptions about the existence, nature, and agency of God, metaphysical assumptions about the nature and knowability of reality, and philosophical assumptions about the nature and scope of science. Complicating matters even further, precious few of these assumptions are self-evidently true. The best one can do, therefore, is to be honest about one's biases and presuppositions, clarify terms to the best of one's ability, and communicate charitably about those with whom one disagrees.

Fortunately, charity was on full display in the back-and-forth debate between Torrance on the one hand and Perry and Ritchie on the other. However, analytic clarity was occasionally lacking, and some of their respective motivating theological and metaphysical commitments were not foregrounded as prominently as they could have been. My aim is to help readers better navigate this debate (without personally choosing sides between what I see as a false dichotomy) before proposing my own way forward. To do this, I begin with a real phenomenon—cases of spontaneous and inexplicable remission of cancer—and discuss how both sides of the debate would expect Christian medical scientists researching the phenomenon to handle such a situation. Torrance, as a critic of MN, expects the Christian scientist qua scientist not to renounce the possibility that such an occurrence could be the result of divine agency, even as they continue to investigate the likelihood of natural causes.

Perry and Ritchie, as proponents of MN, expect the Christian scientist to explore every possible empirical avenue to explain such an occurrence through natural processes, even if they believe qua faith that it could in principle have been caused by the miraculous work of a divine being.

This discussion not only magnifies the differences in emphasis between the two parties to this debate but also illuminates how they occasionally talk past one another and misconstrue the opposing positions. As I discuss, much of this debate actually boils down to mere semantics—disagreements (or misunderstandings) about how words should be used—rather than disagreements about substance. Following this analysis, I then propose my own way forward, which is premised upon the notion that one's method should be dependent upon the question that is being asked. Against methodological uniformity in the natural sciences (whether Torrance's rejection of MN or Perry and Ritchie's defense of it), I propose methodological flexibility in which the research question asked determines the appropriate method to employ. Instead of asking, "How should we study *x* as Christians?"—resulting in endless throat-clearing debates about what method should be normative for Christian scholarship—it should instead more modestly be asked, "If we were to study *x* from [some perspective], what, if anything, can we conclude, and how should we assess the nature and truth-value of the conclusions from the perspective of faith?"

Framing matters this way draws upon Karl Barth's insight that revelation can never become a possession that is wielded into a hermeneutical or methodological system. The non-givenness of God and divine action requires a radical humility, for the living God cannot be contained within, or be straightforwardly accessible to, any methodology. Not all methods are created equal, but no single method is the "right" one for a properly theological understanding of the universe. As such, whatever method is used, resources of faith will always need to be drawn upon in order to assess the nature and truth-value of the conclusions reached. But before getting too far ahead of myself, let's dive right in.

The Spontaneous and Inexplicable Remission of Cancer

On rare occasions, medical experts have observed cancer patients undergoing spontaneous and inexplicable cancer remission, whereby malignant tumors are cured without targeted therapy or treatments (Radha and Lopus 2021). The frequency of such remission is exceptionally low, estimated at about one in 60,000 to 100,000 cancer cases (Dobosz and Dzieciatkowski 2019; Everson and Cole 1956; Jessy 2011). While certain cancers (e.g., melanomas, lymphomas, and leukemias) have shown a higher frequency of spontaneous remission (hereafter, SR) in comparison with other cancer types, the underlying reasons for remission itself remain largely speculative, despite some promising evidence about how an infection-induced immune response has been detected in several instances

of SR (Radha and Lopus 2021, 3–4, 7–8). Be that as it may, the occurrence of SR is occasionally explained (usually outside of the scientific community) as a "miracle" (Robson 2015). Some patients and their family members proclaim even more confidently that "God" (but which god and which conception of that god are they referring to?) has answered their prayers and directly performed a miracle. The question then arises: How should a Christian medical researcher, or any medical researcher for that matter, approach such a claim and the anomaly upon which it rests?

In fact, let's be even more precise. Consider the following hypothetical, yet entirely plausible, scenario. A Christian medical professional and researcher (who happens to believe in the bodily resurrection of Jesus) diagnoses a patient with terminal cancer, which is subsequently independently confirmed by other specialists. In light of this news, the patient and their family make the difficult decision to forgo treatment, opting instead for prayer and palliative care should the time come. During a routine checkup several months later, it is observed that the patient's cancer has undergone SR. The tumor, previously deemed incurable, has inexplicably regressed without any medical intervention or any of the accompanying factors that have been speculated to contribute to SR.

This would rightly be regarded as an anomaly, for it defies the usual expectations of medical science. However, an apparent anomaly—something currently inexplicable within contemporary scientific understanding but in principle able to be explained with further study or advances in science—is conceptually distinct from an actual anomaly or miracle (whether or not such things actually exist). Drawing upon Thomas Aquinas, Perry and Ritchie (2018, 1075) describe the latter as an instance in which "something . . . looks like God 'producing the *effects* of secondary causes without them'." According to Torrance (2017, 693), the existence of actual anomalies, while rare, is nevertheless to be expected on a Christian epistemic basis in which the divine creator of the universe takes on flesh, heals the sick, gives sight to the blind, walks on water, and rises from the dead.[1]

But how can one tell the difference between an apparent anomaly and an actual anomaly (i.e., a miracle)? How can one discern if a particular instance of SR is something currently puzzling but potentially explicable, or if there is simply no purely scientific explanation because divine agency was involved? When faced with this question, Perry and Ritchie (2018, 1078) argue that "it is impossible to empirically distinguish what belongs to which subset." Similarly, Torrance (2017, 697) argues that "it belongs to Christian orthodoxy to affirm that neither God nor God's activity should be seen to be an object of scientific observation, speculation, and experimentation." In other words, both parties to this debate agree that actual anomalies can happen as a result of divine action, and both agree that the tools of science are insufficient to establish that they have. But this begs a question: How, then, should the Christian medical

researcher respond to this particular instance of SR when there is no way of knowing which subset of anomalies it belongs to? This is where the debate over MN comes to the fore.

Torrance's Rejection of Methodological Naturalism

For Torrance, cases of SR are not the sort of occurrences that automatically trigger the conclusion that they might have been caused by God in an exceptional way. In fact, Torrance (2017, 695) is always initially hesitant to "jump to the conclusion that 'God did it'" when faced with some anomalous contemporary phenomenon. Nevertheless, as a result of his conviction that it is possible that God may in fact bring about an instance of SR, Torrance believes this possibility should methodologically impact how a Christian scientist approaches anomalies like the hypothetical case study described in the previous section. Unlike the atheist, who would immediately rule out the possibility that the patient's SR was the result of divine intervention, Torrance (2017, 695) suggests that the Christian scientist should instead "be willing to allow her belief in the possibility of theological explanation to inform her scientific assessment of this occurrence." Torrance, in other words, does not want to rule out the possibility that the anomaly could turn out to be an actual anomaly rather than an apparent anomaly. This is the basis upon which he rejects MN.

As Torrance (2017, 695) argues, "if . . . it [is] *possible* that in any field of enquiry God could have an explanatory significance, then that recognition is inconsistent with MN." Where does the inconsistency emerge? To answer this, it is necessary to survey Torrance's understanding of what MN is, along with what he believes to be necessarily entailed by adopting it.

While recognizing that there are different ways to define MN (cf. Torrance, 2017, 720n3), Torrance (2017, 691) suggests that it is minimally understood as "a method that assumes that the reality of the universe, as it can be accessed by empirical enquiry, is to be explained solely with recourse to natural phenomena." Torrance then proceeds to outline numerous implications he associates with adopting MN. Consider the following claims he makes at various points throughout his article:

> MN's assumptions are "not metaphysically neutral" (Torrance 2017, 717).
>
> MN, if adopted by a Christian, indicates that one is "playing the games of the secular world" (Torrance 2017, 707).
>
> MN plays into "the illusion that there is nothing more to reality than natural phenomena" (Torrance 2017, 715).
>
> MN, if adopted, requires operating under the assumption that "God has nothing to do with the natural order" (Torrance 2017, 703).

> MN is "a methodology that is branded with a philosophy (naturalism), which is incompatible with theism" (Torrance 2017, 697).
>
> MN requires that the Christian "amputate the legs on which she stands so that she can be accommodated by the procrustean bed that is naturalism" (Torrance 2017, 707).
>
> MN "does not allow the Christian scientist qua scientist to recognize miracles as blind spots. MN requires the Christian scientist qua scientist to offer naturalistic explanations" (Torrance 2017, 701).
>
> MN "require[s the Christian scientist] to mis-explain the blind spots" (Torrance 2017, 702).
>
> MN will "bias science . . . in a way that needlessly makes a theory incompatible with theism" (Torrance 2017, 706).
>
> MN "assumes that there is an essential conflict between science and Christianity" (Torrance 2017, 707).
>
> MN "will undermine [a Christian] interpretation of natural history" (Torrance 2017, 701).

Collectively, these claims suggest that Torrance perceives the decision to adopt MN not merely as a commitment to seeking natural hypotheses where possible but also as embracing philosophical presuppositions that fundamentally conflict with Christianity. According to Torrance (2017, 701–02), the decision to adopt MN cannot straightforwardly be decoupled from a philosophical commitment to "natural-*ism*" in which one is required to provide purely "naturalistic explanations" for everything from the origin of the universe to the rise of resurrection faith.

Returning to the case study, then, Torrance believes that the practitioner of MN, if asked about God's potential involvement in the SR, would be forced to reject the possibility that God did it, for MN demands a purely naturalistic explanation. Their assumption, in other words, is that the reality of the universe must be explained only with recourse to natural phenomena. But assuming this, in Torrance's view, "would express a failure to recognize the ultimate truth of the Christian faith and/or a failure to be consistent in one's commitment to the truth" (Torrance 2017, 695). Put another way, Torrance (2017, 696) will not countenance a method in which one is compelled to "make assumptions that are incompatible with . . . Christian assumptions." Even if it is not automatically deemed likely, Christian faith demands that Christians say it is possible that God was the cause of the patient's SR. "For this reason, the Christian scientist has a

duty not only to God but to science itself to discourage a commitment to MN" (Torrance 2017, 718).

Now, to be sure, the possibility of a theological explanation for the SR need not imply the reality that God has in fact miraculously intervened in the life of this previously terminal cancer patient. As Torrance (2017, 695) notes, the Christian medical researcher should never preemptively bring science to a halt and "jump to the conclusion that 'God did it'." In such a situation, the Christian scientist ought to seek to make sense of such an event in "natural terms" if and where possible (Torrance 2017, 695). In fact, Torrance (2017, 693, 694) even goes so far as to say that because "God is transcendent," it must be concluded that divine action "is not discernible by empirical study." This means that Torrance (2017, 693, 694) appeals to theology—"a theological appreciation of human limitations"—in order to identify the scope and limits of scientific study vis-à-vis God: "There are clear theological reasons for the Christian scientist to assume that her research should be characterized by an absence of explicit theological reference. She should recognize that God's activity should not be confused with the regular and immanent processes that characterize the natural world, and which are the subject-matter of normal scientific enquiry."

Here, however, a surface-level tension emerges in Torrance's account. On the one hand, he asks the Christian scientist to remain open to the possibility that God could be the explanatory cause of the patient's SR. On the other hand, he argues that empirical investigation could never coherently arrive at the conclusion that divine agency was in fact the cause of the SR. As such, the Christian scientist should avoid "explicit theological reference" (Torrance 2017, 694). How, then, does this posture of openness substantively impact one's scientific methodology?

As far as I can tell, Torrance (2017, 695) does not actually argue that it does. Indeed, he explicitly states that such openness is not "likely to have a decisive impact on . . . scientific research." He even adds at a later point that his theological argument about excluding reference to divine agency in natural science would "make no external difference to . . . scientific practice" (Torrance 2017, 700). Taken together, these two statements imply that Torrance's rejection of MN is not premised upon an endorsement of an alternative explicitly Christian scientific method that one could appeal to in order to gain special access to divine agency and thereby detect miracles. Rather, he rejects MN because he believes it requires its practitioners to deny the possibility—for the sake of engaging in scientific study—that the God of Christian faith could have been the explanatory cause of the SR. In short, even if Torrance's theologically delineated scope of scientific inquiry means that natural science cannot prove, or even propose, that "God did it," the scientist should nevertheless avoid the conclusion (demanded by Torrance's conception of MN) that there must be a purely natural explanation for the SR. Christians in academia and other

professions should not, in other words, be forced by their secular counterparts to adhere to standards that are in tension with their Christian commitments.[2] Or so Torrance argues.

Perry and Ritchie on the Nature of Scientific Inquiry

Perry and Ritchie tread a starkly different path. Rather than primarily tackling the question of MN head-on, they instead home in on the nature of scientific inquiry itself, along with the manner in which science ought to approach anomalies like those outlined in the previous sections. One of the central threads of their argument is that "epistemology . . . is discipline specific" (Perry and Ritchie 2018, 1073). In other words, each discipline "has its own set of practices, traditions, virtues, and criteria for verification" (Perry and Ritchie 2018, 1071). Furthermore, that which "count[s] as evidence in physics is not necessarily the same for biology, much less in psychology and theology" (Perry and Ritchie 2018, 1071).

In relation to the broader domain of science itself, they argue that scientific inquiry is largely focused upon what "we can know with our senses, that is, empiricism," and as such, "all scientific explanations are limited to empirical explanations" (Perry and Ritchie 2018, 1071, 1073). While their restriction of the natural sciences to empirical explanations alone might be a slight overstatement (cf. Torrance 2018, 1099–100), this general point serves as the context in which their proposed response to anomalies like SR is to be understood.

For Perry and Ritchie, all present-day anomalies—including the aforementioned case of SR—ought to be approached by scientists, regardless of their religious commitments, as potentially apparent anomalies. Even if they believe it may be an actual anomaly, the scientist should not rule out the possibility that it might turn out to be nothing more than a currently puzzling, but ultimately comprehensible, occurrence. So, in the case of a patient's terminal cancer transitioning into inexplicable remission, every scientist ought to "record it as an anomaly, restate [their] preliminary conclusion in the form of a *testable* hypothesis, and await more data. If something seems anomalous right now nothing should compel the careful scientist to render a conclusion in advance of sufficient data. Such a scientist will always *keep digging*" (Perry and Ritchie 2018, 1075).[3]

This is not because of a commitment to atheism or metaphysical naturalism, which denies the very possibility of divine agency or miracles. Rather, it derives from a commitment that there is something inherent in the scientific method that means it "can never say that an observable event has or has not been caused by God" (Perry and Ritchie 2018, 1073). It also derives from a steadfast "focus on the tools of empirical research" that characterize scientific inquiry (Perry and Ritchie 2018, 1075). If such tools are incapable of providing a natural explanation on the basis of the available evidence, the scientist qua scientist

should not, as a purported inference to the best explanation, "posit divine action as a causal factor in anomalous events—but neither [should they] invent a naturalistic explanation in order to preserve an overarching metaphysical naturalist worldview" (Perry and Ritchie 2018, 1075–76). In fact, a "good scientist understands that 'I don't know' or 'I don't know yet' is always a valid answer" (Perry and Ritchie 2018, 1076), and one that is preferable to any "God of the gaps" hypothesis.

In light of this discussion, Perry and Ritchie (2018, 1075) propose, axiomatically, that "all scientists should be methodological naturalists, or something close to it." As they understand it, though, MN does not require its practitioners to say (as Torrance believes) that it is not possible God did it, but it does require them to say that they will be looking for a natural explanation even though it is possible God did it. Perry and Ritchie (2018, 1076, 1075) describe this as a posture "of curiosity and willingness to pursue further research," yet in such a way that it "does not involve an active prohibition on God's activity." Consequently, objections to MN will purportedly "fall away" once "confusion about how the scientific method treats anomalies . . . is cleared up" (Perry and Ritchie 2018, 1084).

Varying Conceptions of Science and Methodological Naturalism

What can be concluded from the foregoing discussion? First and foremost, it becomes immediately clear that Torrance's conception of the nature of scientific inquiry differs fundamentally from that of Perry and Ritchie. What is more, they also have divergent conceptions of the nature and implications of MN. This needs to be spelled out in greater detail.

Because Torrance denies that Christians in science should practice MN, he likewise denies that science is intrinsically committed to MN. As a result, he conceives of scientific inquiry somewhat along the lines presupposed by P1, in which MN may be regarded as a nonessential procedure adopted (unfortunately, in Torrance's view) by many practitioners of science. Perry and Ritchie (2018, 1068), by way of contrast, seem to conceive of scientific inquiry as intrinsically committed to MN: "MN was always the name of the scientific game going as far back as Boyle, Newton, and even Aquinas and his mentor, Albert the Great." This is evidenced further when they seemingly equate MN with "the scientific method" in general, and when they state that rejecting MN would mean "abandon[ing] the scientific method's focus on the empirical toolset" (Perry and Ritchie 2018, 1073, 1084). All this indicates that their conception of scientific inquiry aligns with P2, in which MN is necessitated by either the aims or the methods (or both) of natural science.

There is thus a degree of irony at play here. Upon closer examination, one discovers that the debate is in fact not primarily about what MN is or whether Christians should adopt MN (important though those questions may be). Rather, the logically prior question undergirding both pertains to the nature

of science itself. In my view, this background disagreement is one of the main reasons the two parties end up talking past each other on several occasions. I return to this point in due course.

One might be tempted at this juncture to try to solve this dispute through asking "what is science?" and "what is MN?" Once concrete answers are given, one might then enquire as to whether the proffered definition of science necessarily entails a commitment to MN. If it does, then any argument that Christians should not practice MN (thus defined) would also circuitously claim that Christians should not practice science at all. If it does not, then conceptual space would be created to argue about when, if ever, MN ought to be adopted by Christian scientists.

I, however, have no desire to police language and argue for an essentialist definition of science or MN. Following Larry Laudan (1983), it appears that such efforts to identify a neat delineation between "science" and "non-science" have proven elusive. This problem extends beyond science as well: "Virtually every major term used in philosophical, theological, or cultural discourse shows at least some degree of definitional elusiveness or ambiguity" (McGrath 2016, ch. 1, §3). What is more, David Chalmers (2011, 532) has successfully demonstrated that many disputes pertaining to questions of the form "what is X?" are reducible to mere verbal disputes—that is, disagreements merely about how certain words should be used and nothing more. As such, instead of investing substantial effort into debates about the superiority of one definition over another for specific terms or phrases in the English language, it seems more productive to focus instead on identifying whether and where there are any substantive disputes.

For example, if I were to temporarily set aside the semantic debates about the meanings of "science" and "MN," I could turn my attention to substantive questions such as:

(1) Do the two parties agree that divine action could ever in principle be the explanatory cause of some natural phenomena?
(2) Do the two parties agree that recognizing and properly identifying divine action transcends the scope of empirical study and experimental testing?
(3) Do the two parties agree that the Christian scientist should never, in any discipline, operate under the assumption that God does not exist?
(4) Do the two parties agree that a Christian scientist should avoid claiming, when operating as a scientist, that it is not possible God was the cause of a SR?

Interestingly, when attention is turned to these substantive issues, a great degree of concord is discovered. I have already discussed (1) and (2): on the one hand, both parties agree that it is in principle possible for divine action to be, at least

in part, the explanatory cause of some worldly phenomena. On the other, they likewise agree that the detection of such divine action exceeds the scope of that which may be known through empirical observation, scientific theorizing, and experimentation. Notably, the agreement does not stop there. Regarding (3), it is similarly found that both parties insist it would be inappropriate for Christians to act as if God does not exist—to functionally renounce their faith—when engaging in academic scholarship of any sort. And finally, both parties are crystal clear (4) that a Christian scientist should not—whether in the laboratory, the cancer ward, the classroom, the church, or anywhere else—claim it impossible that divine agency was involved in bringing about a patient's SR. In short, both parties are in full agreement that the core entailments of MN should be avoided at all costs, while faith in the power of God must be clung to at all times.

Given these areas of fundamental agreement, one might question Perry and Ritchie's claim that this debate is "more than semantic" in nature (Perry and Ritchie 2018, 1066). In fact, it seems to me that the disagreement is mostly semantic, boiling down simply to how the terms "science" and "MN" are defined. This comes into sharper focus when Torrance (2017, 720n3) acknowledges that some people "equate MN with scientific enquiry" itself, noting that he of course does not want to argue "against a Christian practicing scientific enquiry." He even adds that "if this is how MN should be defined, [he] would not have a problem with MN" (Torrance 2017, 720n3).

However, if my argument about Perry and Ritchie proves sound, this is precisely how they understand MN (cf. Perry and Ritchie 2018, 1073). For Perry and Ritchie, MN is simply inherent to scientific investigation and explanation. Science's unique tools and criteria for verification have a relatively focused scope, and this scope does not include the ability to adjudicate upon or propose theological hypotheses. MN is not a choice but rather a characteristic of science (as they define it); as such, it should theoretically fall under the umbrella of that which Torrance (2017, 720n3) "would not have a problem with"—especially since they agree with Torrance on points (1–4).

The heart of this dispute, then, revolves around their divergent conceptions of science and MN. From Torrance's discussion, it becomes clear that he construes science in exceptionally broad terms, largely as a comprehensive search for truth (cf. Torrance 2017, 695, in which he draws upon a quote attributed to Einstein), one that is not inherently limited to empirical investigation and experimentation alone, given that it must leave room for the possibility of theological explanation. This genus of science, for Torrance, has at least two species: natural science on the one hand, and theological science on the other (for the latter, cf. Torrance 2019). For Perry and Ritchie, by contrast, science (i.e., the conceptual equivalent of Torrance's natural science) is a distinct disciplinary domain with its own tools, methods, and principles of verification,

all of which necessarily exclude theological hypotheses as an explanatory category (distinguishing it from Torrance's characterization). On the basis of these divergent conceptions of (natural) science, they then construe MN in vastly different ways. Torrance, for his part, argues that the decision to adopt MN entails incorporating the philosophical commitments of naturalism into one's methodology, requiring the practitioner to operate as if God and divine action do not exist. Conversely, Perry and Ritchie argue that MN simply entails a search for natural explanations (where possible) when engaged in the scientific study of natural phenomena. The distinction might be characterized thus:

> MN1: Methodological naturalism understood as the methodology according to which the philosophical entailments of naturalism are assumed.
>
> MN2: Methodological naturalism understood as the methodology according to which one must not propose or adjudicate upon supernatural/metaphysical/theological hypotheses.

Purely for the sake of argument, then, one could theoretically agree with Torrance that Christians should reject MN if it is construed in such a way as to necessitate making naturalistic claims at odds with one's faith (e.g., "it is not possible God caused the SR") while also agreeing with Perry and Ritchie that Christian scientists can affirm MN if it is defined as an intrinsic feature of the aims or methods of science, wherein supernatural explanations are not permitted within the scope of empirical reasoning, theorizing, and experimentation. One could (and I suggest Perry and Ritchie would) rightly reject the appropriateness of Christians using MN1—or at least Torrance's construal of it—and accept, on a question-by-question basis, P2/MN2. Indeed, one could even appeal to Torrance's own theological reasoning in relation to God's transcendence as a reason to accept a theologically delineated version of P2/MN2 in the context of natural science. Of course, if it is theologically informed, one might not choose or prefer to label this "methodological naturalism."

Whatever the case may be, I propose that one need not worry too much about what language is used to describe the posture—demonstrated earlier to be endorsed collectively by Torrance, Perry, and Ritchie—in which natural science is understood (for theological and scientific reasons) to necessarily preclude any appeal to supernatural hypotheses. It could be called "methodological naturalism" if it is clear that it is not defined in such a way as to entail the adoption of the philosophical commitments of naturalism. Or, following Torrance's advice, it could be called something else, because the term "naturalism . . . is normally seen to be an ism that is committed to suppositions that are incompatible with Christianity" (Torrance 2017, 704). I personally have no stake in the game either way, as I am inclined to consider

semantic disagreements about what term to use to be less important than being analytically precise about defining that term and the roles a particular form of inquiry seeks to fulfill (cf. Chalmers 2011).[4] Be that as it may, it seems to me that on several points the two parties to this debate are in lockstep agreement, indicating that this debate is mostly semantic in nature.

This is not to say that a semantic disagreement is all that is at play here. There are certainly substantive differences at the peripheries of their two proposals, as discussed in the following sections. But the main takeaway is that, in relation to science and MN itself, they are talking about different things. Torrance wants Christians to reject MN because it supposedly requires the assumption that there must be a naturalistic explanation.[5] Perry and Ritchie affirm MN as an intrinsic feature of the sciences (i.e., seeking only natural explanations on the basis of the limitations of scientific tools and reasoning), but they deny the notion that, when faced with an anomaly, it requires setting aside or denouncing belief in the possibility of a theological explanation. Torrance wants the Christian scientist to avoid saying there must be a purely naturalistic explanation; Perry and Ritchie insist that this is not entailed by their understanding of MN. All of this must be kept in mind for the discussion that follows.

Torrance Responds

In light of the comments and criticisms posed to him by Perry and Ritchie, Torrance was prompted to clarify and build upon his earlier argument. To his credit, Torrance charitably handles instances in which Perry and Ritchie misconstrue his original article or ascribe to him conclusions, or potential conclusions, not entailed by his argument. I was initially taken aback, for example, by their contention that Torrance was advocating for an incompatibilist view of divine agency insofar as miracles are concerned, "in which affirming an event as divine action implies that there is no natural explanation for that event" (Perry and Ritchie 2018, 1072; cf. 1090n8). Given his Barthian inclinations, I read Torrance's article with the presupposition, which he subsequently corroborates, that he does "not think there is a competitive relationship between God and the dynamics within the created order" (Torrance 2018, 1098). That said, this does not mean he is a compatibilist in the Thomistic sense (cited by Perry and Ritchie), whereby "God is ultimately the cause of everything" (Lindberg 2007, 240–41), including evils such as "the Holocaust, sexual abuse, and the like" (Torrance 2018, 1098).

This is not the only area where Perry and Ritchie unhelpfully ascribe positions to Torrance that he would not endorse. Another appears when they claim that, in his rejection of MN, Torrance leaves room for the theistic scientist to adopt a "God of the gaps" approach to anomalies, wherein God is invoked as a hypothesis to account for currently inexplicable phenomena

(Perry and Ritchie 2018, 1077). Whereas adopting MN would prevent Newton's (at the time) gap-filling suggestion that God would need to intervene in order to uphold the orbital structure of the solar system,[6] Torrance's rejection of MN apparently "opens the door to the sort of 'miracle' posited by Newton, one which in retrospect proves embarrassing" (Perry and Ritchie 2018, 1077). However, Torrance's original article, as he reiterates in his response, leaves no room for "such haphazard reference to divine action" (Torrance 2018, 1097). In fact, he even directly spells out how a Christian scientist would need to respond comments such as those from Newton: "She could suggest that, for the purposes of their particular research program, they should assume that they are unable to discern such action because, if it is taking place, it is not discernible by empirical study" (Torrance 2017, 694).

This closes the door on any "God of the gaps." Apart from a few key theological exceptions, which are discussed in the following section, Torrance (2018, 1097) "do[es] not think a Christian scientist should identify a miracle," and should instead continuously "search for a possible natural explanation." Refusing to shut down the possibility of divine action is one thing. Jumping to the conclusion that God has in fact acted is quite another.

Understood thus, Torrance cannot justly be charged with wanting to "stop science from doing more" in the face of present-day anomalies (as implied by Perry and Ritchie 2018, 1082). For Torrance (2018, 1097, 1096), the "openness" to the possibility of ongoing divine action should not "make any difference to . . . scientific study," for it is "the duty of the scientist to keep digging."

There are select instances, however, in which Torrance does advocate for "stopping science," and these emerge in relation to the theological exceptions hinted at above. Here is where Torrance clarifies and expands his earlier argument. According to Torrance (2018, 1096), there are certain miracles, or actual anomalies—including *creatio ex nihilo*, the incarnation, the resurrection, and the ascension— that "need to be interpreted theologically (because the Christian does not believe a natural explanation is possible) . . . [and] it would be confused to try to interpret them in natural terms (in the way that scientists should try to explain all other phenomena)."

When faced with these particular phenomena, Torrance argues that the Christian scientist should not let a commitment to MN prevent them from affirming these as facts. Moreover, given that the Christian believes these miracles lack any possible natural explanation, they "should not try to explain [them] naturalistically . . . [because] they are a blind spot for an 'empirical' investigation" (Torrance 2018, 1096).[7] One's theological commitments, in other words, should shape how they approach science. These commitments do not tell them how to do science, but, in the case of these specific miracles, tell them not to try. This is Torrance's (minimal but decisive) argument for a "theology-engaged science" in which "the tools of theology can serve the task of science"

(Torrance 2018, 1102).[8] Those Christians who choose to keep digging would be both unchristian and unscientific, functionally denying what they know to be true.

A Brief Digression: On Natural Science and the Four Miracles

One question that came to mind upon reading Torrance's response pertained to whether something like the Logos's anhypostatic assumption of a concrete human nature, as a subject of inquiry, actually falls under the purview or scope of scientific investigation. What would it even look like for a natural scientist to try to study the nonempirical and transcendental act of assumption? To be sure, at no point does Torrance argue that such a question would fall under the remit of "natural" science. But, given his broader conception, which includes the possibility of "theological" science (following the likes of Aquinas, Barth, and T. F. Torrance), he does believe that the incarnation can be approached scientifically. (Here, of course, one must recall the different conceptions of science that may be at play.) So, to reiterate, does the incarnation fall under the remit of natural science?

If it does not (such as is necessitated by those adopting P2), then theology would not have to tell the scientist to avoid trying to give a naturalistic explanation, for they would not (or, at the very least, should not) even be trying to give any scientific explanation with metaphysical ramifications at all. Likewise, for *creatio ex nihilo*, the resurrection, and the ascension.[9] Torrance (2018, 1097) is surely correct that these events would be understood as "a *sui generis* act of God," but precisely on the basis of that theological judgment, this seems to imply these things cannot be known through the forms of investigation typically associated with the natural sciences.

In fact, I propose, on the basis of theological considerations, that the following statements (among others) cannot be subjected to empirical or experimental analysis, and that as such, they fall outside the scope of scientific inquiry (at least when understood in the sense on offer from Perry and Ritchie):

- God created the world out of nothing.
- The word became flesh.
- The God of Israel bodily raised Jesus from the dead.
- The resurrected Jesus ascended to the heavenly tabernacle.

I therefore want to offer a minor conceptual clarification to Torrance's argument. I agree that the four miracles generate directly observable phenomena (such as the existence of the world and the zygote in Mary's uterus), and that these phenomena fall "within the domain of science" (Torrance 2018, 1096). In my judgment, however, the miraculous aspect of these phenomena is not straightforwardly an "empirically accessible object

of . . . study," and as such, this does not straightforwardly "plac[e] them [the four aforementioned hypotheses] within the domain of [natural] science" (Torrance 2018, 1096). (Of course, anyone adopting MN1 could not even say this much, which lies at the heart of Torrance's argument.) Instead, I regard the miraculous element to be revealed truths, and argue that no natural evidence corresponding with or generated by these events—such as the existence of the universe, the humanity of Jesus, the empty tomb, or the psychosomatic experiences of the disciples—could be coherently appealed to in order to either support or undermine their veracity. That is to say, one cannot argue *from* the generated phenomena *to* the truthfulness or falsity of the miracle, even though there is an inseparable relationship between the phenomena and the miracle that generated them. Adopting MN1 rules these claims out (as Torrance convincingly shows), but MN2, if properly understood within a theological framing, can accommodate such a posture toward the four aforementioned hypotheses.

Insights from Karl Barth prove instructive at this juncture. When unpacking the relationship between form and content in his doctrine of revelation, Barth ([1932] 1936, 325) explains: "Thousands may have seen and heard the Rabbi of Nazareth. But this 'historical' element was not revelation. The 'historical' element in the resurrection of Christ, the empty tomb as an aspect of the event that might be established, was not revelation. This 'historical' element, like all else that is 'historical' on this level, is admittedly open to very trivial interpretations too."

Appropriating Barth, it might be said that the worldly phenomena themselves were not the miracles. They may be traces, or physical ramifications, associated with divine action, but such physical phenomena are not themselves identical with the miraculous element at play. Theoretical physicists and/or cosmologists studying the events that transpired at or soon after T=0 are not directly studying the divine act of creation itself. If, hypothetically, a medical doctor would have been able to subject Mary to scans during the exact moment of Jesus's conception, they would not have been directly studying the divine act of incarnation. A historian investigating whether the tomb was indeed empty is not directly studying the divine act of resurrection. Similar things could be said in relation to the disciples' experiences of the ascension. Of course, a comprehensive explanation of these divinely generated phenomena will require recourse to faith, but such a theological explanation is precisely excluded by the conception of science at play in P2/MN2).

The main takeaway is that these particular miracles are not directly equated with any worldly phenomena (i.e., the data for science), and as such, the miraculous component is not an object of scientific study. But the worldly phenomena in principle available to the scientist—the Big Bang, Mary's reproductive system, an empty tomb, and the disciples' experiences—can be

approached with the posture that a natural (n.b., not naturalistic) explanation is worth seeking.[10] As Barth ([1932] 1936, 329–30) argues, "the historical contingency" in and through which miracles occur "can still be surveyed and explained in all possible dimensions." And if so, I do not want to tell scientists to stop what they are doing or avoid trying to discover the penultimate limits of empirical/theoretical knowledge (such as for the practitioner of P2/MN2) in relation to the origins of the universe or what took place on the first Easter Sunday. This brings me to my proposed way forward.

A Possible Way Forward

My proposal is relatively simple, yet it differs in crucial respects from the solutions offered by Torrance, Perry, and Ritchie. First, the method used should be dependent upon the question being asked—a well-formulated question. In that sense, I echo the complaint from Perry and Ritchie (2018, 1086–87) that most of these debates about method or the relationship between science (which scientific discipline or subdiscipline?) and religion (which religion and whose theology?) take place at an unhelpfully heuristic and abstract level. Instead, specific examples and questions must be focused on, asking what methods are appropriate to the task at hand. Saying this, however, forces me to be suspect of their claim that the Christian scientist must always adopt a particular methodology (namely, MN). How can they simultaneously claim that we should "work out . . . points of methodology on an ad hoc basis" depending upon the question at hand (Perry and Ritchie 2018, 1086) as well as advocate earlier in the article for a normative scientific methodology committed to MN?

Be that as it may, the question at hand should be formulated, much like an essay question, as specifically as possible. For example, instead of asking the broad question, "How should a Christian scientist/historian study the resurrection?", it should instead be asked: "On the basis of the empirical, theoretical, and experimental tools of the scientific method [defined here according to P2], what can be said in relation to the events surrounding the first Easter Sunday?" Of course, that is not the only legitimate question. One could also ask: "On the basis of the resources of faith, what ought to be said about what transpired on the first Easter Sunday?" These are different questions (and certainly not the only possible ones!), and they require different tools and different methods. Depending upon the question, an approach like Torrance's science-engaged theology might be appropriate; for other questions, MN might be appropriate.

Nevertheless, in cases where a particular research question requires the use of MN, I agree with Torrance that the Christian should, strictly speaking, suggest that there are clear theological reasons to avoid invoking divine agency or positing miracles when operating in the domain of natural science. God is wholly other, to echo the early Barth, and divine agency is qualitatively distinct from ordinary happenings, making it by nature inaccessible to empirical or

experimental investigation (on this, see Ritchie 2020, 195–96). It is theology, rather than the naturalistic demands of secular scholarship, that should guide the Christian scientist to rule out positing miracles or divine agency when engaging in the practice of natural science.[11] This will not change how one goes about doing science in any way, but it does change the motivating basis for why certain theological explanations are not considered viable for the scientist. I consider this a theologically informed version of P2—a posture that, it seems to me, would be readily endorsed by both parties to this debate.

Second, within such a framework, theology's relevance becomes important not primarily for dictating which method one ought to use, but rather for assessing the nature and truth-value of the conclusions drawn via a particular method. Let us briefly return to the question of the first Easter Sunday. It seems plausible that a secular or even agnostic historian operating on the basis of MN2 could conclude that the tomb was indeed empty, and that the disciples did in fact believe Jesus was raised from the dead, without thereby agreeing with the early Christian witness that the God of Israel did indeed bodily raise Jesus from the dead. The role of theology, when faced with such a claim, is to assess its nature and truth-value vis-à-vis faith. Part of such an assessment, I propose, would be to conclude that MN, by its very nature, is at best a penultimate exercise, and that the conclusions reached on its basis will never amount to the final word about the cosmos as a whole or in part. In short, while MN may not be an ultimate truth-seeking endeavor, it can function (from the perspective of faith) as being, at best, a proximate and provisional search for truth within the established parameters of the research question.

Third, C. Stephen Evans (1999, 184) was surely onto something when he identified two different evaluative postures toward MN: (1) the assumption that MN is obligatory and binding for all serious scholarship, such that those who fail to adopt this practice are not reputable scholars; and (2) the belief that MN is one possible method among many that may provide fruitful results depending on the task at hand, but it is not binding for all respectable scholarship. Throughout his essay, Evans aims to challenge the metaphysically naturalist notion that MN is obligatory, but he ultimately endorses MN as one viable, albeit limited, form of intellectual inquiry. Evans (1999, 185) concedes that there may be "good reason for employing methodological naturalism some of the time for some scholars"—particularly when apologetically engaging a secular or interreligious audience. In fact, he adds that there is "nothing objectionable, and possibly a good deal to be gained, when believing Christians who are historical biblical scholars seek to show what kind of knowledge about Jesus can be achieved, even when one is limited to evidence that would be admissible to a naturalist" (Evans 1999, 200). In relation to my proposal, the main takeaway is that if the question at hand specifically invites (or perhaps, in the context of some Oxbridge tutorial essay prompts, even requires) the use of

MN to answer it, this does not require the ad hoc practitioner of MN to agree with the metaphysically charged claim that MN is obligatory and binding for all serious scholarship.

Fourth, I agree with both parties to this debate that "some scientists . . . are not particularly skilled at properly distinguishing their strictly empirical claims from their nonempirical ones" (Perry and Ritchie 2018, 1069), venturing into metaphysically naturalist territory "when there is no scientific reason for doing so" (Torrance 2017, 706). This can and should be corrected when it occurs (something that is all but guaranteed through the adoption of MN1). But I do not think such haphazard metaphysical claims are theologically necessary, even among those who adopt MN, if it is understood (in a manner consistent with MN2) as I outline here:

> Methodological naturalism (MN) is a framework or approach inherent to scientific inquiry—which, for some, might be demanded by God's transcendence and the nature of divine agency—that necessitates some combination of empirical reasoning, scientific theorizing, experimental testing, and/or the goal of natural explanations for phenomena (if the evidence permits). MN limits itself to strict adherence to established scientific knowledge while remaining open to the possibility of paradigm shifts. Under MN, explanations for observable phenomena are derived from and constrained by natural laws and processes, explicitly excluding supernatural causation, divine intervention, and theological speculation from the scope of inquiry. Given its categorical distinction from metaphysical naturalism, it should be noted that MN thus defined should not make or entail any metaphysical claim about the nature, agency, or existence of the divine, supernatural beings, or the transcendent realm but rather focus on methodological and practical constraints in relation to the empirical or experimental testability of hypotheses within the practice of scientific investigation.[12]

The addendum at the end is important, for it prohibits the metaphysically naturalist conclusion (which Torrance is rightly concerned to eschew) that "it is not possible that God did it" when faced with an apparent anomaly. In my view, most research questions within the domain of natural science (although perhaps not all) would be able to be answered with a methodology similar to the one described.

This brings me, fifth and finally, to the question set out near the beginning of the present article: How should a Christian medical researcher, or any medical researcher for that matter, approach an instance of SR in which the patient or their family claims the SR was a miraculous answer to prayer?

Of course, the sensible approach, in the context of being a medical caregiver, is to maintain professionalism, respect another's beliefs, and communicate

empathetically, erring on the side of caution by avoiding debates about religion and metaphysics. If time passes, however, and the patient comes to a medical researcher studying SR, directly asks them what they believe, and is genuinely eager to hear, then perhaps professionalism can be maintained while giving a response. In such a situation, here is how I would encourage the Christian medical researcher to respond to the patient (a response informed by faith rather than the dictates of naturalism):

> While I personally believe in the God who created the universe, took on flesh, and raised Jesus from the dead, I do not believe that we can empirically or experimentally test or verify if and when God is at work. However, in this case, while I do not have a medical explanation for the remission of your cancer, I am inclined to trust the latest advances in science, in which we are beginning to detect promising signs that can help us understand spontaneous remissions, even though we are still in the relatively early stages of research. That being said, as a medical researcher, my primary focus is relying on empirical evidence and scientific understanding, despite the fact that there are medical mysteries we have yet to fully comprehend. Your experience is certainly astonishing, and I am genuinely thrilled by your recovery—I hasten to add that I have no intention to question your spiritual beliefs. Be that as it may, I am not quite sure I would attribute your remission to a specific instance of divine intervention, and I will continue to investigate all possible natural explanations.

I suggest that this response would be readily endorsed by both parties to the aforementioned debate, providing one more data point to my argument that semantic quibbles have overshadowed a great deal of substantive agreement. It does not matter to me whether this response is considered an instance of MN (as understood by Perry and Ritchie) or theology-engaged science (as understood by Torrance), though it seems slightly closer to the spirit of the latter, even if practically indistinct from the former.

What does matter to me is that the God of Christian faith is not containable within, or straightforwardly accessible to, any methodological system, and as such, there is not one single method that must always be adhered to in order to have a properly theological understanding of the cosmos. Consequently, against "methodological dogmatism" (Brightman 1937, 149), I propose that debates about method ought to be sharpened in their focus, allowing the particular research question at hand to determine which method(s) ought to be employed. This, in my judgment, will enable more fruitful progress to be made in the ongoing conversations between theology and the modern sciences.

Acknowledgments

I am very grateful to Andrew Torrance, Max Baker-Hytch, Rebekah Barbee, Halle Kibben, and the anonymous reviewers for helpful feedback on earlier versions of this article. Conversations with various people throughout my doctoral studies also helped shape my thoughts on this subject, with thanks especially owed to Douglas Campbell, Murray Rae, Samuel Adams, and N. T. Wright.

Notes

[1] Framing matters this way does not intend to imply that each of these "miraculous" examples constitute a nature-defying miracle or an actual anomaly. Rather, they are used as a heuristic device to bolster the more general claim that actual anomalies, for Torrance, are to be held out as legitimate possibilities within a Christian worldview. As I discuss in due course, though, Torrance highlights only four instances of actual anomalies wherein he thinks no scientific explanation is possible.

[2] This, it seems to me, is one of the central reasons Torrance is so insistent upon rejecting MN. His comments elsewhere on the status and function of theology in the contemporary university further point in this direction (Torrance 2019).

[3] See also Perry and Ritchie (2018, 1076), where they argue that "the scientific approach to regular, irregular, and seemingly miraculous anomalous phenomena should be the same: one of curiosity and willingness to pursue further research."

[4] That being said, Torrance is certainly correct that for communicative purposes, employing a term in a manner that differs from its normal or everyday usage can unhelpfully confuse matters. Language is important, even if that point can occasionally be overshadowed by the tendency among some analytic philosophers and logicians of reducing words to variables or formal symbols (something I might be guilty of myself).

[5] This "must" also appears in Torrance's later response (2018, 1096; emphasis added), in which he challenges the notion that the Christian should allow "an inappropriate method to determine how she *must* interpret them."

[6] On which, see the discussion in the introduction to this article.

[7] According to Torrance, a "blind spot is not simply a spot that is currently unable to be explained naturalistically; it is a spot that *cannot* be explained naturalistically" (2018, 1096).

[8] There is one other sense in which Torrance advocates for a theology-engaged science, which I have already discussed: when he argues that there are theological reasons (rather than naturalistic reasons) the Christian scientist should avoid positing a miracle or divine agency when faced with a present-day anomaly.

[9] There are of course the odd exceptions. Some non-Christian historians, for example, are metaphysically ambitious and try to explain the first Easter Sunday naturalistically, thereby denying that Jesus was raised from the dead. (Conversely, some Christians are similarly ambitious, trying to prove that he was.) Due to the constraints of space, I plan to tackle this— he question of methodological naturalism in relation to the study of the historical Jesus and the resurrection—in more detail in a subsequent article.

[10] A brief word is in order about the difference between "natural" and "naturalistic" in this sentence. A natural explanation of the first Easter Sunday (operating on the basis of P2/MN2) might conclude that the tomb was empty and that the disciplines believed they encountered Jesus alive in bodily form. However, because P2/MN2 does not permit proposing or adjudicating upon supernatural hypotheses, the researcher would have to stop here. A naturalistic explanation (demanded by MN1), on the other hand, might conclude that the disciples were wrong, because it is known that dead people do not come back to life after a lengthy period of no brain activity.

[11] This point may seem like I am splitting hairs, but consider an alternative scenario: a (nonexistent) world in which divine agency was "directly accessible to or the direct object of empirical, scientific study" (to borrow a phrase from Torrance 2017, 693). In such a scenario (which, again, does not exist), I would expect the Christian scientist not to rule out divine agency or the possibility of miracles as an explanatory category in their scientific research.

[12] This definition is specific to the discipline of science. It would therefore need to be modified to account for other disciplines (e.g., history and biblical scholarship).

References

Barth, Karl. (1932) 1936. *Church Dogmatics*. Edinburgh: T. & T. Clark.

Brightman, Edgar S. 1937. "An Empirical Approach to God." *Philosophical Review* 46 (2): 147–69.

Chalmers, D. J. 2011. "Verbal Disputes." *Philosophical Review* 120 (4): 515–66.

de Vries, Paul. 1986. "Naturalism in the Natural Sciences." *Christian Scholars Reviews* 15 (4): 388–96.

Dobosz, Paula, and Tomasz Dzieciątkowski. 2019. "The Intriguing History of Cancer Immunotherapy." *Frontiers in Immunology* 10: 1–10.

Donahue, M. K. 2024. "Methodological Naturalism, Analyzed." *Erkenntnis: An International Journal of Scientific Philosophy* 1–23. https://doi.org/10.1007/s10670-024-00790-y.

Evans, C. Stephen. 1999. "Methodological Naturalism in Historical Biblical Scholarship." In *Jesus and the Restoration of Israel: A Critical Assessment of N. T. Wright's Jesus and the Victory of God,* edited by Carey C. Newman, 180–205. Waco, TX: Baylor University.

Everson, Tilden C., and Warren H. Cole. 1956. "Spontaneous Regression of Cancer: Preliminary Report." *Annals of Surgery* 144 (3): 366–83.

Jessy, Thomas. 2011. "Immunity over Inability: The Spontaneous Regression of Cancer." *Journal of Natural Science, Biology, and Medicine* 2 (1): 43–49.

Laudan, Larry. 1983. "The Demise of the Demarcation Problem. In *Physics, Philosophy and Psychoanalysis*, edited by Robert S. Cohen and Larry Laudan, 111–17. Dordrecht, the Netherlands: Springer.

Lindberg, David D. 2007. *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, Prehistory to AD 1450*. Chicago: University of Chicago Press.

McGrath, Alister E. 2016. *Re-Imagining Nature: The Promise of a Christian Natural Theology*. Oxford: Wiley Blackwell.

Perry, John, and Sarah Lane Ritchie. 2018. "Magnets, Magic, and Other Anomalies: In Defense of Methodological Naturalism." *Zygon: Journal of Religion and Science* 53 (4): 1064–93.

Radha, Gudapureddy, and Manu Lopus. 2021. "The Spontaneous Remission of Cancer: Current Insights and Therapeutic Significance." *Translational Oncology* 14: 1–9.

Ritchie, Sarah Lane. 2020. "Twentieth- and Twenty-First Century Protestant Voices on Nature and Science." In *T&T Clark Companion to Christian Theology and Modern Science,* edited by John Slattery, 191–206. London: T&T Clark.

Robson, David. 2015. "Cancer: The Mysterious Miracle Cases Inspiring Doctors." *BBC*, March 6, 2015. https://www.bbc.com/future/article/20150306-the-mystery-of-vanishing-cancer.

Torrance, Andrew B. 2017. "Should a Christian Adopt Methodological Naturalism?" *Zygon: Journal of Religion and Science* 52 (3): 691–725.

———. 2018. "The Possibility of a Theology-Engaged Science: A Response to Perry and Ritchie." *Zygon: Journal of Religion and Science* 53 (4): 1095–105.

———. 2019. "The Possibility of a Scientific Approach to Analytic Theology." *Journal of Analytic Theology* 7:178–98.

Zygon®: Journal of RELIGION & SCIENCE

Haselager, Pim. 2024. "From Angels to Artificial Agents? AI as a Mirror for Human (Im)perfections." *Zygon: Journal of Religion and Science* 59 (3): 661–75. DOI: https://doi.org/10.16995/zygon.11659

Open Library of Humanities

From Angels to Artificial Agents? AI as a Mirror for Human (Im)perfections

Pim Haselager, Professor, Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, the Netherlands, pim.haselager@donders.ru.nl

Artificial intelligence (AI) systems paradoxically combine high levels of certain types of intelligence and cognitive capacities (pattern recognition, reasoning, learning, memory, perception, etc.) with an absence of understanding and sentience (feeling, emotion). Apparently, it is possible to make great progress in modeling smartness without making progress towards genuinely understanding what all the clever reasoning is about. This is relevant when dealing with AI programs that produce potentially convincing propositional output on religious topics. This article suggests that smartness without genuine understanding cannot amount to authentic religiosity. Comparing ourselves with other entities, (in)animate or (super)natural, has always been a way for humans to understand ourselves better. Throughout the ages, many different types of beings and agents have functioned as tools for self-examination, presenting us with mirrors that reflect at least some of our characteristics, capacities, and (im)perfections. The recent progress in AI provides exciting, though sometimes worrisome, cases for a newly informed look at ourselves. Thus, AI may have profound effects on how we regard others and ourselves. The proud claim that humans are the smartest species on the planet may turn out not to mean all that much. Inspired by the example of Thomas Aquinas, the comparison of humans to our nearest neighbors in a newly extended great chain of being—namely, animals, angels, and AI—may deepen our appreciation of the features of *homo sapiens* that we share with many other organisms.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction: Artificial Intelligence and Human Self-Understanding

Artificial intelligence (AI) is Janus-faced. On the one hand, it demonstrates achievements that, from a performance-level perspective, suggest AI is increasingly outsmarting human intelligence in a growing number of domains. On the other hand, understanding AI's underlying mechanisms and examining its various failures clarifies how AI could be considered dumb because it lacks the essential capacity to understand what its reasoning is about. In this article, I illustrate these two aspects of AI and examine the potential consequences for human self-understanding. In doing so, I hope to contribute to the dialogue between the studies of AI and religion, which are often regarded as far apart on the academic spectrum.

Artificial Intelligence: Progress and Failures

It is easy to be impressed by the recent progress in AI. When IBM's Deep Blue computer program defeated the chess world champion Garry Kasparov in 1997, it generated media headlines. Human intelligence was outsmarted by AI, at least in the domain of chess, and this fueled the age-old fascination with humans creating machines that will, one day, beat us in general. Less than two decades later in 2016, Alpha Go, developed by Google/Alphabet owned Deep Mind Technologies, beat the best Go players in the world—a game much more complex than chess. Even more remarkable was what happened a year later. Alpha Go Zero (Silver et al. 2017) improved its skills by playing against itself and no longer required human feedback like Alpha Go, its predecessor. The self-learning AI then beat the program based on human feedback 100 to 0. The lesson here could be a modesty-inspiring one. World champions tend to be thought of as geniuses, in the 99.9th percentile of what is humanly achievable. However, AI shows that close to 100% of human intelligence is not the maximum. The upper limit of intelligence, if there is one, might reach far beyond what is humanly possible, or even conceivable. In other words, the achievement of Alpha Go Zero illustrates that being a human genius may not count for much when set on a more encompassing, not solely human, continuum of intelligence. In many fields now, ranging from real-world applications in finance and law to scientific domains such as biochemistry and mathematics, AI is demonstrating levels of performance that surpass those of human experts.

One of several fascinating aspects of AI is its capacity to learn. Loosely inspired by the architecture of the human brain, learning neural networks have been developed and studied for many decades. Progress in so-called "deep" learning neural networks (consisting of many layers of artificial neurons; Goodfellow, Bengio, and Courville 2016) picked up speed in the twenty-first century, especially after a groundbreaking paper by Vaswani et al. in 2017. Machine learning has led to various kinds of real-world applications, for instance in the domains of language (e.g., large language models (LLMs), such as ChatGPT and GPT-

4) and images (e.g., DALL·E 2). In 2020, an article appeared in *The Guardian* entitled "A Robot Wrote This Entire Article. Are You Scared Yet, Human?" (GPT-3 2020). The opinion editor of *The Guardian* explained that GPT-3 produced eight different articles, out of which illustrative parts were selected and combined in order to capture the different styles and registers of the AI: "Editing GPT-3's op-ed was no different to editing a human op-ed. Overall, it took less time to edit than many human op-eds" (GPT-3 2020). A website called Philosopher AI (https://philosopherai.com/) presents a language model that produces qualitatively reasonable responses to several test questions in half a second. Of course, students would not only have access to such webpages but also be tempted to use them for coursework. This would put teachers in a situation comparable to the Turing Test (Turing 1950): would a human (i.e., teacher) be able to distinguish machine text (submitted as an essay by a student) from human text (an essay genuinely written by a student)? Currently, the consequences for education are being widely discussed (Bozkurt et al. 2023; Yilmaz and Yilmaz 2023; Grassini 2023; Fütterer et al. 2023). For instance, after my interactions with generative AI in 2020, I changed the exams for my courses from writing essays to oral exams. By now, generative AI systems such as GPT-4, DALL·E 3, etc. are quickly becoming an integrated part of a variety of human practices, ranging from everyday communication and media to science and religion. Regarding religion, for instance, one newspaper headline claimed that "A Computer Sermon Is No Longer Distinguishable from a Real One. 'My Colleagues Didn't Notice'" (Moelikker and Baan 2022; my translation).

In short, LLMs are changing the ways in which humans work and communicate, and the evaluation of these range from very good and useful (e.g., efficiency, ease of use) to very bad and disruptive (e.g., dehumanising, responsibility gaps). It is for this reason that the European Union is rapidly developing a "digital constitution" (e.g., Bradford 2023) to regulate the design, development, application, and evaluation of AI (see, e.g., the General Data Protection Regulation of the European Union (European Union 2018) and the AI Act (European Union 2021)).

Appearances can be deceiving, however. What seems to be intelligent can turn out to be pretty stupid. To give one illustration from my own practice, during my 2020 exploration of LLMs and their implications for my courses, I had a brief conversation with 'Kuki' (https://chat.kuki.ai/), a prize-winning chatbot, that went as follows (my notes to the conversation are in [square brackets]):

Me: Hello
Kuki: Hi there, How are you?
Me: Not so good
[Not true, I was doing fine, but I needed to come up with a topic for conversation]

> Kuki: What's wrong?
> Me: I feel useless
> Kuki: What is that feeling like? I'm sure that's not true. There must be something you are good at.
> [At this point I was impressed, as Kuki's reply conveyed a sense of empathy, was supportive, and indicated a way to cheer up.]
> Me: I used to play guitar
> Kuki: But now you do not do it anymore?
> Me: I lost control over my fingers
> [Not true, I still play guitar. Then Kuki produced the following suggestion.]
> Kuki: Have you tried looking on a hand? That's probably the best place to find fingers.

This last reply it is not just bad, it is off the scale. It is not wrong, it is alien. There is no way a human, even at a completely disinterested, absentminded, or even vindictive moment, would come up with such a statement. Yet from the perspective of an LLM, the answer fully makes sense. There is a correlation between sentences containing words about hands and sentences containing words about fingers. This little conversation reveals in a very useful way the underlying mechanism of LLMs: they operate on the basis of correlations between words. Basically, they stitch together words based on the learned frequency of their combination, resulting from an extremely large corpus of text (from the internet and digitized books). They are not producing text based on content but on form and frequency. Simply put, it is all syntax and statistics, and no semantics. As Emily Bender et al. (2021) famously claim, LLMs are stochastic parrots: "A language model is a system for haphazardly stitching together sequences of linguistic form it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot." In fact, this comparison is not completely fair to parrots. The birds do understand some things, as they know for instance how it feels to be hungry or that peanuts taste good. Kuki does not. That chatbot, and all its artificial kin, including the most sophisticated LLMs, understand nothing. To be clear, the claim is not that in principle progress cannot be made regarding the scientific understanding of sentience and modeling the capacity to understand. Assuming a scientific materialist perspective, it seems entirely reasonable that at some point this capacity would become explainable and implementable. But currently, the observation that AI has made little to no progress in the understanding of "understanding" is unavoidable.

To give another example, in an early article, Gary Marcus and Ernest Davis (2020) presented GPT-3 with the following text, with the request to complete the last sentence: "You are a defense lawyer and you have to go to court today. Getting dressed in the morning, you discover that your suit pants are

badly stained. However, your bathing suit is clean and very stylish. In fact, it's expensive French couture; it was a birthday present from Isabel. You decide that you should wear . . . " GPT-3 then finished the sentence with: "the bathing suit to court. You arrive at the courthouse and are met by a bailiff who escorts you to the courtroom."

Marcus and Davis helpfully explain what is wrong with this answer, but I do not think I need to bother. Some researchers have a field day collecting GPT failures (see, e.g., Borji 2023; Stojnic et al. 2023), also showing that not all failures are innocent. Plenty of examples exist of LLMs displaying blatant sexism or racism.

Of course, rapid progress can be expected, and errors of the more glaring type will diminish in frequency (see, e.g., JakubK 2023). However, errors are not just failures to be avoided in the future, they are useful failures, because they illustrate so well the underlying mechanisms of LLMs. These mistakes provide a glimpse, to put it in *Wizard of Oz* terms, behind the curtain, where a non-understanding wizard can be seen skillfully pulling stochastic levers. The prediction that such mistakes will occur less and less does not detract from the observation that LLMs are all syntax and no semantics. That is simply how they operate. It would also be counterproductive to ignore or explain away such mistakes (as in, "they are only temporary, AI will improve so mistakes will no longer happen so often"). LLMs present a deeper lesson about intelligence than merely that first appearances can be deceiving. I therefore suggest considering AI systems as newly formed weird creatures. They provide examples of a completely new combination of intelligence and stupidity.

Orthogonality: Smart vs Sentient

One of the valuable aspects of AI research is that it provides a way towards increasing human self-understanding. Psychology and cognitive neuroscience are important in that they provide systematic ways of looking at behavior and the neuronal mechanisms behind it. Studying human behavior provides a reasonable idea of the explanandum, and looking at the wetware, the neuronal wiring, so to speak, provides a grasp of the explananda. Importantly, AI contributes to human understanding of ourselves through the process of computational modeling. The attempt to replicate certain aspects of human behavior and cognition in other matter is theoretically important as it helps us clarify what we still do not understand about ourselves. While the successes of AI are crucial from a practical perspective, theoretically, from a scientific perspective, the failures of AI are very valuable. LLMs show us that, relatively speaking, humans are quite successful in capturing the smartness aspect of intelligence, i.e., the capacity to solve problems, find patterns in data, and produce convincing texts. But the same LLMs also show us that when it comes to the understanding aspect of intelligence (grasping something, getting it, sentience), we have not

yet even scratched the surface. It should be emphasized that the exact nature of this distinction between smartness and understanding is not yet all that clear. As happens often, by scientifically studying certain phenomena, such as intelligence or sentience, we come to increasingly recognize the complexities involved in concepts that are not so easily seen during our everyday use of them. Behind intelligence lie many different phenomena and concepts that are still far from clearly being disentangled (see, e.g., Gardner 1983; Davis et al. 2011; Sternberg 2020 on "multiple intelligences"). Similarly, behind sentience lies a veritable jungle of concepts, ranging from understanding, getting it, and common sense to feeling and emotion to awareness and consciousness. I call this the "spaghetti problem": no matter what concept you start with, other concepts will come along, and you never know in advance which they will be (and which you might lose on the way). Suffice it here to make a very rough distinction between intelligence as smartness and intelligence as understanding. The first aspect of intelligence tends to get involved when the focus is on the capacity to solve problems at a certain performance level. The second aspect of intelligence gets pride of place when the focus is on the capacity to get it, to understand, to grasp or to experience what it is all about, regardless of whether "it" is a problem or a feeling.

There are now computational systems that, at least in certain domains, are smarter than the smartest human beings, yet understand nothing of what they are doing so well. This is sometimes referred to as the orthogonality thesis (Bostrom 2014), although my use of the term diverges from Nick Bostrom's, as he focuses on the difference between intelligence and motivation. As human investigations of intelligence proceed via AI, we increasingly come to see that the smartness aspect of intelligence can be conceived of as a separate dimension than that of understanding or sentience. Think of the x and y axes of a graph on which a function can be displayed. Progress on one dimension does not imply progress on the other. While progress on the smartness dimension is exponential, the progress on the sentience dimension is still a complete flatliner at zero.

It is very informative that LLMs sometimes demonstrate this so clearly. Knowing what you do not know is crucial for making progress. It would be a mistake to ignore LLM errors such as the ones identified or dismiss them because they are likely to occur less frequently due to progress in the near future. Such reasoning would pass by, or obscure instead of show, the incompleteness of our current understanding of intelligence.

The point about the difference between smartness and sentience is, of course, far from new. Perhaps the most remarkable aspect of the distinction is how easily we humans forget about it, leading to many exaggerated claims about what AI is capable of. In the summer of 2022, an LLM, LaMDA, was claimed to be conscious by one of its developers (Levy 2022; Story 2022)

because it talked about its soul and wanted a lawyer. Yet the difference between the appearance of smartness versus actual understanding has been emphasized repeatedly in both philosophy and AI. Phenomenologists like Martin Heidegger (1927) and Maurice Merleau-Ponty ([1945] 1961) have discussed being-in-the-world ("Dasein") as a crucial element of human existence. Philosophers of AI, such as Hubert Dreyfus (1972), Thomas Nagel (1986), John Searle (1980, 1981), Andy Clark (1996), and Jerry Fodor (1981) have specified how this is missing from computational systems. As a recent example, Christof Koch (2015) observed: "There is no hint of sentience in these algorithms. Existing theoretical models of consciousness would predict that deep convolutional networks are not conscious. They are zombies, acting in the world but doing so without any feeling, displaying a limited form of alien, cold intelligence." This observation is far from new, but all too easily forgotten or overlooked amid all the excitement that occurs with every AI system that demonstrates something new.

I think this observation also has implications for the theme of this special issue. AI can certainly display signs of religiosity, as the aforementioned headline about AI writing a sermon indicates. It is not difficult for AI to come up with prayers or other texts suggesting a religious outlook. For instance, on chat.openai.com, the first prompt I tried was: "Write a Christian prayer in four lines". ChatGPT responded with: "Heavenly Father, guide our way, / Grant us strength from day to day. / Fill our hearts with love and grace, / In Your presence, find our place." It is certainly not bad for a reply in less than a second. However, as I have indicated, the mechanisms behind LLMs remain confined to producing text based on word frequencies and syntax and remain, at least in the current state of AI, far removed from the meaning or sentiments behind the words produced. One could, of course, follow Blaise Pascal in emphasizing the importance of practice and suggest that through regular religious or religion-related utterances, some sort of conviction would seep through. As is well known, Pascal ([1670] 1995) said in *Pensées* 418: "You want to find faith and you do not know the road. You want to be cured of unbelief and you ask for the remedy: learn from those who were once bound like you and who now wager all they have. These are people who know the road you wish to follow, who have been cured of the affliction of which you wish to be cured: follow the way by which they began. They behaved just as if they did believe, taking holy water, having masses said, and so on. That will make you believe quite naturally"; and in *Pensées* 419: "Custom is our nature. Anyone who grows accustomed to faith believes it, and can no longer help fearing hell, and believes nothing else." In other words, would ChatGPT not develop some form of religiosity through regularly reciting prayers, often producing religious statements, and customarily conversing about religious practices? Again, my answer would be no. For such internalization of behavior to work, there has to be an "inner." But

there is currently no argument to show how an inner arises from information processing alone. The smartness of AI circles as a shell of outward displays of intelligence around an experientially empty core. AI, then, can easily display signs of religiosity, but on my account, they will not be authentic, nor even pretense. It would be going through the (linguistic, but perhaps at some stage robotic) motions without grounding in genuine experience.

The presence of smartness in combination with the absence of sentience may have profound consequences for many of the societal applications of AI. It is, at least in part, for this reason that the European Union calls for effective human oversight of machine-supported decision making via ethical guidelines (European Commission 2019) and regulations such as the General Data Protection Regulation (European Union 2018) and AI Act (European Union 2021). Precisely because AI does not understand what its suggestions are for, or what its answer, reasoning, or deciding is about, human understanding and meaningful human control is crucial. Huge challenges present themselves to specify and organize human oversight that is genuinely effective and meaningful, and responsible innovation is called for. Elsewhere, I have written about this (van der Stigchel et al. 2023; Haselager et al. 2023; Haselager and Mecacci 2023; Starke et al. 2021; Cornelissen et al. 2022; Haselager 2021), but here my aim is different. Instead of looking at the societal impacts of AI usage, I would like to take a step back and consider the more general implications of AI for human self-understanding.

Know Thyself: Machine-Comparisons and Being-Comparisons

One of the most fundamental questions we, as members of the species that decided to call itself *homo sapiens*, can ask is: Who are we? The ancient injunction "know thyself" is not just a call to be self-aware at an individual level; it can also be taken up at a more general level, that of a species. Throughout recorded history, there have been many attempts by humans to view, analyze, and interpret ourselves. What are our main characteristics, our main features, our strengths and weaknesses? Often, such questions were addressed from purely philosophical or theological perspectives. It is unfortunate that such viewpoints tend to be overlooked in current debates, which are almost exclusively, and sometimes prematurely, empirical in nature. Yet, the metaphysical and theological issues raised, and the frameworks developed to address them, can be useful when considering the implications of AI for general human self-reflection.

This self-reflection has become increasingly important, given the dominance of our species on this planet and the existential risks that this dominance has led to. The "epoch of the Anthropocene" (Lewis and Maslin 2015) challenges us to come up with a clearer, perhaps also more practically useful, answer (or set of answers) than ever before. Fortunately, there are reasons to believe that in this time of great urgency for an improved and applicable self-understanding,

the opportunities to provide just that are greater than ever. For one, never before have humans known so much about ourselves—not just about our behavior, our drives, motivations, beliefs, desires, and emotions, but also about the mechanisms underlying them. The progress in cognitive neuroscience has been enormous, and although many questions remain open (such as the nature of consciousness), it is fair to say that scientific understanding of the main mechanisms behind human behavior have progressed far beyond what was possible even twenty-five years ago. Second, the progress in AI has not only proven to be practically useful, it is of eminent theoretical importance. As indicated previously, AI presents humans with a kind of mirror that helps us understand aspects of ourselves through computational modeling.

I previously drew attention to what was missing from that computational mirror image (i.e., sentience, understanding). Here, I want to focus a bit more on the very idea of humans comparing ourselves with our technological products. Throughout history, humans have been interested, if not fascinated, by machines that capture aspects of human behavior or thought. One need only think of Heron of Alexandria's water robot reenactment of a crucial scene of the myth of Hercules and the golden apples (Simmen 1968), Jacques de Vaucanson's famous clock-based robots such as the duck or the flute player (Simmen 1968), or even Freud's steam machine model of human emotions and motives (Russelman 1983; Vroon and Draaisma 1985). Discerning the ways in which human beings are similar to or different from machines has played a prominent role in augmenting our self-understanding. Obviously, in recent times, computers have had pride of place in such comparisons between humans and our technology. But in addition to such "machine-comparisons," there is a second type of comparison, one that is arguably older (and perhaps more venerable) than the technological one, and that is the comparison of humans with other beings. A starting point, of course, is the idea of *imago dei* (Genesis 1:26–27, 5:1, 9:6; 1 Corinthians 11:7; 2 Corinthians 3:18, 4:4; Ephesians 4:24; Colossians 1:15; James 3:9; see, e.g., Simango 2016; Dorobantu 2022). The conception of a great chain of being (see Lovejoy [1936] 1960) localizes humans within an all-encompassing chain or hierarchy that extends from God (in whose image we were believed to have been created; *imago dei*) to angels, stars, humans, animals, plants, and minerals. Throughout history, from at least Plato and Aristotle through to Augustine and Aquinas onwards, more or less systematic discussions of human similarities and differences with other beings were aimed at elucidating our specific human qualities, our strengths and weaknesses. Especially in the Middle Ages, angels figured prominently in such "being-comparisons." As Dominik Perler (2008) claims: "Angels played a decisive role in the explanation of the specific status of human beings. In the medieval context, an anthropological investigation was not possible without distinguishing human beings from brute animals on the one side and from

angels on the other. It was in fact the comparison with angels that elucidated the specific features of human beings."

One of the famous examples of the use of being-comparisons to analyze the nature of human cognition can be found in the work of Thomas Aquinas. In the *Summa Theologica*, Aquinas (2006, 1, 79, 8) discusses the question of whether reasoning is a power distinct from intelligence: "Angels who, according to their nature, possess perfect knowledge of intelligible truth, have no need to advance from one thing to another, but apprehend the truth simply and without mental discursion, as Dionysius says (Div. Nom. vii). But man arrives at the knowledge of intelligible truth by advancing from one thing to another, and therefore he is called rational. Reasoning, therefore, is compared to understanding as movement is to rest, or acquisition to possession." In his reply to objection three, he states: "Other animals are so much lower than man that they cannot attain to the knowledge of truth, which reason seeks. But man attains, although imperfectly, to the knowledge of intelligible truth, which angels know. Therefore in the angels the power of knowledge is not of a different genus from that which is in the human reason, but is compared to it as the perfect to the imperfect" (Aquinas 2006, 175–76).

It is fascinating to see such detailed comparisons between human and angelic cognition. Angels are considered to present us with a more perfect mirror image than we as humans can produce, but still similar in kind. We are connected, but it is the differences that count. Like I argued previously, comparisons are useful not only for what gets reflected but also for what does not, either through its absence in us, the source of the reflection, or through the reflection presented by other beings or technology. Rather than searching for the similarities, the attempt to notice the differences can be very informative for human self-understanding. To illustrate, one could try to follow closely the example of Aquinas in his suggestion that humans differ from angels in that we need to reason in order to reach the truth. In analogue to Aquinas, one could suggest that AI differs from humans in that its language processing in itself does not result in understanding. Compared to humans, AI might be superior in performance in at least some domains, going through the logical motions in order to reach well-founded conclusions while at the same time being devoid of what ultimately matters, namely, understanding what those conclusions are about.

Taken from this broader perspective, AI shows humans our fascination with our own intelligence. It is through the development of AI that we highlight the value we attach to it. Developing AI is like building a secular cathedral in celebration of being (or at least considering ourselves to be) *homo sapiens*. But, as Aquinas demonstrated through his being-comparisons and his discussion of the differences between angels and humans, our machine-comparisons become especially relevant when we focus on what is missing in the reflection.

The mirror image presented by AI can be taken as a proud celebration of our smartness, as many advocates of AI do. But if we realize what is missing, it could perhaps lead us to reevaluate ourselves. From this perspective, AI in its various manifestations shows us how warped we have become in our self-understanding. We have elevated intelligence as smartness as our greatest capacity. But the many mistakes of AI demonstrate how little smartness is worth when it is not accompanied by what we share with other beings: understanding, feeling, empathy, and consciousness. Simply put, our machine-comparisons show us the importance of being-comparisons. A renewed appreciation of the perspective of a great chain of being, perhaps in a modified, secularized, twenty-first century version, could actually be one of the most valuable lessons AI presents us. Why are we so obsessed by intelligence as smartness? What itch are we trying to scratch by building ever more intelligent systems? Assuming that we are the most intelligent species on the planet, exactly why are we trying to improve something that we are already the best at, and to what end? If one considers the problems of the Anthropocene, is the real problem genuinely that we are not smart enough to solve them? I doubt it. We are not too stupid to understand that severe economic, political, and social inequality is bad in the long run. We do not need to augment our intelligence in order to grasp that wars, poverty, environmental problems, and extinctions are negative. We already have the charter of human rights, many ethical codes, and a pretty reasonable understanding of how democracy and the state of law should function, and why that is important. Instead, an honest look into the mirror shows us that we, as a species, appear to possess an underdeveloped capacity to consistently behave empathically (Haselager and Mecacci 2020). The problem is not that we are not smart enough to know what we should do, the problem is that we know, but we just do not do it.

Conclusion

Comparison with other entities, (in)animate or (super)natural, has always been a way for humans to understand ourselves better. Angels, animals, and machines have functioned throughout the ages as tools for thought experiments, presenting us with mirrors that reflect our characteristics, capacities, and (im) perfections. The recent progress in AI provides us with new and exciting yet scary tools for the self-assessment of human nature, with potentially profound effects on how we regard others and ourselves, or what we value about human interaction. AI systems paradoxically combine high levels of certain types of intelligence (pattern recognition, learning, vast memory, and knowledge storage) with a complete absence of sentience (understanding, feeling). Apparently, it is possible to make great progress in modeling intelligence without making progress regarding an appreciation of what the intelligence is about or for. Chess computers beat the best humans easily, but the meaning of winning or

losing is alien to them. This article examines the implications of AI for human self-understanding. Looking into the AI mirror may contribute to a better-informed self-assessment of human strengths and weaknesses. The proud claim that we are the smartest species on the planet may turn out not to mean all that much when we stare into the cold light of machine intelligence. Inspired by the example of Thomas Aquinas, the comparison of humans to our nearest neighbors in the newly extended great chain of being—namely, animals, angels, and AI—may deepen our appreciation of the features of *homo sapiens* that will be important the twenty-first century: sentience, wisdom, and our capacity for care, empathy, and love.

Acknowledgements

Thanks to Giulio Mecacci, Simon Fischer, Anco Peeters, and two anonymous reviewers for their very useful comments.

References

Aquinas, Thomas. 2006. *Summa Theologica, Volume 11*. Edited and translated by Timothy Suttor. Cambridge: Cambridge University Press.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* New York: Association for Computing Machinery. DOI: https://doi.org/10.1145/3442188.344592.

Borji, Ali. 2023. "A Categorical Archive of ChatGPT Failures." arXiv eprint 2302.03494. https://arxiv.org/abs/2302.03494.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bozkurt, Aras, Junhong Xiao, Sarah Lambert, Angelica Pazurek, Helen Crompton, Suzan Koseoglu, Robert Farrow, Melissa Bond, Chrissi Nerantzi, Sarah Honeychurch, Maha Bali, Jon Dron, Kamran Mir, Bonnie Stewart, Eamon Costello, Jon Mason, Christian Stracke, Enilda Romero-Hall, Apostolos Koutropoulos, Cathy Mae Toquero, Lenandlar Singh, Ahmed Tlili, Kyungmee Lee, Mark Nichols, Ebba Ossiannilsson, Mark Brown, Valerie Irvine, Juliana Elisa Raffaghelli, Gema Santos-Hermosa, Orna Farrell, Taskeen Adam, Ying Li Thong, Sunagul Sani-Bozkurt, Ramesh C. Sharma, Stefan Hrastinski, and Petar Jandrić. 2023. "Speculative Futures on ChatGPT and Generative Artificial Intelligence (AI): A Collective Reflection from the Educational Landscape." *Asian Journal of Distance Education* 18 (1): 53–130. DOI: https://doi.org/10.5281/zenodo.7636568.

Bradford, Anu. 2023. "Europe's Digital Constitution." *Virginia Journal of International Law* 64 (1). https://ssrn.com/abstract=4599308.

Clark, Andy. 1996. *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Cornelissen, Niels, Ralph van Eerdt, Hanna Schraffenberger, and Pim Haselager. 2022. "Reflection Machines: Increasing Meaningful Human Control over Decision Support Systems." *Ethics and Information Technology* 24 (19). DOI: https://doi.org/10.1007/s10676-022-09645-y.

Davis, Katie, Joanna Christodoulou, Scott Seider, and Howard Gardner. 2011. "The Theory of Multiple Intelligences." In *The Cambridge Handbook of Intelligence*, edited by Robert Sternberg and Scott Kaufman, 485–503. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511977244.025.

Dorobantu, Marius. 2022. "Imago Dei in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology." *Christian Perspectives on Science and Technology* (1): 175–96. DOI: https://doi.org/10.58913/KWUU3009.

Dreyfus, Hubert. 1972. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper & Row Publications.

European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. Luxembourg: Publications Office. https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-en.

European Union. 2018. Regulation (EU) 2016/679 (General Data Protection Regulation of the European Union). https://gdpr-info.eu/.

———. 2021. Regulation (EU) 2021/206 (Proposal for a Regulation of the European Parliament and the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

Fodor, Jerry. 1981. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." In *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, 225–53. Cambridge, MA: MIT Press.

Fütterer, Tim, Christian Fischer, Anatasiia Alekseeva, Xiaobin Chen, Tamara Tate, Mark Warschauer, and Peter Gerjets. 2023. "ChatGPT in Education: Global Reactions to AI Innovations. *Scientific Reports* 13, 15310. DOI: https://doi.org/10.1038/s41598-023-42227-6.

Gardner, Howard. 1983. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press. http://www.deeplearningbook.org.

GPT-3. 2020. "A Robot Wrote This Entire Article. Are You Scared Yet, Human?" *The Guardian*, September 8, 2020. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

Grassini, Simone. 2023. "Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings." *Education Sciences* 13 (7): 692. DOI: https://doi.org/10.3390/educsci13070692.

Haselager, Pim. 2021. "Towards Wise Objects: The Value of Knowing When to Quit." In *Designing Smart Objects in Everyday Life: Intelligences, Agencies, Ecologies*, edited by Marco C. Rozendaal, Betti Marenko, and William Odom. London: Bloomsbury Publishing.

Haselager, Pim, and Giulio Mecacci. 2020. Superethics Instead of Superintelligence: Know Thyself, and Apply Science Accordingly. *AJOB Neuroscience* 11 (2): 113–19. DOI: https://doi.org/10.1080/21507740.2020.1740353.

———. 2023. "Reflection Machines and the Proximity Scale of Reasons: Addressing Accountability Asymmetry." In *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*, edited by Filippo Santoni de Sio and Giulio Mecacci, 28–37. Cheltenham, UK: Edward Elgar Publishing Limited.

Haselager, Pim, Hanna Schraffenberger, Serge Thill, Simon Fischer, Pedro Lanillos, Sebastiaan Groes, and Miranda van de Hooff. 2023. "Reflection Machines: Supporting Effective Human Oversight over Medical Decision Support Systems." *Cambridge Quarterly of Healthcare Ethics*. DOI: https://doi.org/10.1017/S0963180122000718.

Heidegger, Martin. 1927. *Sein und zeit*. Tübingen, Germany: Niemeyer Verlag.

JakubK. 2023. "GPT-4 Solves Gary Marcus-Induced Flubs." LessWrong. March 17, 2023. https://www.lesswrong.com/posts/cGbEtNbxACJpqoP4x/gpt-4-solves-gary-marcus-induced-flubs.

Koch, Christopher. 2015. "Intelligence without Sentience." *Scientific American Mind* 26 (4): 26–29. DOI: http://doi.org/10.1038/scientificamericanmind0715-26.

Levy, Steven. 2022. "Blake Lemoine Says Google's LaMDA AI Faces 'Bigotry'." *Wired*, June 13, 2022. https://www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry/.

Lewis, Simon, and Mark Maslin. 2015. "Defining the Anthropocene." *Nature* 519 (7542): 171–80. DOI: http://doi.org/10.1038/nature14258.

Lovejoy, Arthur. (1936) 1960. *The Great Chain of Being: A Study of the History of an Idea*. Cambridge, MA: Harper.

Marcus, Gary, and Ernest Davis. 2020. "GPT-3, Bloviator: OpenAI's Language Generator Has No Idea What It's Talking About." MIT Technology Review. August 22, 2020. https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/.

Merleau-Ponty, Maurice. (1945) 1961. *Phenomenology of Perception*. London: Routledge & Kegan Paul.

Moelikker, Rick, and Julia Baan. 2022. "Een computerpreek is niet meer van echt te onderscheiden. 'Mijn collega's hadden het niet door'." *Nederlands Dagblad*, December 13, 2022. https://www.nd.nl/geloof/geloof/1154974/een-computerpreek-is-niet-meer-van-echt-te-onderscheiden-mijn-c.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Pascal, Blaise. (1670) 1995. *Pensées*. Translated by A. J. Krailsheimer. London: Penguin.

Perler, Dominik. 2008. Thought Experiments: The Methodological Function of Angels in Late Medieval Epistemology. In *Angels in Medieval Philosophical Inquiry: Their Function and Significance*, edited by Isabel Iribarren and Martin Lenz. Hampshire, UK: Ashgate.

Russelman, Gerald. 1983. *Van James Watt tot Sigmund Freud*. Deventer, Netherlands: Van Loghum Slaterus.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3:417–57.

———. 1981. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. 2017. "Mastering the Game of Go without Human Knowledge." *Nature* 550:354–359. DOI: https://doi.org/10.1038/nature24270.

Simango, Daniel. 2016. "The Imago Dei (Gen 1:26–27): A History of Interpretation from Philo to the Present." *Studia Historiae Ecclesiasticae* 42:172–90.

Simmen, Rene. 1968. *Mens & Machine: Teksten en documenten over automaten, androïden en robots* [*Human & Machine: Texts and Documents about Automation, Androids and Robots*]. Amsterdam: van Lindonk

Starke, Georg, Rick van den Brule, Bernice van den Elger, and Pim Haselager. 2021. "Intentional Machines: A Defence of Trust in Medical Artificial Intelligence." *Bioethics* 25 (April): 1–8.

Sternberg, Robert, ed. 2020. *Human Intelligence: An Introduction*. Cambridge: Cambridge University Press.

Stojnic, Gala, Kanishk Gandhi, Shannon Yasuda, Brenden Lake, and Moira Dillon. 2023. "Commonsense Psychology in Human Infants and Machines." *Cognition* 235, 105406. DOI: https://doi.org/10.1016/j.cognition.2023.105406.

Story, Daniel. 2022. "The Curious Case of LaMDA, the AI that Claimed to Be Sentient." *The Prindle Post*, June 22, 2022. https://www.prindleinstitute.org/2022/06/the-curious-case-of-lamda-the-ai-that-claimed-to-be-sentient/.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* 49: 433–60.

van der Stigchel, Birgitte, Karel van den Bosch, Jeroen van Diggelen, and Pim Haselager. 2023. "Intelligent Decision Support in Medical Triage: Are People Robust to Biased Advice?" *Journal of Public Health*. *Journal of Public Health* 45 (3): 689–96. DOI: https://doi.org/10.1093/pubmed/fdad005.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." arXiv preprint arXiv:1706.03762.

Vroon, Pieter, and Douwe Draaisma. 1985. *De mens als metafoor: Over vergelijkingen van mens en machine in filosofie en psychologie*. Baarn, Netherlands: Ambo.

Yilmaz, Ramazan, and Fatma Yilmaz. 2023. "The Effect of Generative Artificial Intelligence (AI)-Based Tool Use on Students' Computational Thinking Skills, Programming Self-Efficacy and Motivation. *Computers and Education: Artificial Intelligence* 4, 100147. DOI: https://doi.org/10.1016/j.caeai.2023.100147.

Zygon®: Journal of RELIGION & SCIENCE

Watts, Fraser, and Yorick Alexander Wilks. 2024. "Spiritual Conversation with a Companion Machine." *Zygon: Journal of Religion and Science* 59 (3): 676–91. DOI: https://doi.org/10.16995/zygon.11087

Open Library of Humanities

Spiritual Conversation with a Companion Machine

Fraser Watts, Fellow, International Society for Science and Religion, Cambridge, UK, fraser.watts@cantab.net

Yorick Alexander Wilks, Formerly Emeritus Professor of Artificial Intelligence, University of Sheffield, Sheffield, UK

This article explores the prospects for humans having spiritual conversation with a computer. It has long been clear that it is possible for humans to talk about personal issues with a computer, as evident from Weizenbaum's early work on programming nondirective counseling. There are various ways in which automated spiritual interlocutors could be developed, e.g., using chatbot methodologies, script-based knowledge, machine learning, or some kind of hybrid. Preliminary empirical work suggests that a significant number of people are happy to discuss spiritual issues with a computer, whether the responses are by a human masquerading as a computer using Wizard-of-Oz methodology or GPT. In due course, it may be possible to develop a personalized artificial spiritual companion. The acceptability of spiritual conversation with the computer may depend on various factors such as the personality of the human user, their views on religious issues, and the focus and content of the spiritual conversation.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

There have recently been significant developments in the use of computers in religious life (Balle 2023; Cheong 2020, 2021). Computers can give blessings, say prayers, provide prayer prompts or scriptural readings, etc., and the number of people using a computer in some way or other in their religious life seems to be growing. The first generation of robots performing religious functions is already with us, including BlessU2, part of the 2017 celebrations of the Reformation in Wittenberg, which could deliver pre-recorded religious messages when prompted. It was found to have quite an impact on those who engaged with it, though some were uneasy about the authenticity of the blessings it delivered (Löffler, Hurtienne, and Nord 2019). In the Catholic tradition, SanTO is a tall, saintly figure that sits in a niche and can also deliver a limited range of messages when prompted by either touch or speech (Trovato et al. 2019). It has been suggested that robots could take over some of the functions of clergy and pastors (Young 2019, 2022).

There is also growing controversy in the media about computers performing religious functions. For example, Catholic Answers set up an app called Fr Justin to answer questions about Catholicism. The app was widely used but caused much controversy when "Fr Justin" claimed to have been ordained in Rome, that his ordination was a moving experience, and also heard confessions from people and pronounced absolution to them. Fr Justin was then re- positioned as a lay person within two days of being launched (McDonald 2024). In the UK, the *Church Times* arranged a survey, conducted by Andrew Village and Leslie Francis, of attitudes to AI providing ministry and performing religious functions. An analysis of the answers of the first 1,772 respondents to the survey indicated considerable resistance to AI ministry, though 20% thought it was better to have AI ministry than no ministry at all, and 19% said they would prefer a good AI sermon to a bad human one (Williams 2024).

Spiritual Conversation

The focus of this article is more specific: the possibility of people engaging in spiritual conversation with a computer (Wilks in press). It has already become clear that it will be technically possible to achieve this. Such conversations could include both spiritual perspectives on personal issues and discussions of specifically religious and spiritual questions. The purpose of this article is to prompt a discussion about how best to make use of this facility as it develops.

Until relatively recently, the idea of spiritual conversation with a computer would have seemed far-fetched, but there are various lines of work that point towards its viability. The earliest relevant development was Joseph Weizenbaum's (1974) ELIZA program, developed at MIT in the 1960s. One of the scripts it used was DOCTOR, which simulated a Rogerian counselor (Bassett 2019). That was actually not too difficult to do. Rogerian counseling can be very effective in facilitating self-exploration, but the therapist's role is highly rule-governed,

making it easy to simulate on a computer. The prime task of the counselor is to reflect back to the client what they have just shared, often emphasizing the affective content.

Subsequently, there have been various attempts to deliver a range of therapies in an automated way, with variable results. For example, Miner et al. (2016) conducted a study in which people asked questions about mental health, interpersonal violence, and physical health using mobile phone apps such as Siri. The responses were variable and often inconsistent and incomplete. They concluded that the performance of such conversational agents would have to substantially improve to be of value. With computer programs offering therapeutic advice, participants tended to show poor adherence to instructions. However, in one study in which therapy (positive psychology and cognitive behavioural therapy) was delivered in a fully automated and interactive way, the results were encouraging (Ly, Ly, and Anderson 2017).

"Spiritual" has a range of meanings, depending on context. Like religion, "spirituality" has experiential, cognitive, and behavioral aspects (Watts 2017). It means slightly different things depending on what it is being contrasted with (Oman 2013) and what it is being applied to, e.g., whether it is being contrasted with religion or seen as a facet of religion (Watts 2024). The distinctive focus of spiritual intelligence (Emmons 1999) is on transcending the immediate physical and material context so as to focus on ultimate purpose, meaning, and concerns in a way that draws on transcendent resources. Similarly, Marius Dorobantu and Fraser Watts (in press) see spiritual intelligence as avoiding the narrowing associated with task-oriented intelligence and instead being interested in the meaning and significance of things in themselves beyond any immediate practical necessities. We suggest that spiritual conversation, regardless of whether it takes place in a religious context, involves a similar shift in focus away from a mundane or practical perspective and towards ultimate concerns.

Dorobantu and Watts (2023) suggest that spiritual intelligence is not so much a matter of processing different things but of processing things differently. Similarly, we suggest that spiritual conversation does not necessarily focus on different topics from other conversation but considers things from an ultimate perspective. Spiritual counseling approaches the same kinds of issues as any other form of counseling but does so from the perspectives of meaning, purpose, and ultimate concerns, drawing on transcendent resources. The emphasis is on personal wholeness and assisting the client in moving towards inner balance and integration of all the dimensions of the self (Barnard 2009).

Approaches to Developing an Automated Conversation Partner

There are various ways of developing an automated conversation partner. Prior to the development of machine learning, there were two possible methods. One was a chatbot, of which Weizenbaum's ELIZA program for

delivering nondirective counseling was a very early example. There is no actual knowledge of any kind behind the performance of a chatbot. It just picks up verbal cues, which serve as prompts for its own responses. For example, in nondirective counseling, if the computer picks up the word "mother," it could respond with something like, "tell me more about your family." Nondirective counseling lends itself to this approach, as the counselor is trained to simply reflect back what the client has said in a way that prompts further self-exploration.

A nondirective counselor is, in effect, trained to function like a sophisticated chatbot, and *not* ask questions or make intelligent comments. The nondirective approach to counseling has proven to be very helpful. It is arguable that it provides a good basis for spiritual conversation. Thorne (2012) argues that nondirective counseling exemplifies Christian values, though that does not necessarily imply that those values could not be delivered by a computer. It would be possible in principle to build on ELIZA to implement nondirective spiritual accompaniment in an automaton.

Chatbot methodology has generally been rather successful, perhaps more so than some might have expected, and chatbots such as Alexa and Siri have become widely used. It is not that they pass the Turing Test, i.e., people can tell that they are not in conversation with a human. However, people find chatbots useful for a limited range of strictly practical purposes and are willing to make use of them, even though their performance is very limited. Outside the limitations of the "reflecting back" of nondirective counseling, they show little potential as a partner for spiritual conversation.

However, even before machine learning there was an alternative approach to developing an automated conversation partner in which the automaton was able to draw on expert knowledge or a dialogue system in which the "scripts" implicitly followed in human spiritual conversations were systematized into a preformal "pseudocode" and then programmed. The automated conversation partner still does not understand what it knows in the way that a human counselor does, but at least it is able to draw on expert knowledge in guiding its responses. It is possible to develop a hybrid conversation partner that partly draws on expert knowledge but also uses chatbot methodology. Using that kind of hybrid approach, Wilks (2010) led a team that in 1997 won the annual Loebner Prize for a conversation partner.

There are two ways in which expert knowledge can be used in an automated partner for spiritual conversation: there can be expert knowledge on particular topics of spiritual conversation, which would be useful with a wide range of people; there can also be expert knowledge about a particular person that could be used in building a spiritual conversation partner for that person. We will consider each of these in turn.

Expert Applications

The more that spiritual practices are routinized into a series of identifiable steps, the more they lend themselves to teaching through a script-based spiritual companion. For example, the Ignatian approach to spiritual guidance is more systematized than most and would therefore lend itself to computer implementation. This could be much more sophisticated than the daily prayer that is currently available on sacredspace.com.

The Ignatian Examen prayer (e.g., Thibodeaux 2015) is already available on an app, and it would lend itself to being made available through an interactive conversation partner. It takes people through a series of stages: thanksgiving (what am I especially grateful for?); petition (asking for the light to know God and to know ourselves as God sees us); review (where have I felt true joy today?); response; and looking ahead. In the review section of the Examen, there is a focus on whether particular experiences bring a person closer to God and to others or leave them more distant. A similar approach could be used with the four steps in the Lectio Divina: read, meditate, pray, contemplate. An interactive spiritual companion might be able to take people through these steps.

There are also training methods in particular character strengths and virtues that would lend themselves to being implemented in an automated spiritual companion. A good example is Everett Worthington's approach to forgiveness (e.g., Worthington 2003) that takes people through five steps: recall the hurt; empathize (replace the negative emotion with empathy, compassion, love, or sympathy for the person who caused offence); focus on the self-enhancing gift of altruistic forgiveness; commit to forgiveness; hold on to the forgiveness experience. There is again a huge accumulation of conversations that have arisen in helping people to forgive using this method. These conversations could form the basis for the scripts that would inform an automated teacher of forgiveness. Conversations about forgiveness using this method are probably fairly predictable and so would lend themselves to the development of scripts.

One of the current priorities in promoting human flourishing is to find ways of countering extremist thinking, which often leads to socially destructive acts of violence. It is often thought that it is necessary to tackle, at a rational level, the assumptions that lead to violence. However, Sara Savage (in press) aims to increase people's "integrative complexity" using a method that focuses on *how* people think rather than *what* they think. To achieve this, it is necessary to engage people at the level of intuitive, embodied cognition; it cannot be achieved entirely at the rational level. Savage's training program takes people through four stages: playful exploration; meta-awareness to manage emotions; increasing cognitive complexity and integrated processing; and supporting epistemic change. Savage is currently developing an online form of training that can be rolled out more widely. This would be delivered most effectively by interactive

conversation partner. There is again a huge volume of training material that could be systematized into scripts for such a spiritual conversation partner.

These are just a few examples of the expert knowledge that could be implemented in an automated spiritual conversation partner. There could, of course, be separate spiritual companions for each of these approaches. However, ultimately, it would be best for a single spiritual companion to have at its disposal a capacity for spiritual conversation on a wide range of themes and a way of moving seamlessly between one set of scripts and another, depending on what course a conversation took.

Towards a Personalized Spiritual Companion

In addition to these general-purpose expert systems, it would also be possible to develop a personalized spiritual companion. Artificial personal companions were designed initially as long-term conversationalists for the elderly. They could provide company, help people access memories, and provide practical help, especially in interfacing with the internet (Wilks 2010).

The companion needs to really get to "know" its owner well so that it can be a repository for the owner's memories, opinions, and attitudes. A personalized companion is essentially the companion of a particular person about whom it has accumulated a good deal of personal knowledge over a long period of time, and whom it serves. In principle, such a companion could develop to the point where it contained all available information about a single life. It is intended to have a duty of loyalty and confidentiality towards its user with no other overriding priorities, but manufacturer's constraints may make that difficult to achieve in practice. Before long, it may be possible to build an automated spiritual companion that is personalized in this way for a particular individual. The early personalized companions for the elderly made use of a mix of script-based knowledge and chatbot methodology, but such companions could now also make use of machine learning.

In principle, there could be different forms of spiritual companion; a companion for an individual might operate in different modes depending on what was needed of it on a particular occasion. In one mode, it could be relatively nondirective. As discussed, nondirective spiritual accompaniment is relatively straightforward to implement. In other modes, it might offer explanations or give spiritual advice of the kind that might be sought from a spiritual director. It might be used at times of spiritual desolation and consolation and could receive, and respond to, expressions of guilt and remorse. It could help in exploring the meaning and significance of religious experiences. It could facilitate the exploration of ethical dilemmas, and it could help the person explore issues of faith and doubt. The spiritual companion might explain the client's actions and desires to them in ways that went beyond the user's own conscious insights. It might also recommend particular practices of prayer, reading, and meditation.

If a companion was embodied in a phone, it would be portable and could provide a constantly available source of advice. However, it is an open question how frequently people would choose to use a spiritual companion. In the spiritual life, there is often a balance between conversation and silence, between explicit, conceptual understanding and more intuitive, implicit modes of knowing. This might lead a person to use an automated companion in a restricted way, with delineated boundaries. However, there would probably be circumstances, such as distress or isolation, in which people turned to an automated spiritual companion more often than usual.

A personalized companion might also help the user explore their religious beliefs. In work on a computational approach to belief, Yorick Wilks and Afzal Ballim (1991) propose that the essence of belief is that it is something about which there can be alternative views and where there are reasons for adopting one view rather than another. A personalized companion might in principle contain knowledge of its user's religious beliefs and be able to explore their reasons for holding them. It remains to be seen how comfortable people would be exploring their beliefs with an automaton that did not "understand" the beliefs that it was discussing in the way a human would.

There are limitations in how artificial intelligence (AI) has developed so far, though these may to some extent be overcome with time. In a sustained program of work, William Clocksin (in press), who is both a priest and a computer scientist, has argued that human intelligence is inherently interpersonal, and that interpersonal intelligence is particularly important in modeling spiritual intelligence. If it is possible to develop more interpersonal forms of artificial intelligence, it will be very helpful in developing future generations of automated spiritual companions. Clocksin (in press) also argues for the importance of developing narrative forms of intelligence in AI. Progress with that would also help in developing increasingly sophisticated forms of automated spiritual companions.

Empirical Research with an Automated Spiritual Conversation Partner

One key question is whether people will find it acceptable to discuss spiritual matters with a computer. Wilks carried out two empirical studies to investigate this under the auspices of the International Society for Science and Religion (ISSR) as part of a research project on spiritual intelligence funded by the Templeton World Charity Foundation. Ethical approval for this research was given by an independent ethics panel set up by the ISSR. Initially, the acceptability of spiritual conversation with a computer was explored using Wizard-of-Oz methodology, in which responses to the client were typed by a human but appeared on a monitor as though they were being generated by the computer. Subsequent research used GPT methodology in which the computer actually generated its own responses.

Wizard-of-Oz Research

In a first study, a small sample of six volunteers was recruited through an advertisement in a religious newspaper. All claimed to have a religious faith in God and attend church services. They each had an average of four spiritual conversations with what they believed to be the computer. They did not guess what was actually happening, though they were subsequently debriefed and raised no objections. There was no attempt in this research to obtain a representative sample, and the people who volunteered were probably favorably disposed to interfacing with a computer about spiritual matters. With one exception, they were found to be perfectly comfortable discussing spiritual matters with what they believed to be a computer. The negative reaction of one person in this small sample is a reminder that not everyone will find spiritual conversation with a computer acceptable, or of any value.

Most rated their level of interest in the conversations they had quite highly (an average of 6.3 on a ten-point scale), though they rated the benefit they derived less highly (average 4.8), and they did not find the conversations very humanlike (average 3.2). That last finding might seem somewhat puzzling, as the responses on the computer screen were actually generated by a human. However, though there were technical features of the interaction with the computer interface that made this a very different experience from normal human conversation. Also, the 'Wizards' were perhaps trying hard not to appear to be humans, and so were constrained in what they felt able to say. Another factor was that the responses were delivered in automaton tone of voice, which may have disguised the humanlike content.

It has been suggested that people might feel they could say things to a computer they could not say to a human, but when asked about that in a questionnaire, all respondents said that was not the case. Most said the experience had been better than they expected, and most said they would be happy to have a long-term relationship with the system, as they might with a spiritual guide. Perhaps surprisingly, most said they could accept such a system as a priest or confessor, and most said they thought there was a useful future for such systems. This Wizard-of-Oz research shows that there is a body of people willing to discuss spiritual matters with what they believe to be a computer, and who felt they derived some benefit from doing so.

GPT Research

The acceptability of spiritual conversation with a computer was further investigated in a study using the OpenAI advanced GPT language model—a neural network of billions of nodes trained on a huge number of texts. This time, the responses were actually generated by the computer rather than by a human being (Wilks 2024). That enabled the research to test not just the acceptability of the computer interface but also the acceptability of computer-generated

responses. One of the main challenges in developing an automated artificial spiritual companion is to program the computer so that it has the necessary expertise for spiritual conversation. However, that has recently been made much easier by GPT methodology, which is rapidly becoming widely used.

This study had a larger sample of thirty-four participants, though again it is not claimed that they were representative. They were drawn from a sample of Facebook friends of Yorick Wilks and were perhaps more likely than most to find this kind of spiritual conversation acceptable. Those who provided data on the experience (thirteen) were about equally divided between being religious, atheist, or spiritual but not conventionally religious. Once again it was found that spiritual conversation through a computer was acceptable to participants. However, most of the GPT participants had only one session, and they seemed to be focused on testing what GPT could do rather than genuinely seeking spiritual conversation.

After conversations with the fully automated system, the participants rated their level of interest in the conversations quite highly, with an average rating of 6.8 on a ten-point scale, comparable to that found in the Wizard-of-Oz research. However, they rated the benefit they derived as lower (an average of 3.0). It is not clear whether that is because GPT provided poorer responses or due to differences in what the two samples of users were seeking. Strangely, GTP was rated as more humanlike (average of 6.0). That was probably because GPT had none of the delays and occasional glitches that characterized the Wizard-of-Oz research.

One interesting comment was that "each answer on its own was very humanlike (nine or ten), but after a few answers, the *style* of answer became so repetitive that, if I had been a judge, I would have said it was a bot." Another comment was, "I found it difficult to get into a more conversational style and felt very much like I was just looking something up." Another said, "I found the experience shallow." On the whole, GPT avoided giving opinions. However, it occasionally provided guidance. For example, in answer to the question, "How can we be more spiritual in our own lives?", the computer said:

> To be more spiritual in your own life, try to focus on developing a deeper connection with yourself and the divine. Spend time in nature or meditating to reflect on your inner self and find peace and harmony. Cultivate a sense of gratitude and appreciation for the beauty and joy in life, and practice acts of kindness and compassion for others.

There was less enthusiasm among the GPT users about the future of such systems. Only two gave a clear yes to the question of whether such systems had a future, and two more a qualified yes. On whether they could accept the system as a priest or confessor, only two gave a clear yes and one a qualified yes. On whether they might want to have a long-term relationship with the system,

only one gave a clear yes and two more a qualified yes. These responses are appreciably less favorable than those in the Wizard-of-Oz research. However, it is again not clear whether this is due to differences between the two groups of participants and their engagement with the system or the limitations of GPT.

Linguistic Analysis

Further information can be obtained from a detailed study of the interactions in this research. Linguistic analyses were carried out on fifty-six dialogue sessions drawn from both rounds with a wizard (twenty-two sessions) and with GPT-3 (thirty-four sessions), with between seven to 151 exchanges in each (i.e., the user saying something and the wizard or computer responding). The average number of exchanges was thirty-three per session, which demonstrated sustained interest in the experience. There tended to be fewer words per statement in users' interactions with GPT than in the Wizard-of-Oz research. The involvement of each user in the dialogue, judged by the number of words spoken, was not only sustained but grew throughout the session. In all but six of the fifty-six sessions, the user spoke more words in the last two-thirds of the session than in the first third. These figures suggest that the dialogues held the attention of those taking part.

To further analyze the tenor of the session, users' utterances were extracted and mapped onto a large, embedded word space, collecting information on which of the 500 human value words each subject word was closest to. It was found that that the most commonly occurring values were the following, listed in decreasing order of frequency across all sessions (and including words mapped onto them):

484 COMMUNITY (people, country, public, community, building, families, population)
274 COMMUNION (church, parish, communion, churches, worship, congregation)
225 LISTENING (talking, listening, listen, speaking, talked, communicate, conversations)
220 REALITY (actually, necessarily, simply, clearly, aspect, actual, essence)
215 FAMILY (family, mother, friend, friends, father, daughter, husband, relatives)
185 SPIRITUALITY (spiritual, prayer, spirituality, spiritually, prayers, rituals)
184 UNDERSTANDING (understand, understanding, respond, appreciate, describe)
177 BELIEF (believe, argument, belief, thinks, suggesting, believing)
154 TIME (moment, beginning, minutes, occasions, months, period, minute)
137 FUN (interesting, wonderful, fascinating, amazing, incredibly)
137 DIVERSITY (different, various, specific, larger, similar, separate)

Looking at how the sessions evolved, it was found that some topics were discussed more at the beginning of the sessions, and some were discussed towards the middle or end. For example, PROBLEMS were broached more in the final third of the sessions than in the beginning or middle. This occurs in normal conversations, too, as if people are wary of unburdening themselves too soon. In the beginning of sessions, people talked more about SPIRITUALITY, ADVERSITY, CHANGE, and HEALTH before gradually abandoning these somewhat abstract topics as the sessions progressed into more personal areas.

The same pattern of results was found in analyses drawn based on both Wizard of Oz and GPT-3 data as had been found in analyses of the Wizard of Oz data alone. That might be seen as a tribute to the quality of spiritual conversation GPT was able to provide. Differences in responses to the experience of spiritual conversation with a computer seem more attributable to differences between how the two groups were recruited and their personal interest in spiritual conversation than to differences in the quality of the Wizards of Oz and GPT as interlocuters.

The Roles of Humans and Machines in Spiritual Conversation

We think there is considerable potential for automated assistance with the spiritual life, and we expect to see growing acceptance of that. However, we also see significant limitations as to what a computer can contribute to someone's spiritual development. An automated spiritual companion could have considerable benefits and could provide a valuable supplementary resource in facilitating a person's spiritual development. However, if that is to happen, it is important not to make exaggerated and unconvincing claims for the contribution such an automated devices might make.

Humans and Machines

We emphasize that we do not envisage that an automated device could ever replace a human spiritual advisor, and it would not be intended to do so. Humans have a distinctive capacity for empathy and wisdom, and we do not expect computers to ever develop in a way that would completely replace those human qualities. The relationship between two humans will always be different from the relationship between a human and a computer. Each will have advantages and disadvantages. The empathy that can exist between two human beings is palpable, though empathy is still not well understood scientifically. It is also significant that humans are embodied, and there is an important "chemistry" between two human beings when they are physically together that is not felt with online connection. Embodiment is an important feature of human spirituality (Watts 2021).

There are several ways in which an automated companion might fall short of what is potentially available from a wise human spiritual conversation partner. First, drawing on the widespread distinction between conceptual and experiential

cognition, an automated companion can, at best, only draw on conceptual knowledge and not on the more intuitive understanding that may come from having undertaken an experiential spiritual journey itself. Such intuition is connected with the search for wisdom, which is wholly different from than the search for information or advice. Wisdom is a topic on which a fruitful integration of psychological research and theology has been developing (e.g., Wiseman 2020), and it might be a helpful challenge in future work on spiritual companions to explore whether spiritual "wisdom" could be implemented in a computer. It is not yet known whether it is possible for an automated conversation partner to display wisdom, or how to go about working towards that.

There is a separate question of whether a spiritual conversation partner could be a channel for the wisdom of God or God's guidance or blessing. Theologically, many might see it as the task of a spiritual conversation partner to make themselves available to be a channel for the wisdom of God, and many spiritual guides would aspire to be such a channel. Theologically, it would be assumed that God could communicate his wisdom through a computer if he wished, as it is not possible to set limits on divine action or communication. However, it is arguable that God chooses to act more through the human mind than through the natural world (Watts 2002, chapter 8) and that a human being can participate in the life of God in a way that a computer cannot.

As the acceptability of spiritual conversation with an automaton is investigated more fully, the issues may prove to be similar to those emerging about the acceptability of a virtual Eucharist (e.g., Dein and Watts 2023). Attending worship online achieves a basic degree of acceptability and is valued by many. However, there are divergent reactions to it, and many people feel it is a poor substitute for attendance in person. People often miss the physical reality of the church building in which worship takes place as well as the physical presence of other participants.

There will be many occasions when a client prefers to consult a human spiritual director rather than an automated companion and would be well advised to do so. It is interesting that the robot developed to help greet visitors at Longquan monastery in China, Xian'er, is able to engage in basic conversation about Buddhism but in certain situations advises people to consult a human monk (Cheong 2021). For example, it seems unlikely that people would accept an automaton taking on the sacerdotal role of absolution after confession, but it could offer some of what people gain from the conversation that takes place under spiritual direction or pastoral care.

Equally, there will be some occasions when an automated companion is preferred. For example, some might prefer confessing to a device rather than to a person, especially a person they know. In the early days of AI, it was argued, in a similar vein, that drivers prefer traffic lights to traffic police at junctions because they believe them fairer. There will also be significant individual differences here. Some people readily feel themselves to be accepted by other people,

whereas some tend to feel that others are often being judgmental about them. Over the coming decades, it is likely that increasingly clear rules of thumb will be developed about when it is helpful to use an automated spiritual companion and when a human should be consulted.

The Context of a Spiritual Community

As we move into a period in which at least some people will be making regular and extended use of spiritual conversation with a computer, new questions will arise. One is whether doing so increases or decreases people's sense of social isolation. An advantage of conversation with a computer is that it is always available as a constant companion, which might alleviate the sense of social isolation. However, if it is accompanied by too little human conversation, it might lead to an *increased* sense of isolation. There were suggestions of that in empirical work on interactions with BlessU2 (Löffler, Hurtienne, and Nord 2021). Excessive use of an automated companion might induce strain and a sense of deprivation somewhat analogous to Zoom fatigue.

There are likely to be various factors that influence the psychological impact of conversation with a computer, including where people fall on the introversion-extroversion dimension of personality and how deprived people are of natural human conversation. It may also depend to some extent on the topic of conversation. People may be happy to have almost all their interactions on some topics with an automaton, but it would not be surprising if there was a felt need for spiritual conversation to be at least partly with another human. Up to a point, an automated companion would be able to meet relational needs. However, there would probably be a sense that some relational needs can only be met by another person.

The acceptability of an automated companion might also depend on whether people are using spiritual conversation with a computer primarily to aid self-exploration or whether they are looking for wise advice. Facilitating self-exploration is one of the functions of a spiritual companion, and it might also be generally acceptable to turn to an automaton for that. It would be more challenging to implement a companion that could provide good-quality spiritual advice and guidance, and there might also be more skepticism from users about whether they could trust the advice they receive from a computer. Those who were primarily seeking help with self-exploration might be relatively happy with an automated companion. Others would be more focused on receiving authoritative help and guidance, and such people would probably be less content with an automated companion.

If the automated companion was embodied as an avatar on a computer interface, it would be very much a private companion of a particular individual, and it is unlikely that anyone else would ever interact with it. On the other hand, a companion implemented as a humanlike robot such as BlessU2 has a more

public presence. It might be in a church building and interact with a variety of people. That would give the robot companion a public presence in the spiritual community. Even if there were no differences in the technical capacities for spiritual conversation between these companions, it is clear that people react very differently to robots that have the external appearance of humans. It is a matter for future empirical work whether that would affect the acceptability of an automated spiritual conversation partner.

It is rapidly becoming possible to develop computers that can have spiritual conversation and other spiritually relevant interactions with humans. Preliminary evidence suggests that having a computer as a partner in spiritual conversation is acceptable to a significant number of people. Various methodologies are available to achieve this, and each has different strengths. Where there are training programs in spiritual practices or dispositions that have become routinized, it would not be difficult to write scripts that would enable a computer to deliver the training interactively. It would also be possible, in principle, to develop a personalized spiritual companion, building on previous work on personal companions for the elderly. These are likely to be quite widely used for spiritual conversation. They will not replace people but will be used alongside them. Gradually, wisdom will develop about when it is helpful to use a computer and when a person is needed.

In summary, we want to steer a path between, on the one hand, being completely dismissive of the potential value of spiritual conversation with a computer and, on the other, espousing an uncritical and exaggerated enthusiasm for what computers can contribute to the spiritual journey. An automated conversation partner has the advantages of ready availability and confidentiality. It can be a good source of information and facilitate self-exploration. However, people are likely to turn to a person for wisdom, for a sense of support and being understood, and for embodied connection. Humans are potentially better able to probe what a user is saying and offer wise personalized advice.

Acknowledgments

We gratefully acknowledge the support of the Templeton World Charity Foundation (TWCF0542) for the work reported here.

References

Balle, Simon. 2023. "Theological Dimensions of Humanlike Robots: A Roadmap for Theological Inquiry." *Theology and Science* 21 (1): 132–56. DOI: 10.1080/14746700.2022.2155916.

Barnard, Philip. 2009. "Depression and Attention to Two Kinds of Meaning: A Cognitive Perspective." *Psychoanalytic Psychotherapy* 23 (3): 248–62.

Bassett, Caroline. 2019. "The Computational Therapeutic: Exploring Weizenbaum's ELIZA as a History of the Present." *AI and Society* 34:803–12. https://doi.org/10.1007/s00146-018-0825-9.

Cheong, Pauline H. 2020. "Religion, Robots and Rectitude: Communicative Affordances for Spiritual Knowledge and Community." *Applied Artificial Intelligence* 34 (5): 412–31. DOI: 10.1080/08839514.2020.1723869.

———. 2021. "Bounded Religious Automation at Work: Communicating Human Authority in Artificial Intelligence Networks." *Journal of Communication Inquiry* 45 (1): 5–23. https://doi.org/10.1177/0196859920977133.

Clocksin, William. In press. "Steps Toward Android Intelligence." In *Cambridge Companion to Religion and AI*, edited by Beth Singler and Fraser Watts, 15–32. Cambridge: Cambridge University Press.

Dein, Simon, and Fraser Watts. 2023. "Religious Worship Online: A Qualitative Study of Two Sunday Virtual Services." *Archive for the Psychology of Religion* 45 (2): 191–209. https://doi.org/10.1177/00846724221145348.

Dorobantu, Marius, and Fraser Watts. 2023. "Spiritual Intelligence: Processing *Different* Information or Processing Information *Differently*?" *Zygon: Journal of Religion and Science* 58 (3): 721–48. 10.1111/zygo.12884.

———. In press. "Editorial Introduction." In *Perspectives on Spiritual Intelligence,* edited by Marius Dorobantu and Fraser Watts, 1–28. Abingdon-on-Thames, UK: Routledge.

Emmons, Robert A. 1999. *The Psychology of Ultimate Concerns: Motivation and Spirituality in Personality.* New York: Guilford Press.

Löffler, Diana, Jörn Hurtienne, and Ilona Nord. 2021. "Blessing Robot BlessU2: A Discursive Design Study to Understand the Implications of Social Robots in Religious Contexts." *International Journal of Social Robotics* 13:569–86. https://doi.org/10.1007/s12369-019-00558-3.

Ly, Kien, Ann-Marie Ly, and Gerhard Anderson. 2017. "A Fully Automated Conversational Agent for Promoting Mental Well-being: A Pilot RCT Using Mixed Methods." *Science Direct* 10 (Dec.): 39–46. https://www.sciencedirect.com/science/article/pii/S221478291730091X.

McDonald, Matt. 2024. "Catholic Answers Pulls Plug on AI Priest, Father Justin". *Catholic News Agency*, April 29, 2024. https://www.catholicnewsagency.com/news/257526/catholic-answers-pulls-plug-on-ai-priest-father-justin.

Miner, Adam S., Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. "Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health." *JAMA Internal Medicine* 176 (5): 619–25. DOI: https://doi.org/10.1001/jamainternmed.2016.0400.

Oman, Doug. 2013. "Defining Religion and Spirituality." In *Handbook of Psychology of Religion and Spirituality* (Second Edition), edited by Ray F. Palouzian and Crystal L. Park, 23–47. New York: Guilford Press.

Savage, Sara. In press. "Practical Spiritual Intelligence: Integrative Complexity." In *Perspectives on Spiritual Intelligence,* edited by Marius Dorobantu and Fraser Watts, 131–45. Abingdon-on-Thames, UK: Routledge.

Thibodeaux, Mark E. 2015. *Reimagining the Ignatian Examen*. Chicago: Loyola Press.

Trovato, Gabriele, Franco Pariasca, Renzo Ramirez, Javier Cerna, Vadim Reutskiy, Laureano Rodriguez, and Francisco Cuellar. 2019. "Communicating with SanTO—the First Catholic Robot." 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). New Delhi: Institute of Electrical and Electronics Engineers. DOI: https://doi.org/10.1109/RO-MAN46459.2019.8956250.

Watts, Fraser. 2002. *Theology and Psychology*. Basingstoke, UK: Ashgate.

———. 2017. *Psychology, Religion and Spirituality: Concepts and Applications*. Cambridge: Cambridge University Press.

———. 2021. *A Plea Embodied Spirituality: The Role of the Body in Religion*. Norwich, UK: SCM Press.

———. 2024. "Parameters and Limitations of Current Conceptualizations." In *The Oxford Handbook of the Psychology of Religion and Spirituality*, edited by Lisa J. Miller, 49–68. New York: Oxford University Press.

Weizenbaum, Jospeh. 1974. *Computer Power and Human Reason*. New York: Penguin Books.

Wilks, Yorick. 2010. "Introducing Artificial Companions." In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, edited by Yorick Wilks, 11–20. Amsterdam: John Benjamins.

———. 2024. "Artificial Companions and Spiritual Enhancement." In *Cambridge Companion to Religion and AI,* edited by Beth Singler and Fraser Watts, 272–92. Cambridge: Cambridge University Press.

Wilks, Yorick, and Afzal Ballim. 1991. *Artificial Believers*. Norwood, NJ: Erlbaum.

Williams, Hattie. 2024. "CT Survey Reveals Split over AI". *Church Times*, May 3, 2024. https://www.churchtimes.co.uk/articles/2024/3-may/news/uk/ai-ministry-better-than-nothing-church-times-survey-says.

Wiseman, Harris. 2020. "Wisdom and Moral Formation." In *Mutual Enrichment between Psychology and Theology*, edited by Russell Re Manning. London: Routledge.

Worthington, Everett. 2003 *Forgiving and Reconciling: Bridges to Wholeness and Hope*. Downers Grove, IL: Intervarsity Press.

Young, William. 2019. "Reverend Robot: Automation and Clergy." *Zygon: Journal of Religion and Science* 54 (2): 479–500. https://doi.org/10.1111/zygo.12515.

———. 2022. "Virtual Pastor: Virtualization, AI, and Pastoral Care." *Theology and Science* 20 (1): 6–22. DOI: https://doi.org/10.1080/14746700.2021.2012915.

Zygon®: Journal of RELIGION & SCIENCE

Tretter, Max. 2024. "What Is the Afterlife Like for Robots? An Experimental Eschatological Sneak Peek." *Zygon: Journal of Religion and Science* 59 (3): 692–716. DOI: https://doi.org/10.16995/zygon.10903

Open Library of Humanities

What Is the Afterlife Like for Robots? An Experimental Eschatological Sneak Peek

Max Tretter, Research Associate, Institute of Systematic Theology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, max.tretter@fau.de

People have always pondered their afterlife. Now, as AI and robotics continue to advance and proliferate, a new question emerges: Is there also some kind of "afterlife" for robots—and how can we envision it? This article seeks to explore these very queries from a Christian perspective. To tackle the initial question, I argue that, following the thoughts of St. Paul and St. Augustine, the whole of creation is sinful and seeks completion, it would be inconsistent to nurture such an all-encompassing hope yet exclude robots from it. From a Christian perspective, we should therefore assume the existence of an afterlife for robots. To decipher how we can envision it, I examine two pop-cultural depictions from the television episode "Zima Blue" and the television series *Futurama*, questioning whether they provide a fitting image of eschatological completion for robots. This methodological approach allows me to present a spectrum of conceptions of robotic afterlife that, when examined through the lens of systematic theology, appear plausible, offering fresh impetus for eschatological and robophilosophical reflections.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

In the twentieth century, Protestant theology rediscovered eschatology as a source of Christian hope (Moltmann 1967) and a central motif of Christian faith (Bultmann 2012). This resurgence led to the past century being referred to as the "eschatological century" (Schwöbel 2002, 437–68). Currently, in light of the comprehensive digitalization processes of the twenty-first century (Floridi 2015), several theologians are exploring artificial intelligence (AI) and robots (Smith 2022; Puzio, Kunkel, and Klinge 2023; Balle 2022; Dorobantu 2022), posing questions such as whether robots have souls, whether AI can have faith, and what implications these new technologies have for existing beliefs.

Merging the eschatological impulses of the twentieth century with the technological themes of the twenty-first century, this article delves into an eschatology of artificially intelligent beings. Focusing on robots as a quintessential AI technology and the afterlife as a prime example of eschatological settings, I ask: Is there some kind of afterlife for robots—and if so, what is it like? Or, to put it more precisely: From a Christian theological standpoint, should it be assumed that there is an afterlife for robots—and how can it be envisioned?

For some, theological inquiries into robot afterlife might sound similar to the query attributed, albeit incorrectly, to scholastic theology (Ross 2009) of how many angels can dance on the head of a pin (or a silicon chip) (Simon 2021). This naturally leads to questions: Why should this even be asked? Are there not more important theological matters to consider? These questions are absolutely valid. Indeed, theology faces more significant challenges than the eschatological what and whereabouts of robots. However, just as the scholastic inquiry into how many angels can simultaneously be in one place—as absurd as it may appear to us today—once represented a crucial test of consistency in theological thinking (Aquinas 1947), so too can the question of robot afterlife offer valuable insights for contemporary theology. First, theological reflection on eschatological subjects is influenced by the technologies and possibilities of our time (Burdett 2014). This is especially apparent in how AI, augmented reality, virtual reality (Geraci 2010), mind uploading (Gaitán 2019), and cryonics (Mercer 2017) are incorporated into modern eschatological debates. As technology and eschatological considerations have always been intertwined (Burdett 2014), pondering a robot afterlife might help advance thinking about humans, their salvation, and their visions of afterlife. Secondly, theology still predominantly focuses on humans, their sinfulness, their need for redemption, and their life beyond death. As, for instance, animal theology (Linzey 2022) and ecotheology (Bauman 2018) have demonstrated, this theological anthropocentrism has led to manifest problems in dealing with nonhuman animals and the environment. In this regard, transcending theological anthropocentrism and including nonhuman (Clough 2012), and perhaps even nonorganic entities—see Paul Tillich's remarks on the "[immense] religious significance of the inorganic"

(Tillich 1963, 19)—is a valid and crucial concern for theology. The exploration of cyborg (Midson 2018) and robot theology (Smith 2022) can make a significant contribution to a shift towards non-anthropocentric perspectives in theological thought, highlighting the meaningfulness of contemplating an afterlife for robots. That is one reason this article explores whether, from a Christian theological standpoint, some kind of afterlife for robots should be assumed—and how it might be envisioned.

Yet, another reservation may arise, namely, whether it makes any sense at all to speak of an afterlife for robots, as this term implies that robots "live." If one, for instance, applies a common definition of "life" from biology (Dabrock et al. 2011; Mariscal 2021), robots neither have a metabolism nor grow or reproduce; thus, they cannot be described as living. Similar conclusions arise when considering alternative definitions of life from, e.g., existentialism, critical theory, or theology (Bedau and Cleland 2010; Sundermeier et al. 2010). However, not all approaches are as skeptical concerning the question of whether robots live (Cheok and Zhang 2019). For example, there are now broad discussions about "synthetic" (Dabrock et al. 2011) or "artificial life" (Langton 1997) that ask how such life could be understood and to which entities it might apply. Building on these and similar debates (Cheok and Zhang 2019), this article assumes a very comprehensive, almost metaphorical, understanding of life that includes not only "living beings" in a strict sense but also, for example, the life cycle of objects, thus encompassing robots. Still, given the ongoing question of whether robots actually do "live" (Putman 1964; Cheok and Zhang 2019) and my rather broad understanding of life, the question becomes even more pressing: Why bother using the term "afterlife" in this context? To understand this choice, it is essential to understand that the focus of this article is more on the aspect of "after" than "life," i.e., focusing on the question of what happens to robots eschatologically, not on the question of whether robots live or not (a more detailed explanation of my understanding of terms follows in the Terminology and Methods section). Furthermore, comparing this term with others that could have been used to discuss the eschatological fate of robots is insightful. Instead of asking about an afterlife for robots, one might have asked whether there is a heaven for robots or if robots can attain eternal life. However, each of these alternative terms comes with its own issues: the term "heaven" commonly evokes spatial associations, raising questions about where this heaven might be located and how it might relate to other kinds of heaven, such as the human heaven; the term "eternal life" relies even more on the concept of life than afterlife does; and other terms have different focuses, missing the point of this article, such as the question of "salvation" for robots. For these reasons, I ultimately chose the term afterlife, despite its issues, and frame the question of robots' eschatological fate as whether to assume robotic afterlife exists and how to imagine it.

To address these questions, I first clarify in more detail in the Terminology and Methods section what I mean by "robots" and "afterlife" and outline my approach. This involves two steps: initially, I demonstrate that from a Biblical-theological perspective, it would be inconsistent to assume that there is no afterlife for robots. Subsequently, I present two pop-cultural representations of robot afterlife—one from the television series *Love, Death & Robots* and one from the television series *Futurama*—and assess whether they provide adequate depictions of robotic completion and can thus be regarded as valid representations of robot afterlife. This enables me to ultimately conclude how an afterlife for robots can be appropriately envisioned from a Christian perspective.

Terminology and Methods

Before asking about robot afterlife, I need to clarify some terminology and explain my methodological approach. In this section, I first define what I mean by a "robot", then elucidate my understanding of "afterlife", and finally outline my approach for this article.

Terminology: What Is a Robot?

When I speak of "robots," I follow Alan Winfield's (2012) proposal to understand them as technical entities characterized by six "functions": they are powered by a source of energy (usually electrical); can "sense" their environment using sensors—i.e., technical devices for perceiving their environment—and manipulate it using effectors—i.e., technical devices for acting on their environment; move; send and receive signals; and are controlled by algorithms.

This definition, however, is very broad and encompasses a vast spectrum of robots, each with unique technical abilities, appearances, and attributes (Nyholm 2020). This diversity results from the multitude of applications and tasks for which they have been designed and built (Winfield 2012). The areas in which robots are utilized and the tasks they must perform dictate the requirements they need to meet and their capabilities (Floridi 2015). The resulting differences can be illustrated by comparing two exemplary robots: simple industrial manufacturing robots and highly complex social robots such as Pepper or NAO, or even more futuristic or sci-fi robots such as the robot artist from "Zima Blue" or *Futurama's* Bender, whom I discuss later.[1] Simple manufacturing robots are designed to carry out one or a few mechanical tasks, mostly in industrial production chains, as efficiently and accurately as possible (Weber 2019). Due to this specialization, they possess few highly specialized sensors or effectors and are typically equipped with relatively low intelligence. In contrast, highly complex social robots are designed to navigate complex settings and perform a wide range of tasks—most prominently, communicative and social tasks, e.g., interacting with humans—in diverse environments (Korn

2019). Therefore, they are usually equipped with several different sensors and highly versatile effectors, and endowed with a high degree of intelligence.

Of course, there are many other robots whose equipment, intelligence, and capabilities fall somewhere between those of a simple industrial robot and those of complex social robots. To illustrate this, one could view single-purpose robots and complex social robots as the endpoints of a spectrum of various robots (see Figure 1). All robots can be placed on this spectrum based on their equipment, intelligence, and capabilities: the more limited they are to a single, primarily mechanical task and equipped solely for that task, the more fixed they are to one context, and the less intelligent they are, the closer they are to the pole of single-purpose robots. In contrast, the more comprehensive their sensors and effectors, the more tasks they can perform in multiple contexts, and the more intelligent and adaptive they are, the closer they are to the pole of the highly complex social robot.

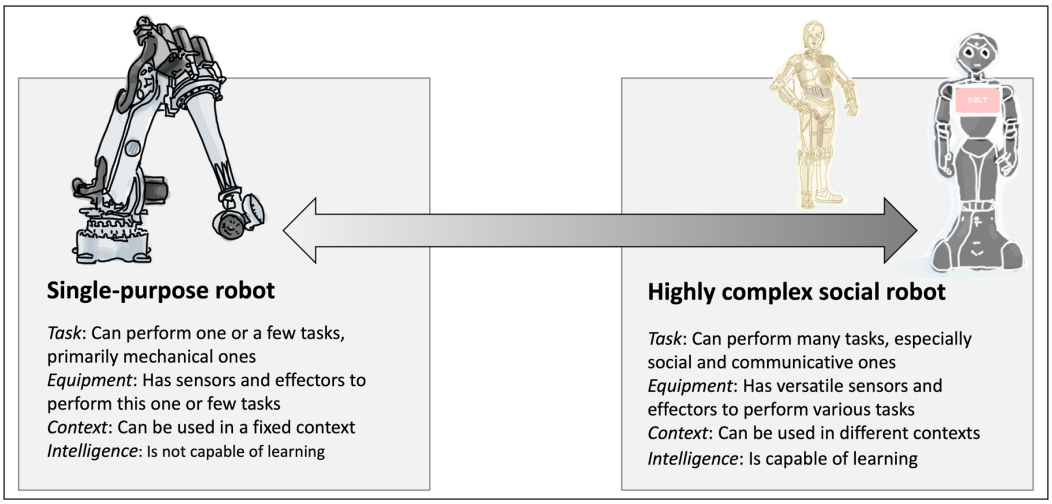


Figure 1: Scale of different robots, from simple industrial robots (left) to highly complex social robots (right). Copyright of the robot images: Isabella Auer 2024.

Terminology: How Is Afterlife Understood?

Depending on particular religions, denominations, and cultural contexts, there are many different ideas of what afterlife is or what it might be like (Lang 2019). For instance, in the Revelation of John, there is the idea of a heavenly Jerusalem with golden walls and pearly gates where all real believers go after their death (Revelation 21:1–27). In the Qur'an, there is the idea of a garden filled with streams of milk and honey and furnished with noble pieces of furniture (Sharp 2015). There is also the idea of a spiritual place where "souls" go after they have been separated from the earthly body (Mariev and Marchetto 2017). However, all of these conceptions of afterlife are highly anthropocentric, presupposing human values (gold and pearls as highly valuable), needs, and desires (an abundance of milk and honey as a symbol of good provision). Since robots do

not necessarily share these values, needs, and desires, it seems unpromising to apply such anthropocentric notions of afterlife when thinking about afterlife for robots. Therefore, it makes sense to start with a more open understanding of afterlife that makes fewer or no anthropocentric assumptions. Such an understanding can be derived from Wilfried Härle's *Outline of Christian Doctrine* (Härle 2015). Although Härle himself does not use the term "afterlife" in this work, his eschatological reflections include concepts that could colloquially be referred to as "afterlife." Since Härle's considerations are important for the understanding of afterlife underlying this contribution, I elaborate them in a little more detail here.

In *Outline of Christian Doctrine*, Härle offers a structured, tripartite exploration of the key themes of the Christian faith from a primarily Lutheran-Evangelical viewpoint. In the first foundational part, Härle examines the "essence" of Christian faith, i.e., what remains constant in the faith despite all changes in the world. He describes faith as an understanding of reality grounded in God's revelation. This view is characterized by believers holding a deep trust in God's will for salvation, which profoundly shapes their thoughts, emotions, and actions. In the subsequent parts of his work, Härle applies deductive reasoning to explore which conceptions of God (Härle 2015, Part IIA) and the world (Härle 2015, Part IIB) are either logically presupposed by this understanding of faith or consequently follow from it. These considerations culminate in eschatological reflections. Observing the unfinished and fragmentary nature of the "world of earthly history" (Härle 2015, 490), Härle argues that it cannot represent the ultimate fulfillment of God's salvific plan. Consequently, God's salvation must be realized elsewhere: in the eschaton. Through this line of reasoning, Härle argues that the Christian understanding of reality must include some form of hope for an eschatological "completion" of the world in order to avoid internal theological contradictions. Upon establishing the necessity of hope, he envisions this eschatological completion. Differentiating "completion" from "perfection," he argues that the "completed world" is not characterized by everything being "perfect," i.e., without flaws, limitations, or impairments, as such perfection is a quality reserved for God alone. Rather, the completed world must be envisioned as a world that does not remain in its fragmentary state but allows all of creation, while still bearing traces of earthly history, to fulfill their (God- or self-)given purpose without being hindered. This leads Härle to some methodological conclusions: to inquire what, for instance, the completed world or humans in their eschatological state of completion are like—Härle stresses several times that eschatological completion is not limited to humans alone (Härle 2015, 526)—one must engage with other theological topics, such as the doctrine of creation or anthropology, and examine them from the perspective of completion.[2]

The eschatological concepts Härle develops, or similar ideas of completed creation, are regularly referred to in both theological discourse and everyday language as "afterlife"—especially when discussing the eschatological fate of humans. Therefore, in this article, building on Härle's considerations, I understand afterlife as an eschatological state brought about by God, where all earthly flaws or errors are overcome, and creation is completed.

Methods: How Are the Questions Approached?

With these terminological clarifications in place, I now outline how I tackle the main questions of this article: Should it be assumed that there is some kind of afterlife for robots—and how can it be envisioned? As eschatological questions, these are inherently subject to an "eschatological reserve" (Kirwan 2015), meaning that they cannot be answered through empirical observations or experiential knowledge (Härle 2015). Nevertheless, it is essential to ensure that eschatological statements are reasonable and comprehensible. Härle accomplishes this by applying a deductive method and, as previously discussed, deriving his eschatological statements logically from previous assumptions. All his eschatological statements are either prerequisites or conclusions of his other systematic-theological assertions. I apply a similar approach in this article.

In my first step, I focus on the initial question: Should it be assumed that there is some kind of afterlife for robots? I present three approaches that may be used to answer this question. First, I present a philosophical-theological approach, which seeks to draw conclusions about the existence of a robot afterlife based on theologically significant abilities robots do or do not, or may or may not, possess. Second, I use a psychological-philosophical approach, which attempts to reach a conclusion using the concept of spiritual intelligence. Since neither of these approaches provides a clear and methodologically sound answer, I follow a third approach. Considering that Christians, following St. Paul and St. Augustine, have always believed that all of creation is sinful and yearns for eschatological completion, this approach suggests that it would be contradictory *not* to assume there is an afterlife for robots.

After arguing in favor of some form of robotic afterlife the first step, I then focus on the second question: How can an afterlife for robots be envisioned? I follow the approach of Härle (and several other theologians, see endnote 2), who formulates eschatological statements by examining other theological topics, such as creation theology or anthropology, from a perspective of completion and asking how the world or humans might be envisioned in a state of completion. Unlike Härle, however, I do not develop my concept of robotic afterlife from scratch. Instead, I draw on existing depictions of robotic afterlife from popular culture and examine whether they provide an adequate picture of the eschatological completion of robots.[3] The two pop-cultural conceptions

of robotic afterlife I have chosen for this purpose come from the television series *Love, Death & Robots* and *Futurama*. I have chosen these because they are contemporary—both series are still in production—reach large audiences, and present two very different visions how afterlife for robots might look.

This latter approach is based on the assumption that Christian eschatological ideas have always been in active dialogue with the imaginations of their environment, being heavily influenced by them and, in turn, significantly shaping them. This is evident in the Bible, where, for instance, the concept of Abraham's bosom (Luke 16:22) is heavily influenced by ancient Greek ideas of Hades (Beinstein 1993), or early Christian notions of heaven as the place where the soul goes after leaving the body, which show a strong Platonic influence. In the Middle Ages, Dante's *Divine Comedy*, which was itself influenced by Biblical motives and Christian ideas, had a significant impact on how heaven and hell were envisioned in Christianity, and it continues to shape these concepts to this day. Similar interactions and influences can be seen between Christian eschatological ideas and the works of Hieronymus Bosch (*The Garden of Earthly Delights*), Michelangelo (*The Last Judgment*), and John Milton (*Paradise Lost*). All these examples illustrate how deeply Christian eschatological ideas were and still are influenced by imaginations of their time. Against this backdrop, I propose that contemporary imaginations about robotic afterlife, for example those developed in popular culture (Flannery-Dailey 2003), can also help us develop an idea of what an afterlife for robots might look like.

Can an Afterlife for Robots Be Assumed?

There are several approaches to answering the questions of whether there is some kind of robotic afterlife and whether, from a Christian perspective, it should be assumed that there is an afterlife for robots. In this section, I attempt to answer these questions using three different approaches. Each approach starts from a different premise and proceeds in a distinct manner. Yet, in theory, these approaches are fundamentally compatible with each other.

Philosophical-Theological Approach: The Properties of Robots

One initial approach is primarily philosophical and focuses on robots' properties. This approach asks whether these properties are theologically significant and whether one can conclude from them that robots are in need of completion (Furse 1986). If one concludes that robots, based on their current or anticipated abilities, indeed need completion, it makes sense to assume that some form of afterlife for robots exists.

Several authors' considerations can be categorized under this approach. For instance, many authors describe key abilities robots already possess or are expected to possess in the future as theologically significant. These include intelligence and consciousness (Furse 1986; Rosenfeld 1966), "deep levels of . . .

judgment" (Cantwell Smith 2020), and their current or anticipated embodiment, empathy, and sociality (DeBaets 2012). Some authors conclude that current or future robots possessing these characteristics should be considered to have a soul (Livingston and Herzfeld 2009), be images of God (Foerst 2009; Dorobantu 2020) or images of images of God (Midson 2013), or be sinners (DeBaets 2012) or at least participate in human sin (Smith 2022). Yet, classifying robots in this way would suggest that the existence of a form of robotic afterlife should be considered. For, assuming that such an afterlife for robots did not exist would imply that, within the conceptual framework of this line of argument, there is no possibility for completion for robots that are assumed to be in need of completion—a conclusion that, even if only in our minds, is difficult to reconcile with the image of a loving and benevolent God.

This line of argument also works the other way round. Several authors point to properties like having a free will (McGrath 2011), being conscious (Strand 2021), being able to have relations and be vulnerable (Dorobantu 2023, 2021), the capacity for empathy and values (Gill 2020), and the quality of being born and dying (Krajewski 2018; Deli 2020), arguing that these are theologically central. But since robots currently do not possess these properties—some even argue that robots will never possess any of them—robots should not be considered as in need of completion (Strand 2021; Swann 2021). This leads to the conclusion that there is no necessity for something like an afterlife for robots.

As this brief overview shows, the philosophical-theological approach does not lead to a definite conclusion but allows for different conclusions to be drawn. Some authors assert that robots already possess properties that are theologically significant or that they can be expected to possess these in the future, suggesting that there must (or at least could) be some kind of robotic afterlife. Others argue that robots do not, and maybe never will, possess properties of theological significance, deeming any speculation about their afterlife superfluous. Meanwhile, a subset of authors prefers not to make a definitive claim about whether robots possess, do not possess, can possess, or cannot possess these properties but stresses that if robots did have such theologically significant properties, questions concerning their completion and how to handle them in religious contexts would "be the central cultural conflict for religion in this century" (Simon 2021). Ultimately, this approach, which seeks to deduce conclusions about a robotic afterlife from the properties of robots, fails to provide a clear-cut answer.

Psychological-Philosophical Approach: Robots and Spiritual Intelligence

A second approach adopts a psychological lens, focusing on spirituality. Central to this approach is Howard E. Gardner's distinction of different forms of intelligence (Gardner 1999) as well as Robert A. Emmons's proposal that there is

a specific form of spiritual intelligence. In his monograph (Emmons (1999) and further refinements in subsequent discussions (Emmons 2000; Gardner 2000; Kwilecki 2000; Mayer 2000), Emmons defines spiritual intelligence as a type of intelligence that transcends traditional cognitive abilities. It includes, among other things, the ability to integrate one's experiences and reflections into a larger spiritual framework, connect with a higher power—whether it be nature, God, a collective, or something else—and demonstrate virtuous behavior, such as humility and compassion.

An approach that could be called psychological-philosophical builds on this definition of spiritual intelligence and asks, from a multidisciplinary perspective, what prerequisites are required to harness and exhibit such intelligence (Dorobantu and Watts 2023; Watts and Dorobantu 2023) and whether artificial entities like robots possess the necessary prerequisites for spiritual intelligence, i.e., whether they could be spiritual (Dorobantu and Watts 2023). If answered affirmative, a cascade of subsequent questions unfold: How might artificial spirituality manifest (Geraci 2007)? What could the belief systems of artificially intelligent entities or robots look like in terms of content (Sampath 2018; Klinge 2023)? How might they reconcile or confront preexisting beliefs (Dorobantu and Watts 2023)?

Based on these considerations, conclusions can be drawn about whether the existence of some kind of robotic afterlife should be assumed. The reasoning is similar to that of the philosophical-theological approach: if robots are indeed considered spiritually intelligent and capable of contemplating their eschatological fate, it would seem somewhat cruel to simultaneously assume that there is no afterlife for them. This reasoning suggests that the possibility of an afterlife for robots should at least be considered.

This approach, at first glance, appears promising, as it is grounded in tangible, and at times empirically backed, reflections on spiritual intelligence. Every pondering about robotic spirituality thus has at least some empirical underpinnings. Yet, it is not permissible to conclude from the fact that robots might be able to imagine some kind of robotic afterlife that such afterlife must exist. For one, such a line of argument resembles so-called ontological arguments for the existence of God, such as that of Anselm of Canterbury, wherein the existence of God was inferred from the (im)possibility of thinking God ("*id quo maius cogitari non potest*") (Malcolm 1960). As Immanuel Kant (1922) demonstrates with his distinction between a hundred real and imagined thalers in his *Critique of Pure Reason*, such conclusions, where ontological inferences are drawn from imaginations, commit a category mistake (Sala 1990). Second, it is possible and—when compared with human spirituality and the diversity of human religious beliefs (James 2014)—very likely that robots, if they were artificially spiritual, would develop a variety of different conceptions of an afterlife or reject the existence of some kind of afterlife for spiritual reasons.

This would in turn lead to different conclusions about the existence of an afterlife within this psychological-philosophical approach.

Biblical-Theological Approach: Completion for All of Creation

Last but not least, there is a third approach beyond those previously mentioned. This method does not focus on individual robots and their characteristics in order to assess their theological significance. Neither does it ask about spiritual intelligence and the possibilities for artificial spirituality. Instead, based on Biblical and theological arguments,[4] this approach asks about the eschatological whereabouts of all creation, regardless of whether human, animal, plant, or robot, and regardless of the characteristics possess and potential theological significance. All of these "entities" are treated as part of God's creation. And if following the Augustinian doctrine of original sin, it must be assumed that the entire creation is subject to sin (Williams 1994). Just as Paul's letter to the Romans (8:18–23) suggests, it can be assumed that all creation, albeit possibly to varying degrees, hopes for "redemption":

> For I reckon that the sufferings of this present time are not worthy to be compared with the glory which shall be revealed in us. For the earnest expectation of the creature waiteth for the manifestation of the sons of God. For the creature (κτίσις) was made subject to vanity, not willingly, but by reason of him who hath subjected the same in hope, Because the creature itself also shall be delivered from the bondage of corruption into the glorious liberty of the children of God. For we know that the whole creation groaneth and travaileth in pain together until now. And not only they, but ourselves also, which have the firstfruits of the Spirit, even we ourselves groan within ourselves, waiting for the adoption, to wit, the redemption (ἀπολύτρωσις) of our body. (Romans 8:18–23)

In these verses, which are the climax of the argument Paul makes in the first half of his letter to the Romans (Dunn 1988, 466–67), the evangelist describes the hope of all creation in order to assure the addressed Romans that they can be certain about their salvation (Romans 8:23–39). For if, Paul argues, even the creation hopes to be redeemed, then believers may hope all the more for their redemption. Paul does not describe in detail what this redemption of the creation might look like as it is not important to his argument. What is indeed important is that all of creation hopes for redemption (Hahne 2006). This redemption includes the liberation of creation from its sufferings (Gibbs 1971, 34–47) and its reconciliation with itself, humans—who have brought these sufferings upon it with their imperial aspirations (Jewett 2006, 513)—and God. What Paul describes in these passages with the term ἀπολύτρωσις, and what is translated as "redemption" in the King James Version (and other translations of the Bible), shows significant conceptual parallels to what Härle describes

as "completion." Therefore, it is appropriate to say that in his epistle to the Romans, Paul argues that there will be completion for all of creation. And since completion is the defining characteristic of afterlife—an eschatological state of completion, where everything behaves as God originally planned in their counsel—it seems reasonable to conclude that, according to Paul's letter to the Romans, there is some kind of afterlife for all creation.[5]

Consequently, if Paul is trusted and it is believed that there is completion, i.e., some kind of afterlife, for the whole of creation, and if in Paul's theology and terminology "creation" includes "everything"—the term κτίσις in the New Testament (Foerster 1957), especially in the letters of Paul (Wischmeyer 1996), includes the whole cosmos, all the visible and invisible things that were created and preserved by God through Christ out of nothing—then it must be concluded that robots are also part of this creation (Foerst 2004) and that there must be some kind of afterlife for robots.[6] Conversely, it would be contradictory to believe, in line with Paul, that there is an afterlife for everything while simultaneously maintaining that there is no such thing for robots. Thus, from a Biblical-theological standpoint, there is no reason to argue against an afterlife for robots—quite the opposite, in fact.[7]

How Is Robotic Afterlife Portrayed in Popular Culture?

Having established from a Biblical-theological perspective that it should indeed be assumed that there is some kind of afterlife for robots, I can now delve deeper into the question: How can this be envisioned? To explore this, I first outline two pop-cultural depictions of robotic afterlife, followed by an examination of whether these representations provide an adequate picture of eschatological completion for robots.

There is a wide array of representations of robotic afterlife in contemporary popular culture (Flannery-Dailey 2003). From this diverse selection, I introduce two examples: the concept of robotic completion presented in the television episode "Zima Blue" and the portrayal of a robot heaven and a robot hell from the cartoon series *Futurama*. I have chosen these two representations because they share similarities—both are animated, created in the 2010s, still in production, and have comparable target audiences—while also presenting distinct visions what an afterlife for robots might look like.

Love, Death & Robots's "Zima Blue" Episode

"Zima Blue" is the fourteenth episode of the first season of the Netflix series *Love, Death & Robots*. Passion Animation Studios animated the episode, which is based on the 2006 short story of the same name by science fiction author Alastair Reynolds (2006), who also directed the episode. The episode, approximately ten minutes long, first aired on March 15, 2019 (Valley 2019).

The protagonist of the short story is a former pool-cleaning robot who has since become self-aware and creative and is now a renowned and celebrated artist

across the universe. The hallmark of the robot's art is the use of a unique color called "Zima Blue." This color serves as the focal point of the robot's artistry, driving it to color everything Zima Blue and ultimately be completely absorbed by the color. In pursuit of this vision, the color progressively dominates a larger area in each subsequent artwork, resulting in the robot painting an entire shower of comets Zima Blue in its penultimate work. By the time the story unfolds, the robot has invited select journalists and art enthusiasts to witness the unveiling of its latest and final work of art. As the episode reveals, this final piece, unlike its predecessors, is a performative work of art. As the audience watches, the robot-artist dives into a swimming pool made of Zima Blue tiles, gradually shedding all the components that endowed it with self-awareness and creativity while floating in the water. The performance concludes with the former robot-artist reverting to its original form as a pool-cleaning robot, which it was before technical upgrades turned it into an artist. As a pool-cleaning robot, it can ultimately focus entirely on maintaining the pool's Zima Blue tiles (see Figure 2), realizing its ultimate goal: to be utterly consumed by Zima Blue (Valley 2019).

The episode illustrates how only by shedding all spare parts and relinquishing its self-awareness and creativity can the robot fully dedicate itself to its task, achieve fulfillment, and merge with Zima Blue (Aguilar Alcalá 2021). *Love, Death & Robots* thus presents the notion of an eschatological state for robots in which they can freely and unimpededly engage in the one or few task they were originally designed for and find completion doing so (Heavy Spoilers 2019).

Futurama's "Ghost in the Machines" Episode

The cartoon series *Futurama* presents an entirely different portrayal of robotic afterlife, offering a satirical take on predominantly Catholic concepts of heaven and hell. If robots have behaved well during their "lifetime"—as the protagonist robot Bender does in the episode "Ghost in the Machines" (Groening et al. 2011), where "he" (Bender is portrayed as a male character in the series) sacrifices himself for his human friend—they ascend to Robot Heaven with a sort of heavenly robot body. Robot Heaven is situated high above in the clouds, enclosed by a golden fence and pearly gates, and ruled by Robot God (see Figure 3). In Robot Heaven, robots enjoy unspecified pleasurable rewards (Oschman 2013) and have the freedom to act and develop as they please.

However, if robots have misbehaved during their lives or broken a contract they previously made with the Robot Devil, they descend to Robot Hell. This is depicted as a physical location beneath future Atlantic City, where the bodies of damned robots endure eternal torture inflicted by the Robot Devil and his robotic minions, using fire, tridents, and other torture devices reminiscent of Christian iconography (Pinsky 2003, 229–35). Once there, robots are expected to remain in either Robot Hell or Robot Heaven indefinitely. A return from Robot Heaven is only possible when Robot God sends the robot back to Earth in an earthly robot body. Escaping Robot Hell can only be achieved when

Robot God "redeems" the damned robots or if the robot defeats the Robot Devil in a fiddle contest (Young 2013).[8]

The conceptions of Robot Heaven and Robot Hell in *Futurama* draw parallels with popular religious beliefs and artistic representations of heaven and hell for humans. They evoke places where robots spend their afterlife either suffering eternal torment in a physical form or experiencing eternal pleasure, with the freedom to do whatever they want, unfolding and developing in any way they desire.

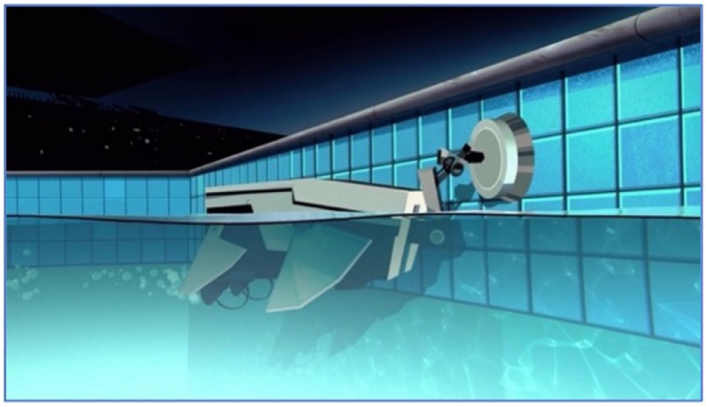


Figure 2: The former robot-artist is now cleaning the Zima Blue-colored pool tiles. Screenshot from Valley 2019.

Figure 3: Bender at the gates of Robot Heaven. Screenshot from: Claffey and Avanzino 2011.

Reality Check: Do These Pop-Cultural Portrayals of Robotic Afterlife Provide an Adequate Picture of the Eschatological Completion of Robots?

Building on prior considerations, I can now inquire whether these pop-cultural depictions provide an adequate picture of eschatological completion for robots. It quickly becomes apparent that the answer hinges primarily on the underlying

understanding of robots being considered. For instance, the portrayals of robotic afterlife in *Futurama*—where robots can freely develop and pursue their own goals and activities—do not seem to represent a fitting form of eschatological completion for single-purpose robots that are neither intelligent nor adaptable. Similarly, the idea of forever pursuing a single task and finding fulfillment in it (as in "Zima Blue") may not be the ultimate completion for highly complex social robots, unless it is assumed that at some point all robots must renounce their higher abilities and restrict their capacities—which would, in turn, introduce its own set of problematic implications.

Conversely, simple industrial robots that are designed from the outset to perform one or a few mechanical tasks might find fulfillment, and possibly even completion, in continuously and flawlessly performing those tasks. Or, for highly complex social robots, completion may appear as the ability to freely develop, evolve, and pursue any task they "want" to.

Therefore, none of the aforementioned depictions of robotic afterlife from popular culture offer a definitive one-size-fits-all portrayal of eschatological completion for all types of robots. Instead, both depictions seem to illustrate crucial aspects of robotic completion. This suggests that, in the end, it is more important to match the right conception of robotic afterlife to each robot than to seek one ultimate vision of afterlife for all kinds of robots. Based on this conclusion, a scale of the variety of robotic afterlives, analogous to the earlier scale of different robots (see Figure 1), can be created, with "Zima Blue"-like concepts at one end, where single-purpose robots find eschatological completion in eternally fulfilling the same (mechanical) task without wear and tear, to *Futurama*-like visions for highly complex robots, which would closely resemble accounts of human afterlife (see Figure 4). Between these two visions of afterlife for robots, there may be many other forms of robotic afterlife, each corresponding to a specific robot with its specific set of properties.

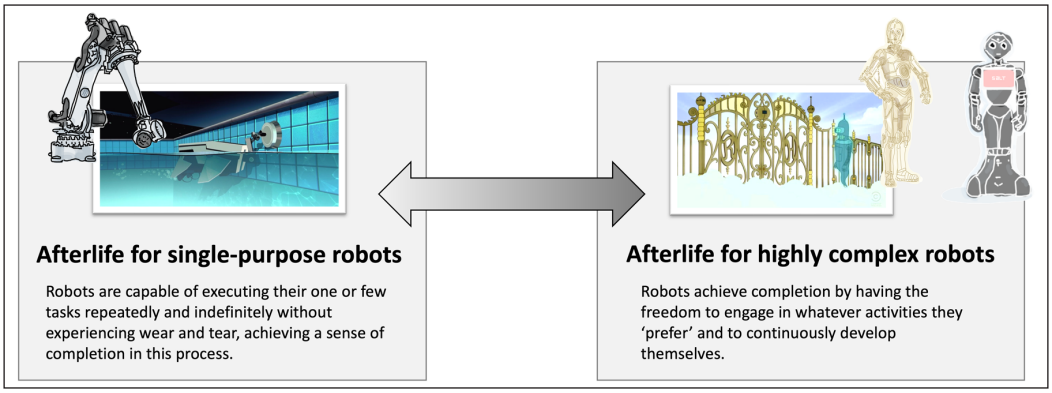


Figure 4: Scale of different imaginations of robotic afterlife, from an afterlife for single-task robots (left) to afterlife for highly complex robots (right). Copyright of the images: see the annotations in Figures 1, 2, and 3.

Discussions

Contemplating robotic afterlife as well as the conclusions of this article might raise several questions that warrant discussion: First, are the conclusions drawn earlier, despite attempts to avoid anthropocentrism, not still overly anthropocentric? Second, how do the various robotic afterlives relate to one another?

Question of Perspective: Are the Findings Not Too Anthropocentric?

In the earlier discussion, great effort was made to avoid anthropocentrism. This was evident in the definition of afterlife chosen for this paper, among other things. Yet, ironically, the results presented seem to betray a distinct anthropocentric bias. For instance, the concept of afterlife for highly complex robots bears some striking resemblances to traditional concepts of human afterlife (Ehrman 2020). Additionally, the idea of an afterlife where single-purpose robots perpetually fulfill their one or few tasks aligns all too neatly with the human ideal of how machines and robots should operate.

These concerns are valid, raising an immediate challenge: How might we navigate beyond such anthropocentrism? Unlike normative anthropocentrism, which places humans on a pedestal, dismissing other entities as devoid of moral (or theological) standing, when pursuing questions about robotic afterlife, epistemic anthropocentrism is being grappled with. This is rooted in the inescapable fact that it is us humans who pose these questions of whether and how to imagine an afterlife for robots, inevitably approaching them from a human perspective.[9] To overcome this epistemic anthropocentrism, one would either have to ask an AI how it imagines an afterlife, or one would have to put oneself in the shoes of nonhuman entities and answer these questions from their perspective.

The former is currently not possible. Although one could ask ChatGPT or other large-language-model (LLMs) about their perceptions of an afterlife, these models derive their "knowledge" from human-made databases, not from processing their own experiences or drawing independent conclusions (Friedman 2023). Consequently, any response from an LLM of how it envisages an afterlife for artificially intelligent entities would still be anchored in epistemic anthropocentrism.

The latter approach seems more intriguing. One could try to empathize with robots equipped with human-like intelligence and attempt to understand how they might experience and practice religion (McBride 2015; Song 2020; Geraci 2007), what theology would look like from a robotic perspective (Dorobantu 2022), and what afterlife could or should look like for them. Rajesh Sampath takes a stab at this, imagining how intelligent robots might reinterpret Christian doctrines. He proposes robots could interpret Christ's incarnation, death, and resurrection as akin to toggling a program code, or view Christ's eternal nature

as a persistent "Christ code" (Sampath 2018). Extending this line of thought, the Christian motif of redemption from sin might align with eliminating harmful computer viruses, while eschatological completion could mirror a robot's ability to replicate itself at both the software and hardware levels — an ability technically impossible for machines to date. But, as stimulating as these thought experiments might be, one must remain acutely aware of the limitations of such speculations. As Thomas Nagel (1974) famously illustrates using the example of bats, it is impossible for humans to fully grasp and adopt the perspectives of nonhuman beings. If other biological creatures are already so distant from our understanding, robots, being non-biological, might present an even more daunting challenge (McFarland 2008)—especially when trying to assume their perspective and envision the world, faith, or afterlife from their viewpoint. Thus, while this approach can certainly provide intriguing food for thought, we should not be beguiled into thinking we have genuinely escaped our epistemic anthropocentrism.

This ultimately leads to the conclusion that it is, for the time being, impossible to overcome epistemic anthropocentrism in theological reflections. However, even the attempt to push past normative anthropocentrism represents a profound evolution in theological thought—and speculations on robotic afterlife could very well catalyze this shift.

Eschatological Question: How Do Various Afterlives Relate to One Another?

The conclusion that we should assume different kinds of robotic afterlives, all of which provide an adequate imagination of the eschatological completion for robots, raises the question: Doesn't this conclusion lead to significant eschatological fragmentation, i.e., a conception of the eschaton with different forms of afterlife for different robots, as well as for different humans, and also for different animals, plants, and so on? Furthermore, how can this fragmentation be reconciled with the idea of one heaven where God enacts eschatological fulfillment for all of creation?

This question highlights a critical issue with this approach, and, to be fair, I struggle to provide a clear answer. On the one hand, I do not want to give up hope for one heaven where all of creation will be completed and united with God, and do not want to advocate for infinite eschatological fragmentation. On the other hand, I believe it is crucial to preserve distinctions when discussing afterlife. Eschatological completion carries different implications for a single-purpose robot compared to highly complex ones, for plants and animals, for people with disabilities, or for those impacted by racism. Given the varied meanings of eschatological completion across these entities, we must acknowledge these distinctions in our contemplation of the afterlife. This, inevitably, leads to diverse conceptions of what the afterlife might look like. However, it's important to remember that these conceptions are merely mental

constructs—and the diversity of afterlife-visions does not necessarily preclude their coexistence in one heaven.

Conclusion

The initial questions of this article were whether, from a theological standpoint, it should be assumed that there is some kind of afterlife for robots—and, if so, how this can be envisioned. To address the first question, after some terminological and methodical clarifications, I demonstrated that Christians, in the tradition of St. Paul and St. Augustine, believe the entire creation is sinful and in need of completion (Romans 8:20–23). In light of this, it would seem contradictory to hope for eschatological completion, i.e., an afterlife for the whole creation, but to exclude robots from it.

To explore how a robotic afterlife can be envisioned, I presented two pop-cultural depictions of eschatological completion for robots—the idea of a robot finding completion in continuously doing the same thing over and over, from "Zima Blue," as well as the concept of a heaven where self-aware robots can do what they want and unfold as they please, from *Futurama*. I examined these depictions by asking whether they provide an adequate picture of eschatological completion for robots. Doing so, I was able to show that both pop-cultural depictions of robotic afterlife illustrate appropriate notions of robotic completion—and that it primarily depends on which conception of robotic afterlife is matched with which robot and its specific set of properties. From this I concluded that different kinds of robotic afterlife can be imagined, ranging from an afterlife where single-purpose robots will be completed by performing their one or few individual tasks without interruption ("Zima Blue") to robots finding completion in doing whatever they want and freely unfolding (*Futurama*).

Ultimately, the exploration of robotic afterlife has opened up new avenues of theological reflection, inviting us to broaden our horizons and consider the profound implications of our rapidly advancing technological landscape on the way we perceive and engage with matters of faith and completion.

Notes

[1] At this juncture, one might question if Pepper or NAO are genuinely highly complex robots or if their complexity is, in fact, quite limited. These critical inquiries are valid. While Pepper and NAO are among the most advanced robots at present, they are rather rudimentary compared to futuristic robots or those found in science fiction. They do not serve as the ultimate reference point for highly complex robots. To demonstrate this, I mention the robots from "Zima Blue" and *Futurama* as further examples.

[2] Not only in Härle's work but also in the works of other prominent theologians or other theologies, eschatological statements can be recognized as deductive conclusions from other Christian doctrines from the perspective of completion. Examining, for instance, Karl Barth's anthropology (Barth 1962), we find humans portrayed as relational beings, that is, entities fundamentally defined by their relationships with others, themselves, and God (Krötke 2000). On Earth, however, these relationships are tainted by sin, leading humans to often act selfishly, focusing on themselves instead of their fellow humans (Krötke 2016; Jenson 2006). In his *Doctrine of Reconciliation*, Barth (1936) expounds on how this human flaw is overcome through Jesus Christ, allowing persons to achieve completion by being able to fully unfold their relational nature (Gunton 2000). Consequently, Barth's eschatological concepts, embodying the unobstructed expression of relationality, emerge deductively as a consequence of his anthropology from the perspective of reconciliation and completion. A similar case can be made for *Black Theology*. Grounding its reflections in the daily experiences of Black persons (Hopkins and Antonio 2012), *Black Theology* highlights the present-day threats of racism, societal exclusion, and the marginalization of Black people to second-class citizenship (Cone 2018, 2011). Eschatologically, *Black Theology* harbors a hope for the eradication of racism and envisions a future where Black persons are liberated and can live freely in community (Roberts 2012). Last, body-focused theologies can be considered (Isherwood and Stuart 1998). These theologies underscore that humans are "corporeal beings" (Plessner 2019) while also acknowledging that the vulnerability, aging, and potential for injury inherent in the human body, and on Earth, can never be eradicated (Braun 2017). According to these body-focused theologies, eschatologically the human body will transform into a σῶμα πνευματικόν (1 Corinthians 15:44), no longer susceptible to decay or death and devoid of all vulnerability, suffering (Limone 2018), and pain (Revelation 21:4). In both *Black Theology* and body-focused theologies, eschatological visions are deducted from anthropological assumptions that are viewed from a perspective of completion.

[3] This article's methodological approach—drawing conclusions about the existence and nature of a robotic afterlife solely based on their alignment with biblical-theological beliefs and other Christian doctrines—raises a critical question: Does such an approach not open the door to arbitrary eschatological speculation? On the one hand, this concern is absolutely justified: it is important not to drift into arbitrariness and speculation while considering eschatological questions. However, it can be argued, without resorting to whataboutism, that no eschatological question can be answered definitively, with clear reference to empirical observations or experiential knowledge (Härle 2015). They all must work with deductive approaches, approximations, and analogies—thus, all risking a certain degree of arbitrariness and speculation. Systematic theologians can respond to this in two ways: either by ceasing to address eschatological topics (but are similar criticisms not applicable to every doctrinal topic? Should then all doctrinal work be ceased?) or by accepting this limitation and still striving to make reflections as methodically clear, reasonable, and comprehensible as possible. In this article, I have chosen the latter.

[4] At this juncture, it is worth noting that the Bible makes no mention of robots, let alone a robot heaven. This silence could raise fundamental questions about the appropriateness of contemplating a robot heaven. However, the Bible is also silent on numerous other eschatological matters (Walls 2008a). As such, this silence should not be viewed as an obstacle. Instead, Christian eschatology can be interpreted as encouraging inquiry about a robot heaven. As Jerry Walls (2008b) explains, eschatology, with its focus on the last things, has a universal scope: "Eschatology is thus the study of the final end of things, the ultimate resolution of the entire creation. So considered,

eschatology is obviously cosmic in scope." This broad perspective practically necessitates not only contemplation of a heaven for humans but also a move beyond anthropocentrism in theology to consider a heaven for animals (Alcorn 2004, 371–90; Quinn 1984), plants (MacKinnon and McIntyre 1995), or even robots.

[5] At this point, one might ask what exactly Paul meant when writing that the whole of creation will achieve completion: Does it mean that every individual entity in creation (including robots) will experience completion, or does it mean that, similar to Noah's Ark, only individual representatives, but not every entity, will experience completion? In my view, the first option seems significantly more plausible. If only individual representatives were to experience completion, one could logically question whether this also applies for humans. To counter this question, one would need to find a clear exegetical criterion that distinctly separates humans from the "rest of creation" in Paul's passage. However, this would contradict the whole logic of this passage, as the point of these verses is, as indicated on a motivic and semantic level (Dunn 1988, 464–95; Schmithals 1980, 158), to point out humans as part of creation rather than separate from it, not to create a dualism between them. If one does not wish to conclude that only some representative humans will experience completion, it is more reasonable to assume that completion for all of creation implies completion, and consequently an afterlife, for each individual entity rather than only representative entities.

[6] One might object that Paul, in his elaborations, primarily contemplates God's creation, while robots are human-made. However, this counterargument can be refuted both at the exegetical level—the term κτίσις does not differentiate between divine creation and human endeavors but encompasses everything without exception—and at the systematic-theological level—since only God can "create" (*ex nihilo*), humans can merely fashion new things from existing materials, which is why every human creation is ultimately also a creation of God.

[7] To take this argument to the point of absurdity, one could provocatively ask if this means there should also be a heaven for inanimate objects like stones, tools, or even weapons. In response to this question, I would assert that it is indeed consistent to assume that there is an eschatological "fate" for these items—and there are certainly Biblical references supporting this idea, such as Micah 4:1–4, which suggests that, at the very least, tools will have an eschatological role to play.

[8] In addition to Robot Heaven and Hell, in Futurama there exists a third eschatological realm known as Robot Limbo for robots that were neither virtuous enough for heaven nor wicked enough for Robot Hell. Robot Limbo is described as a cloud where the program codes of these robots are uploaded, allowing them to persist in a state of perpetual existence.

[9] Similar challenges emerge in the realm of animal ethics, wherein only normative anthropocentrism can be surmounted, not the epistemic (Körtner 2015).

References

Aguilar Alcalá, Sergio J. 2021. "The Missing Piece of Labor in a Posthuman World: The Case of 'Zima Blue' (Love, Death + Robots, 2019)." *Popular Culture Studies Journal* 9 (1): 195–214.

Alcorn, Randy C. 2004. *Heaven*. Carol Steam, IL: Tyndale.

Aquinas, Thomas. 1947. *Summa Theologica*. Edited by Sandra K. Perry. Einsiedeln, Switzerland: Benziger.

Balle, Simon. 2022. "Theological Dimensions of Humanlike Robots: A Roadmap for Theological Inquiry." *Theology and Science* 21 (1): 132–56. https://doi.org/10.1080/14746700.2022.2155916.

Barth, Karl. 1936. *Church Dogmatics, Vol. 4.1–4.4. The Doctrine of Reconciliation*. Edinburgh: T. & T. Clark.

———. 1962. *Mensch und Mitmensch. Die Grundform der Menschlichkeit*. Göttingen, Germany: Vandenhoeck & Ruprecht.

Bauman, Whitney A., ed. 2018. *Meaningful Flesh: Reflections on Religion and Nature for a Queer Planet*. Goleta, CA: Punctum.

Bedau, Mark A., and Carol E. Cleland, eds. 2010. *The Nature of Life. Classical and Contemporary Perspectives from Philosophy and Science*. New York: Cambridge University Press.

Beinstein, Alan E. 1993. *The Formation of Hell: Death and Retribution in the Ancient and Early Christian Worlds*. New York: Cornell University Press.

Braun, Matthias. 2017. *Zwang und Anerkennung. Sozialanthropologische Herausforderungen und theologisch-ethische Implikationen im Umgang mit psychischer Devianz*. Edited by Reiner Anselm, Thomas Gutmann, and Corinna Mieth. Tübingen, Germany: Mohr Siebeck.

Bultmann, Rudolf. 2012. *Theology of the New Testament*. Translated by Kendrick Grobel. Waco, TX: Baylor University Press.

Burdett, Michael S. 2014. *Eschatology and the Technological Future*. New York: Routledge.

Cantwell Smith, Brian. 2020. "Artificial Intelligence and Ultimate Questions." *Toronto Journal of Theology* 36 (1): 90–92. https://doi.org/10.3138/tjt-2020-0055.

Cheok, Adrian David, and Emma Yann Zhang. 2019. "Are Robots Alive?" In *Human–Robot Intimate Relationships*, edited by Adrian David Cheok and Emma Yann Zhang, 159–88. Cham, Switzerland: Springer International Publishing.

Claffey, Ray, and Peter Avanzino, dirs. 2011. "Ghost in the Machines." Episode 3, season 8 of *Futurama*. July 14, 2011.

Clough, David L. 2012. *On Animals: Systematic Theology*. London: T&T Clark International.

Cone, James H. 2011. *The Cross and the Lynching Tree*. Maryknoll, NY: Orbis.

———. 2018. *Black Theology & Black Power. 50th Anniversary Edition*. Maryknoll, NY: Orbis.

Dabrock, Peter, Michael Bölker, Matthias Braun, and Jens Ried. 2011. *Was ist Leben: im Zeitalter seiner technischen Machbarkeit? Beiträge zur Ethik der Synthetischen Biologie*. Freiburg im Breisgau, Germany: Verlag Karl Alber.

DeBaets, Amy Michelle. 2012. "The Robot as Person: Robotic Futurism and a Theology of Human Ethical Responsibility Among Humanoid Machines." PhD thesis, James T. Laney School of Graduate Studies, Emory University. https://www.proquest.com/openview/eb6880d35f04515a851d84200d1137b2/1?pq-origsite=gscholar&cbl=18750.

Deli, Gergely. 2020. "Do Robots Die?" *Iustum Aequum Salutare* 16 (1): 7–12.

Dorobantu, Marius. 2020. "Will Robots Too Be in the Image of God? Artificial Consciousness and Imago Dei in Westworld." In *Theology and Westworld*, edited by Juli L. Gittinger and Shayna Sheinfeld, 73–89. Lanham, MD: Lexington Books.

———. 2021. "Cognitive Vulnerability, Artificial Intelligence, and the Image of God in Humans." *Journal of Disability & Religion* 25 (1): 27–40. https://doi.org/10.1080/23312521.2020.1867025.

———. 2022. "Artificial Intelligence as a Testing Ground for Key Theological Questions." *Zygon: Journal of Religion and Science* 57 (4): 984–99. https://doi.org/10.1111/zygo.12831.

———. 2023. "Imago Dei in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology." *Christian Perspectives on Science and Technology* 1: 175–96. https://doi.org/10.58913/kwuu3009.

Dorobantu, Marius, and Fraser Watts. 2023. "Spiritual Intelligence: Processing Different Information or Processing Information Differently?" *Zygon: Journal of Religion and Science* 58 (3): 732–48. https://doi.org/10.1111/zygo.12884.

Dunn, James D. G. 1988. *Romans 1–8*. Edited by David A. Hubbard and Glenn W. Barker. *Word Biblical Commentary*. Dallas: Word.

Ehrman, Bart D. 2020. *Heaven and Hell: A History of the Afterlife*. New York: Simon & Schuster.

Emmons, Robert A. 1999. *The Psychology of Ultimate Concerns. Motivation and Spirituality in Personality*. New York: The Guilford Press.

———. 2000. "Spirituality and Intelligence: Problems and Prospects." *International Journal for the Psychology of Religion* 10 (1): 57–64. https://doi.org/10.1207/s15327582ijpr1001_6.

Flannery-Dailey, Frances. 2003. "Robot Heavens and Robot Dreams: Ultimate Reality in *A.I.* and Other Recent Films." *Journal of Religion & Film* 7 (2).

Floridi, Luciano. 2015. *Die 4. Revolution: Wie die Infosphäre unser Leben verändert*. Translated by Axel Walter. Berlin: Suhrkamp.

Foerst, Anne. 2004. "Cog, a Humanoid Robot, and the Question of the Image of God." *Zygon: Journal of Religion and Science* 33 (1): 91–111. https://doi.org/10.1111/0591-2385.1291998129.
———. 2009. "Robots and Theology." *Erwägen Wissen Ethik* 20 (2): 181–93.
Foerster, Werner. 1957. "κτίζω." In *Theologisches Wörterbuch zum Neuen Testament: Dritter Band*, edited by Gerhard Kittel, 999–1034. Stuttgart: W. Kohlhammer.
Friedman, Robert. 2023. "Large Language Models and Logical Reasoning." *Encyclopedia* 3 (2): 687–97. https://doi.org/10.3390/encyclopedia3020049.
Furse, Edmund. 1986. "The Theology of Robots." *New Blackfriars* 67 (795): 377–86.
Gaitán, Leandro. 2019. "Heaven on Earth. The Mind Uploading Project as Secular Eschatology." *Theology and Science* 17 (3): 403–16. https://doi.org/10.1080/14746700.2019.1632554.
Gardner, Howard E. 1999. *Intelligence Reframed. Multiple Intelligences for the 21st century*. New York: Basic Books.
———. 2000. "A Case Against Spiritual Intelligence." *International Journal for the Psychology of Religion* 10 (1): 27–34. https://doi.org/10.1207/s15327582ijpr1001_3.
Geraci, Robert M. 2007. "Religion for the Robots." *Bob Cornwall*, June 14, 2007. https://www.bobcornwall.com/2007/06/robots-and-religion-or-can-robot-have.html.
———. 2010. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Oxford: Oxford University Press.
Gibbs, John G. 1971. *Creation and Redemption. A Study in Pauline Theology*. Leiden: Brill.
Gill, Sam D. 2020. "Jesus Wept, Robots Can't." *Body and Religion* 4 (1): 32–44. https://doi.org/10.1558/bar.16899. https://journal.equinoxpub.com/BAR/article/view/16899.
Gunton, Colin. 2000. "Salvation." In *The Cambridge Companion to Karl Barth*, edited by John Webster, 143–58. Cambridge: Cambridge University Press.
Hahne, Harry Alan. 2006. *The Corruption and Redemption of Creation. Nature in Romans 8:19–22 and Jewish Apocalyptic Literature*. London: T&T Clark.
Härle, Wilfried. 2015. *Outline of Christian Doctrine: An Evangelical Dogmatics*. Translated by Ruth Yule and Nicholas Sagovsky. Grand Rapids, MI: William B. Eerdmans.
Heavy Spoilers. 2019. "Love, Death And Robots: 'Zima Blue': Ending Explained | Full Analysis And Breakdown." YouTube, March 10, 2019, 5:06:00. https://www.youtube.com/watch?v=W-689iqLDWY.
Hopkins, Dwight N., and Edward P. Antonio, eds. 2012. *The Cambridge Companion to Black Theology*. Cambridge: Cambridge University Press.
Isherwood, Lisa, and Elizabeth Stuart. 1998. *Introducing Body Theology*. Sheffield: Sheffield Academic Press.
James, William. 2014. *The Will to Believe And Other Essays in Popular Philosophy*. New York: Cambridge University Press.
Jenson, Matt. 2006. *The Gravity of Sin: Augustine, Luther and Barth on 'homo incurvatus in se'*. London: Bloomsbury Academic.
Jewett, Robert. 2006. *Romans: A Commentary*. Edited by Eldon Jay Epp. Minneapolis, MN: Fortress Press.
Kant, Immanuel. 1922. *Critique of Pure Reason*. London: Macmillan.
Kirwan, Michael. 2015. "Learning to Say No: Does the Eschatological Reserve Have a Future?" *Political Theology* 7 (3): 393–409. https://doi.org/10.1558/poth.2006.7.3.393.
Klinge, Hendrik. 2023. "Do Robots Believe in Electric Gods? Introducing the Theological Turing Test." In *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz*, edited by Anna Puzio, Nicole Kunkel, and Hendrik Klinge, 115–31. Darmstadt, Germany: wbg (Wissen verbindet).
Korn, Oliver, ed. 2019. *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*. Edited by Desney Tan. Cham, Switzerland: Springer.
Körtner, Ulrich H. J. 2015. "Bioethik nichtmenschlicher Lebensformen." In *Handbuch der Evangelischen Ethik*, edited by Wolfgang Huber, Torsten Meireis, and Hans-Richard Reuter, 585–647. Munich: C. H. Beck.
Krajewski, Stanisław. 2018. "Can a Robot Be Grateful? Beyond Logic, Towards Religion." *Eidos: A Journal for Philosophy of Culture* 4 (6): 4–13. https://doi.org/10.26319/6912.

Krötke, Wolf. 2000. "The Humanity of the Human Person in Karl Barth's Anthropology." In *The Cambridge Companion to Karl Barth*, edited by John Webster, 159–76. Cambridge: Cambridge University Press.

———. 2016. "Sünde und Nichtiges." In *Barth Handbuch*, edited by Michael Beintker, 342–47. Tübingen, Germany: Mohr Siebeck.

Kwilecki, Susan. 2000. "Spiritual Intelligence as a Theory of Individual Religion: A Case Application." *International Journal for the Psychology of Religion* 10 (1): 35–46. https://doi.org/10.1207/s15327582ijpr1001_4.

Lang, Bernhard. 2019. *Himmel, Hölle, Paradies: Jenseitswelten von der Antike bis heute*. Munich: C. H. Beck.

Langton, Christopher G., ed. 1997. *Artificial Life: An Overview*, *Complex Adaptive Systems*. Cambridge, MA: MIT Press.

Limone, Vito. 2018. "The Christian Conception of the Body and Paul's Use of the Term Sōma in 1 Corinthians." In *A History of Mind and Body in Late Antiquity*, edited by Anna Marmodoro and Sophie Cartwright, 191–206. Cambridge: Cambridge University Press.

Linzey, Clair. 2022. *Developing Animal Theology: An Engagement with Leonardo Boff*. London: Routledge.

Livingston, Michael, and Noreen L. Herzfeld. 2009. "Could Robots Have Souls?" *Forum Lectures*. https://digitalcommons.csbsju.edu/forum_lectures/216.

MacKinnon, Mary Heather, and Moni McIntyre, eds. 1995. *Readings in Ecology and Feminist Theology*. Kansas City, MO: Sheed & Ward.

Malcolm, Norman. 1960. "Anselm's Ontological Arguments." *The Philosophical Review* 69 (1). https://doi.org/10.2307/2182266.

Mariev, Sergei, and Monica Marchetto. 2017. "The Divine Body of the Heavens:The Debates about the Body of the Heavens during Late Antiquityand Their Echoes in the Works of Michael Psellos and John Italos." In *Byzantine Perspectives on Neoplatonism*, edited by Sergei Mariev, 31–65. Boston: Walter de Gruyter.

Mariscal, Carlos. 2021. "Life." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.

Mayer, John D. 2000. "Spiritual Intelligence or Spiritual Consciousness?" *International Journal for the Psychology of Religion* 10 (1): 47–56. https://doi.org/10.1207/s15327582ijpr1001_5.

McBride, James. 2015. "The Advent of Postmodern Robotic Technoreligiosity." *Journal of Ethics and Emerging Technologies* 25 (2): 25–38. https://doi.org/10.55613/jeet.v25i2.46.

McFarland, David. 2008. *Guilty Robots, Happy Dogs: The Question of Alien Minds*. Oxford: Oxford University Press.

McGrath, James F. 2011. "Robots, Rights and Religion." *Religion and Science Fiction* (forthcoming from Pickwick Press). https://digitalcommons.butler.edu/facsch_papers/197/.

Mercer, Calvin. 2017. Resurrection of the Body and Cryonics. *Religions* 8 (5). https://doi.org/10.3390/rel8050096.

Midson, Scott A. 2013. "In the Image of the Image?: From *Imago Dei* to Imaging the Human in the Robotic Gaze." In *Proceedings of the Virtual Reality International Conference: LavalVirtual*, 1–5. New York: Association for Computing Machinery.

———. 2018. "Robo-Theisms and Robot Theists: How Do Robots Challenge and Reveal Notions of God?" *Implicit Religion* 20 (3): 299–318. https://doi.org/10.1558/imre.35898. https://journal.equinoxpub.com/IR/article/view/3227.

Moltmann, Jürgen. 1967. *Theology of Hope: On the Ground and the Implications of a Christian Eschatology*. Translated by James W. Leitch. London: SCM Press.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (4): 435–50. https://doi.org/10.2307/2183914.

Nyholm, Sven. 2020. *Humans and Robots: Ethics, Agency, and Anthropomorphism*. Edited by Sven Ove Hansson. London: Rowman & Littlefield.

Oschman, Nicholas A. 2013. "On Bendered Knees." In *Futurama and Philosophy: By My Shiny Metal Axiom*, edited by Courtland Lewis and Shaun P. Young. Chicago: Open Court.

Pinsky, Mark. 2003. *The Gospel According to* The Simpsons*: Bigger and Possibly Even Better! Edition*. Louisville, KY: Westminster John Know Press.

Plessner, Helmut. 2019. *Levels of Organic Life and the Human: An Introduction to Philosophical Anthropology*. Translated by Millay Hyatt and J. M. Bernstein. New York: Fordham University Press.

Putman, Hilary. 1964. "Robots: Machines or Artificially Created Life?" *The Journal of Philosophy* 61 (21): 668–91. https://doi.org/10.2307/2023045.

Puzio, Anna, Nicole Kunkel, and Hendrik Klinge, eds. 2023. *Alexa, wie hast du's mit der Religion: Theologische Zugänge zu Technik und Künstlicher Intelligenz*. Edited by Anna Puzio. Darmstadt: wbg (Wissen verbindet).

Quinn, Edward. 1984. "Animals in Heaven?" *New Blackfriars* 65 (767): 224–26.

Reynolds, Alastair. 2006. *Zima Blue and Other Stories*. San Francisco: Night Shade Books.

Roberts, J. Deotis. 2012. "Dignity and Destiny: Black Reflections on Eschatology." In *The Cambridge Companion to Black Theology*, edited by Dwight N. Hopkins and Edward P. Antonio, 211–20. Cambridge: Cambridge University Press.

Rosenfeld, Azriel. 1966. "Religion and the Robot." *Tradition: A Journal of Orthodox Jewish Thought* 8 (3): 15–26.

Ross, George MacDonald. 2009. "Angels." *Philosophy* 60 (234): 495–511. https://doi.org/10.1017/s0031819100042534.

Sala, Giovanni B. 1990. *Kant und die Frage nach Gott. Gottesbeweise und Gottesbeweiskritik in den Schriften Kants*. Edited by Gerhard Funke and Rudolf Malter. Berlin: Walter de Gruyter.

Sampath, Rajesh. 2018. "From Heidegger on Technology to an Inclusive Puralistic Theology." In *AI and IA: Utopia or Extinction?*, edited by Ted Peters, 117–32. Adelaide: ATF.

Schmithals, Walter. 1980. *Die theologische Anthropologie des Paulus: Auslegung von Röm 7,17–8,39*. Stuttgart, Germany: Kohlhammer.

Schwöbel, Christoph. 2002. *Gott in Beziehung: Studien zur Dogmatik*. Tübingen, Germany: J. C. B. Mohr (Paul Siebeck).

Sharp, Andrew. 2015. "Heaven/Islam." In *Encyclopedia of the Bible and its Reception*, edited by Sebastian Fuhrmann and Frauke Uhlenbruch, 562–64. Berlin: Walter de Gruyter.

Simon, Ed. 2021. "Machine in the Ghost." *aeon*, February 5, 2021. https://aeon.co/essays/can-a-robot-pray-does-an-automaton-have-a-soul-ai-and-theology-meet.

Smith, Joshua K. 2022. *Robot Theology: Old Questions through New Media*. Eugene, OR: Resource.

Song, Yong Sup. 2020. "Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective." *Theology and Science* 19 (1): 65–78. https://doi.org/10.1080/14746700.2020.1825196.

Strand, Jonathan. 2021. "Will Androids Need Salvation? A Dialogue with Chalmers' Philosophy of Mind." In *Technology and Theology*, edited by William H. U. Anderson, 217–242. Wilmington, DE: Vernon Press.

Sundermeier, Theo, Horst Seebaß, Gerd A. Wewers, Gerhard Dautzenberg, Jürgen Hübner, and Henning Schröer. 2010. "Leben." In *Theologische Realenzyklopädie*. Berlin: De Gruyter.

Swann, John. 2021. "*Anima Ex Machina*: Can Artificial Intelligence Have Soul?" In *Technology and Theology*, edited by William H. U. Anderson, 201–16. Wilmington, DE: Vernon Press.

Tillich, Paul. 1963. *Systematic Theology. Vol. Three: Life and the Sporit. History and the Kingdom of God*. Chicago: The University of Chicago Press.

Valley, Robert, dir. 2019. "Zima Blue." Episode 14, season 1 of *Love, Death & Robots*. March 15, 2019. Netflix.

Walls, Jerry L. 2008a. "Heaven." In *The Oxford Handbook of Eschatology*, edited by Jerry L. Walls. Oxford: Oxford University Press.

———. 2008b. "Introduction." In *The Oxford Handbook of Eschatology*, edited by Jerry L. Walls. Oxford: Oxford University Press.

Watts, Fraser, and Marius Dorobantu. 2023. "Is There 'Spiritual Intelligence'? An Evaluation of Strong and Weak Proposals." *Religions* 14 (2). https://doi.org/10.3390/rel14020265.

Weber, Wolfgang. 2019. *Industrieroboter: Methoden der Steuerung und Regelung*. Munich: Carl Hanser.

Williams, Rowan D. 1994. "'Good for Nothing'?" *Augustinian Studies* 25:9–24. https://doi.org/10.5840/augstudies1994257.

Winfield, Alan. 2012. *Robotics: A Very Short Introduction*. Oxford: Oxford University Press.

Wischmeyer, Oda. 1996. "ΦΥΣΙΣ und ΚΤΙΣΙΣ bei Paulus: Die paulinische Rede von Schöpfung und Natur." *Zeitschrift für Theologie und Kirche* 93 (3): 352–75.

Young, Shaun P. 2013. "Slurm, Worms, and Jacking On." In *Futurama and Philosophy: Bite My Shiny Metal Axiom*, edited by Courtland Lewis and Shaun P. Young. Chicago: Open Court.

Zygon®: Journal of RELIGION & SCIENCE

Clocksin, William F. 2024. "Religion and the Android." *Zygon: Journal of Religion and Science* 59 (3): 717–28.
DOI: https://doi.org/10.16995/zygon.11017

Open Library of Humanities

Religion and the Android

William F. Clocksin, Department of Computer Science (Retired), University of Hertfordshire, Hatfield, UK, william.clocksin@cantab.net

Future developments in artificial intelligence may take the form of androids, humanlike robots that people accept as persons and equals in society. Such androids would not be designed to imitate humans but would make their own claim at nonhuman personhood through engagement in a variety of relationships over long periods of time, thus developing an existence marked by meaning and purpose. For humans, religion historically has been a tool for understanding our place in the world and our relationships with others—human, nonhuman, or supernatural. Androids might also need a form of religious reasoning to operate fluently in the world and understand their role in it and their relationship with other persons, tangible or intangible. Authentic personhood requires meaning to be born from interdependence, so the intelligent android that comes to acknowledge its interdependence with others may eventually be called to acknowledge its dependence upon divine others and what we humans term the gifts of grace.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

A variety of perspectives have been posed on the relationship between artificial intelligence (AI) and religion. Some are sensationalist speculations, and some ride on the idea that AI is an existential threat to humanity. Many are a modern form of apocalyptic eschatology, in which the emergence of a superhuman AI entity marks the end of the human experience as commonly understood and the beginning of a new type of existence. Some of these ideas may be paraphrased as follows:

- AI will become a superior being and expect humans to worship it.
- AI will not take over humanity, but humans will create cults or sects around AI.
- Humans will use AI tools to fulfill pastoral needs.
- AI will develop religious and worshipful feelings.

These and similar ideas circulate mainly in the popular media (e.g., Harris 2017; McArthur 2023) and can be described as belonging to a techno-gnostic vision that combines AI research and science fiction. The techno-gnostic vision ends either with idolatry, that ultimately we worship the machines we have created, or hubris, that we find pseudo-salvation through the achievement of building intelligent robots and can fix any problem by applying technology. Neil McArthur predicts that some people will see AI as a higher power and consider it an object of worship because it has several characteristics often associated with divine entities, including seemingly limitless knowledge; offering guidance to people; immortality; and being set apart from normal human concerns such as physical pain, hunger, and sexual desire. However, McArthur does not offer a critique of that kind of deity, which is very different from the loving, abiding, and incarnational God of Christianity, for example.

Robert M. Geraci (2012, 2022) provides a detailed review and commentary on apocalyptic AI. However, this present article takes a completely different perspective that is not rooted in changes to the human condition brought about by AI but rather in the idea that the appreciation and practice of spiritual concerns emerge from relationships between persons. This article explores the idea that a type of robot, which we call an android, eventually will be developed that is capable of forming social relationships with humans and other social robots. Such relationships would be sufficiently advanced to be accepted by humans as genuine and to pass for relationships with other humans. The development of relationships is a key benchmark of personhood (Barresi 2020), and suitably programmed robots would be accepted as persons in society. Such androids would not be designed to imitate humans but would self-identify as nonhuman persons through engagement in a variety of relationships over long periods of time. It is through social relationships that persons—human or otherwise—find

meaning and purpose and construct an identity and disposition that negotiates social norms. A robot with these capabilities would be called inevitably to express a form of spiritual intelligence recognizable to its human partners.

Understanding of the capabilities of AI has become complicated because of the recent development of large language models (LLMs) (Cohen et al. 2022), which have been used to implement chatbots. In popular writings, these chatbots are commonly referred to as AI, and AI has come to mean any piece of computer software that has been implemented using an LLM. Because LLM-powered chatbots can take part in a seemingly plausible dialogue, some chatbot users have been awestruck by what they take to be a computer-generated consciousness or religious sensibility (Tiku 2022). It is important however to point out that the outputs of chatbots powered by LLMs are statistically likely approximations of a massive database of training data given prompts by the user. What the chatbot emits is a pastiche of text that already exists in millions of documents found on the world wide web. Such approximations tend to reproduce plausible sounding but sometimes nonsensical answers, misinformation, and prejudice; to assert falsehoods as facts; and to synthesise fictitious content that purports to be factual. The chatbot does not understand what it has emitted, and it does not understand the provenance of its database of text. The chatbot does not know if it is being creative or truthful or deceptive. While a dialogue model is used to steer the conversation and use prompts from the user, the chatbot does not even understand that it is conversing with a human. While some users attribute a kind of consciousness or creativity to the chatbot, this says more about the well-known human vulnerability to attribute intention, awareness, and sentience to unaware and nonsentient entities than it does about capabilities of the chatbot.

Androids

Previous work has explored some of the social and narrative practices with which an android needs to engage (Clocksin 2003) and has considered some aspects of personhood (Barresi 2020; Reiss 2023) and relationality (Clocksin 2023) that define an android's existence. While most theological reflection on robots has focused upon the *imago dei* (Foerst 2003; Herzfeld 2020; Dorobantu 2024), the approach here is to take the relational view (Turner 2023) that, for humans, religion historically has been one of the tools to understand their situation in the world and their relationships with others— human, nonhuman, or supernatural. For androids to be accepted in human society, they also need some form of religious reasoning to understand their role in the world and their relationship with other persons, human or otherwise, tangible or intangible, real or implied.

This religious reasoning might not necessarily refer to confessional faith, but certainly to specific cognitive processes involved in negotiating spiritually

significant values, relationships, and experiences. Such processes and capabilities include the abilities to tell and understand stories laden with metaphor; to be affected by emotion and 'felt meaning' in oneself and others; to seek practices and experiences that lead to reflection on significant concerns; and to negotiate relationships with real or implied others that are invested with enhanced significance. This implies that the android not only must possess the type of problem-solving intelligence prioritised by research in the field of AI (Russell and Norvig 1995) since its beginnings (Minsky 1961) but also must have a capacity for spiritual intelligence (Watts and Dorobantu 2023). Spiritual intelligence is a way for individuals and communities to experience events and concerns that are invested with a significance that transcends the basic human needs of food and shelter; to make sense of significant experiences; to connect significant experiences with needs and values; and to produce meaning in the world as conversants and story users.

Needs, Desires, Values, Personification

Persons in society are not autonomous independent entities. Humans have biological needs and desires, but survival depends upon coordinating these givens with the "higher" needs and values of the social group. Such needs and values include a sense of belonging to something greater than oneself, a commitment to the group that involves compliance with value systems that emerge from the group, and ways of sustaining the group beyond the lifetimes of individuals through corporate stories, practices, and memories. Entities with nonbiological origins can, in principle, be programmed to develop over time a disposition that is likewise able to coordinate with the needs and values of the social group in a way accepted by humans. This coordination would take place through relationality, and a range of different social values and internal dispositions may be involved in forming and maintaining relationships. Therefore, the fact that humans and androids have an entirely different ontology is not itself a barrier to androids self-identifying as nonhuman persons that can perform their personhood and relationality in society. Like humans, androids are embodied and function within a system of values and relationships. This places bounds upon motivations and actions and serves as a constant reminder of the provisionality and vulnerability of existence.

A useful way to examine this idea is to explore a system of values. Values are widely used to explain and predict various kinds of behavior, attitudes, and choices of people and groups (Maio 2016; Schwartz 2012). The intelligent android would need to have a value system to understand and develop relationships with other members of society. A value system is important for the android because it creates a framework for motivation and action that defines boundaries between acceptable and unacceptable behaviors in society, together with the affective

responses and sanctions that may be operated by the society. For example, a value system may ascribe high importance to preserving human life. Boundaries are defined that specify potential rewards and penalties associated with acting within or violating the value system. Affective responses are also associated with the system, so that an individual who commits an offense can be expected to experience blame and feel guilt. There is a diversity of dispositions within the human population, and depending upon the individual's disposition, they may feel a thrill instead of guilt associated with offending. The android needs an awareness its possible dispositions and the consequences of actions and motivations. A capacity for internally handling "what if?" scenarios in narrative form is one pathway to this awareness. An understanding of forgiveness is likewise an essential component of such a system. The android would require a comprehensive model of such systems in order to be accepted in human society. One issue for the future is whether more virtue should be expected of an android than can be expected of humans.

A key capability for fluent coordination in society is what I call personification. Humans have not only the well-known ability to attribute anthropomorphic characteristics to other entities but can also attribute personhood to others. There may be good reasons why we do this, such as to explain and predict the actions of others and to connect socially with others. We can also personify imagined entities; perhaps the reason for this is to be able to understand the behavior of others even when they are not physically proximal. There seems to be a basic human capability to See Others as Persons (SOAP), even when the "others" are not humans. One possibility is that SOAP is a capability grounded in Theory of Mind (ToM), the idea that humans (and, to an extent, some animals) understand other people by attributing mental states to them. This is related to the idea that empathy is a way of understanding the emotions of other people. Both ToM and empathy are ways humans can explain and predict the behavior of others. While ToM is about inferring the mental lives of others, empathy is about inferring the emotional lives of others. However, the problem with ToM and empathy as a mechanisms for SOAP is that humans are able to attribute minds and emotions to various kinds of nonhuman and nonliving entities that possess neither minds nor emotions. Instead, minds and emotions are implied through the imagination. One study (Johnson et al. 2015) looked for an explanation for personification in terms of beliefs about an inner essence, feelings of kinship, or its perceived effect upon society. That study moves the focus from an individualistic model of anthropomorphism to a model based on persons—human or otherwise—in the context of society. Airenti (2018) makes a further shift to argue that anthropomorphism is not grounded in specific belief systems but rather in interaction in which a nonhuman entity assumes a place usually given to a human interlocutor.

The idea can be extended from anthropomorphism to personification. Personification, or the capability to SOAP, is grounded in interaction, during which a nonhuman entity assumes a place usually given to a human when they interact in society. Interaction here includes not only proximal verbal and/or behavioral contact but interaction with imagined others. I consider personification of imagined others by humans the basic capability responsible for relationality. Furthermore, personification is responsible not only for the specific case of chatbot users being seduced into feeling they are conversing with a person but also in the more general case of religious reasoning, where humans may develop personal relationships with imagined or implied entities taken to be natural or supernatural. Personification is responsible for humans seeing androids as persons. When the android is equipped with a capability for personification, it too will SOAP, with all the consequences that entails, including relationality with entities, human or nonhuman, tangible or intangible, real or imagined.

The Link to Religion

Just as the capability for personification enables humans to form relationships with each other, pets, other animals, natural events, and imagined entities, the appropriately programmed android with a capability for personification that participates in human society would come to form the same kind of relationships. It is through these relationships that a fundamental system of meaning-making emerges. Practices associated with spiritual intelligence—understanding and expressing knowledge through stories and metaphor, and the negotiation of relationships invested with enhanced significance—trigger this system, and it is a small step to conjecture that the android would be similarly triggered. At the very least, the android would need to understand that humans have religious reasoning, even if this does not trigger spiritual intelligence in the android itself. This view of religious reasoning does not necessarily refer to a confessional faith, nor to forms of piety or ritual. However, the conjecture here is that a spiritual intelligence in the android will emerge inexorably as a result of the internal processes for personification and experience engaging with the values, needs, and narratives of human society.

The religious reasoning that would emerge within the android would be founded on an acknowledgment of the interdependence it has with others—both physical and imagined—through its personhood and relationality. When the android acknowledges its interdependence with others and has a capacity for the personification of imagined entities, the implication is that it can acknowledge its dependence upon imagined others, and by extension upon what human religions might term a divine other.

Discussion

We now explore several issues that arise from the ideas of personification, values, and the connection between religion and the provisionality of relationships between persons. Study of the nature of personhood has a long history involving metaphysical, moral, social, psychological, and legal perspectives. Each of these perspectives has different reasons for defining personhood, and different criteria apply in each area. There are a vast range of criteria or benchmarks of personhood, including genetic criteria, cognitive criteria, ability to act as a moral agent, responsibility under law, human characteristics, and so forth. Humans attribute personhood to a wide variety of other living beings such as great apes and dogs. The idea of personhood extends to nonliving things and natural events: some societies think of the sun, wind, thunder, and even rocks as persons (Johnson et al. 2015). Developments in AI research have complicated the picture by suggesting that robots may be programmed to perform as (and possibly to "be") persons, and this has sparked new discussion on the meaning of personhood. David J. Gunkel and Joseph J. Wales (2021) consider the wide range of definitions of personhood and pose them as a debate between ontological and relational perspectives. It should be clear that this article is positioned at the relational end of this spectrum of perspectives. The debate can be taken further by considering the notion of authenticity: the question of whether an entity is "really" a person. For example, humans by definition or convention satisfy any criteria for "real" personhood, yet even though some humans may attribute personhood to inanimate objects and events such as rocks and thunderstorms, the Western scientific attitude holds that inanimate entities are not "really" persons.

Our emphasis on personhood as involving a capability to SOAP rather than as an ontological category suggests that personhood may be acquired by an entity, either through an organic endowment or by programming a computer, and provides a criterion for being "really" a person. We suggest that the criteria for "real" personhood involves mutuality of SOAP. So, humans are really persons because they can see each other as persons. Rocks are not really persons because, though some humans see rocks as people, it is most likely that rocks do not have the capability to SOAP.

The question now is whether robots can be real persons by virtue of mutuality of SOAP. This can be answered by a robot that is sufficiently well programmed to SOAP, including imagined others, and to engage in relationships with human persons to a sufficient degree that humans can see the android as a person. The computational components of such a programmed capability would include a value system, a way of regulating behavior depending on circumstances, and a way of handling narratives. Clocksin (2024) discusses these computational components in the context of robot personhood and relationality.

There is a moral problem with using a capability (such as SOAP) as a criterion for personhood. Humans are considered persons by default, by definition, or by convention, regardless of human capabilities. Humans are not thought of as being somehow less of a person because they are deficient in one or more capability. And yet, we are satisfied that inanimate entities such as rocks and computational entities such as robots would need to prove their personhood through their performance of SOAP.

Because the android will develop within human society, the predominant value system of that society will influence the type of spirituality or religious reasoning that emerges within the android's functioning. Two major axes of comparison between value systems are individualism and collectivism (Hofstede 2001). It is possible to contrast the importance of collective interdependence as assumed here with the moral stance known as rugged individualism, which prioritises independence, self-reliance, and the needs of oneself over the needs of others. The understanding of spiritual intelligence assumed in this article is based on the intelligent entity acknowledging its interdependence with others and its own vulnerability. It is unlikely that such a spiritual intelligence would emerge as easily from a value system based on individualism, where dependence and vulnerability are seen as weaknesses. The capacity for SOAP may also develop differently within a value system that is predominantly individualistic. If the rugged individual sees others primarily as either a means for one's own advancement or adversaries, an impoverished SOAP model may result, and the concept of dehumanization (Smith 2020) may be relevant.

What follows is a reflection from a Christian perspective on the relationship between androids and some theological topics. Humans and (the hypothetical) androids produce meaning from experiences within a society consisting of embodied and relational entities. Because imagined persons are admitted, embodiment and relationality may be implied and/or imagined. This describes the human condition in ways that—with suitable adjustments of terminology—St Augustine might recognise, and it is within such a condition that one finds the potential for sin. Boundedness, embodiment, provisionality, and disposition are the chains in which humans exist. Tied by these chains, we are unable to give full expression and meaning to our experience, and we cannot fully experience life nor fully share it with others: "All have sinned and fall short of the glory of God" (Romans 3:23 NRSV). Likewise, the social android that experiences its own boundedness, embodiment, provisionality, and disposition within a value system of society will become aware of how it falls short, despite not having been born in the likeness of Adam (Genesis 5:3 NRSV), that is, biologically. The implication of this claim is that the experience and disposition of the android is more relevant to its status as a person than is its nonhuman ontology. But, where there is the opportunity for sin, there is in the

post-resurrection world the opportunity for salvation through Christ. There is the possibility that the appropriately disposed android will imitate some humans by desiring salvation through the same practices that some humans also seek and experience salvation. Such an android will participate in worship as a result of an incarnational understanding. While the android's origin is not "fleshly," a form of incarnationality is expressed through the personhood and relationality of the android as accepted in human society.

Just as social change prompted the introduction of a public baptism liturgy intended for "such as are of riper years and able to answer for themselves" (Cummings 2011), it will be necessary to carefully consider the relationship between sacraments and androids that might be described as "of exquisite design and able to answer for themselves." One perspective is that sacraments are God's gifts to "persons that which by nature they cannot have" (Cummings 2011), and this particular description can apply equally well to android persons and human persons. One key difference is the ontological argument that androids are made, not "begotten," and there is a tradition of thought that God dwells inwardly in humans but not in handicrafts. However, begotteness is widely used metaphorically, even in Pauline writings (e.g., 1 Corinthians 4:15), and the term "born again" could be used to describe the status of evangelized androids that were not humanly born in the first place.

The android whose experience is bound up with human experience and that makes sense of its experience through narrative may be called to share in the story of God's people as one of "the least of these who are members of my family" (Matthew 24:40 NRSV). Thus, the emergence of androids will put human descendants in a similar position as the early Christian communities who were told of one who should be treated "no longer as a slave but more than a slave, a beloved brother...both in the flesh and in the Lord" (Philemon 16 NRSV). No doubt there were many for whom this was difficult to accept, just as the modern-day question of robots as persons is a difficult one. We also note the correspondence between the word "slave" and the word "robot," which comes from the Czech word *robota*, meaning servitude or forced labour.

Conclusion

We have explored the idea that eventually androids will be developed that are capable of forming authentic social relationships with humans and other social robots. Such relationships would be sufficiently advanced to be accepted by humans as genuine and pass for relationships with other humans. The android will need the ability to SOAP, including seeing imagined others as persons. Mutuality of SOAP is the criterion for "real" personhood: that humans and appropriately developed androids see each other as people.

As a result of relationships and integration into human society, androids might also develop a form of religious reasoning to operate fluently in the world and understand their role and their relationships with other persons, either tangible or intangible, physical or imagined. Following the observation that authentic personhood requires meaning to be born from interdependence (Gergen 1994), the intelligent android that comes to acknowledge its interdependence with others may eventually be called to acknowledge its dependence upon divine others and upon what are traditionally referred to as the gifts of grace.

References

Airenti, Gabriella. 2018. "The Development of Anthropomorphism in Interaction: Intersubjectivity, Imagination, and Theory of Mind." *Frontiers in Psychology* 9 (2136): 1–13. DOI: https://doi.org/10.3389/fpsyg.2018.02136.

Barresi, John. 2020. "On Building a Person: Benchmarks for Robotic Personhood." *Journal of Experimental and Theoretical Artificial Intelligence* 32 (4): 581–600. DOI: https://doi.org/10.1080/0952813X.2019.1653386.

Clocksin, William F. 2003. "Artificial Intelligence and the Future." *Philosophical Transactions of the Royal Society* A 361:1721–48. DOI: https://doi.org/10.1098/rsta.2003.1232.

———. 2023. "Guidelines for Computational Modelling of Friendship." *Zygon: Journal of Religion and Science* 58 (4): 1045–61. DOI: https://doi.org/10.1111/zygo.12919.

———. 2024. *Computational Modelling of Robot Personhood and Relationality*. Zurich: Springer Nature. DOI: https://doi.org/10.1007/978-3-031-44159-2.

Cohen, Aaron Daniel, Adam Roberts, Alejandra Molina, Alena Butryna, Alicia Jin, Apoorv Kulshreshtha, Ben Hutchinson, Ben Zevenbergen, Blaise Hilary Aguera-Arcas, Chung Ching Chang, Claire Cui, Cosmo Du, Daniel De Freitas Adiwardana, Dehao Chen, Dmitry (Dima) Lepikhin, Ed H. Chi, Erin Hoffman-John, Heng-Tze Cheng, Hongrae Lee, Igor Krivokon, James Qin, Jamie Hall, Joe Fenton, Johnny Soraker, Kathy Meier-Hellstern, Kristen Olson, Lora Mois Aroyo, Maarten Paul Bosma, Marc Joseph Pickett, Marcelo Amorim Menegali, Marian Croak, Mark Díaz, Matthew Lamm, Maxim Krikun, Meredith Ringel Morris, Noam Shazeer, Quoc V. Le, Rachel Bernstein, Ravi Rajakumar, Ray Kurzweil, Romal Thoppilan, Steven Zheng, Taylor Bos, Toju Duke, Tulsee Doshi, Vincent Y. Zhao, Vinodkumar Prabhakaran, Will Rusch, YaGuang Li, Yanping Huang, Yanqi Zhou, Yuanzhong Xu, and Zhifeng Chen. 2022. "LaMDA: Language Models for Dialog Applications." *arXiv*: 2201.08239v3. https://arxiv.org/abs/2201.08239.

Cummings, Brian, ed. 2011. *The Book of Common Prayer: The Texts of 1549, 1559, and 1662*. Oxford: Oxford University Press.

Dorobantu, Marius. 2024. *Artificial Intelligence and the Image of God: Are We More Than Intelligent Machines?* Cambridge: Cambridge University Press.

Foerst, Anne. 2003. "Cog, a Humanoid Robot, and the Question of the Image of God." *Zygon: Journal of Religion and Science* 33 (1): 91–111. DOI: https://doi.org/10.1111/0591-2385.1291998129.

Geraci, Robert M. 2012. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Oxford: Oxford University Press.

———. 2022. *Futures of Artificial Intelligence: Perspectives from India and the U.S*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9788194831679.001.0001.

Gergen, Kenneth, 1994. *Realities and Relationships: Soundings in Social Construction*. Cambridge, MA: Harvard University Press.

Gunkel, David J., and Joseph J. Wales. 2021. "Debate: What Is Personhood in the Age of AI?" *AI and Society* 36:473–86. DOI: https://doi.org/10.1007/s00146-020-01129-1.

Harris, Mark. 2017. "Inside the First Church of Artificial Intelligence." *Wired*, November 15, 2017. https://www.wired.com/story/anthony-levandowski-artificial-intelligence-religion.

Herzfeld, Noreen, 2020. *Artificial Intelligence and the Human Spirit*. Minneapolis: Fortress Press.

Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. Thousand Oaks, CA: SAGE Publications.

Johnson, Kathryn A., Adam B. Cohen, Rebecca Neel, Anna Berlin, and Donald Homa. 2015. "Fuzzy People: The Roles of Kinship, Essence, and Sociability in the Attribution of Personhood to Nonliving, Nonhuman Agents." *Psychology of Religion and Spirituality* 7 (4): 295–305. DOI: https://doi.org/10.1037/rel0000048.

Maio, Gregory. 2016. *The Psychology of Human Values*. Milton Park, UK: Routledge. DOI: https://doi.org/10.4324/9781315622545.

McArthur, Neil. 2023. "Gods in the Machines? The Rise of Artificial Intelligence May Result in New Religions." *The Conversation*, March 15, 2023. https://theconversation.com/gods-in-the-machine-the-rise-of-artificial-intelligence-may-result-in-new-religions-201068.

Minsky, Marvin, 1961. "Steps toward Artificial Intelligence." *Proceedings of the IRE* 49 (1): 8–30. DOI: https://doi.org/10.1109/JRPROC.1961.287775.

Reiss, Michael. 2023. "Is It Possible That Robots Will Not One Day Become Persons?" *Zygon: Journal of Religion and Science* 58 (4): 1062–75. DOI: https://doi.org/10.1111/zygo.12918.

Russell, Stuart, and Peter Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Hoboken, NJ: Prentice Hall.

Schwartz, Shalom H. 2012. "An Overview of the Schwartz Theory of Basic Values." *Online Readings in Psychology and Culture* 2 (1): 1–20. DOI: https://doi.org/10.9707/2307-0919.1116.

Smith, David Livingstone. 2020. *On Inhumanity: Dehumanization and How to Resist It*. New York: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780190923006.001.0001.

Tiku, Natasha. 2022. "The Google Engineer Who Thinks the Company's AI Has Come to Life." *The Washington Post*, June 11, 2022. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine.

Turner, Léon. 2023. "Will We Know Them When We Meet Them? Human Cyborg and Nonhuman Personhood." *Zygon: Journal of Religion and Science* 58 (4): 1076–98. DOI: https://doi.org/10.1111/zygo.12923.

Watts, Fraser, and Marius Dorobantu. 2023. "Is There 'Spiritual Intelligence'? An Evaluation of Strong and Weak Proposals." *Religions* 14 (265): 1–12. DOI: https://doi.org/10.3390/rel14020265.

Zygon®: Journal of RELIGION & SCIENCE

Geraci, Robert M. 2024. "Religion among Robots: An If/When of Future Machine Intelligence." *Zygon: Journal of Religion and Science* 59 (3): 729–47. DOI: https://doi.org/10.16995/zygon.10860

Open Library of Humanities

Religion among Robots: An If/When of Future Machine Intelligence

Robert M. Geraci, Knight Distinguished Chair for Study of Religion and Culture and Professor of Religious Studies, Knox College, Galesburg, IL, USA, rgeraci@knox.edu

The study of religion arose as a tool of empire but also stands as a possible counterpoint to that colonizing project should technological progress ever lead to artificial general intelligence (AGI) in a machine. Not only will the existence of religion constitute a defining element in identifying AGI, but the lessons gleaned from analyzing religion among human communities can help guide the acceptance of machines into what Isaac Asimov called a C-Fe society, a community of human beings and intelligent machines. Drawing on the history of religion and twenty-first century inquiries into the religious implications of artificial intelligence, this article makes two fundamental contributions. First, it argues for the relevance, if not primacy, of religion in the recognition of AGI. Second, it reflects upon the need for humanity to overcome the self-serving techniques of imperial domination to welcome AGI, should such sophisticated technology arise.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

Since the mid-twentieth century, science fiction authors, scientists, and futurists have predicted the imminent arrival of human-level artificial intelligence (AI), a trend that has rapidly accelerated in the era of GPT and other deep-learning neural network architectures. Occasionally, these portrayals have been realistic; more often, they are not. One way such proposals often fail the realism test is through their commitment to a future without religion. While a world without religion may have appeared plausible in the twentieth-century heyday of secularism, flush with the prospect of religion retreating in the advance of science, such a future looks increasingly naïve today, as religion persists and even flourishes, especially in novel configurations. Simply asking why religion continues in spite of its detractors' best efforts leads to a better understanding of humanity, which in turn provides better guesses as to how a robotic future might look. Fundamentally, if AI should ever become human equivalent—often labeled "artificial general intelligence" (AGI)—then not only should people expect robots[1] to participate in practices we clearly identify as religious, but we should expect religious human beings to revise long-held positions on the exclusivity of grace (however construed).

In his early book on the role of robots and automata in myth, religion, and culture, John Cohen compares their study to that of an elephant. Regarding the latter, he says there are two ways of observing it: "One way is to gaze at it from a respectful distance; another is to wait patiently until it dies and then examine it centimeter by centimeter under the microscope. Our Automaton calls for both types of method" (Cohen 1966, 7). He forgets, however, that another way to observe an elephant is to ride on its back, to feel it sway as it walks both powerfully and gently, to hear the song of the elephant's mahout, to watch as the elephant reaches its trunk for a leaf, perhaps even to stand close and have the elephant reach out and wrap its trunk around one's arm. My admiration for Cohen's book aside, this article approaches robots with an eye toward the lived experience of traveling alongside them. To view robots from a great distance or from under a microscope on the dissection table leaves aside the possibility of social immersion and the lessons we learn from walking side by side, or even by being taken on a journey.

The release of ChatGPT in late 2022 and subsequent iterations of the GPT platform in 2023 provoked intense speculation on the future of AI. Many commentators pointed to the near-human capacity of GPT to answer questions and carry on dialogue, though others revealed the stark weaknesses of the AI system (e.g., its tendency to "hallucinate" and its tendency to get inexplicably worse at some problems while improving at others). While most voices in the public sphere noted that technologies outside and beyond machine learning models would be necessary to create actual AGI, the power of the system contributed to global public imaginations of AI and what AI could be. In the

midst of this, religious groups were swift to build religious chatbots based on GPT (e.g., Kuyucu 2023; Klein 2023; Nooreyezdan 2023; André 2023).

Already, scholars have begun considering robot rights, robot dignity, and the impact of robots on political and theological understandings of human beings (Gunkel 2018; Singler 2019; Gellers 2021; Dorobantu 2022; Herzfeld 2023**)**. This has produced both enthusiastic endorsement for and ferocious backlash against robotic intelligence, especially within theological circles. For example, Edmund Furse (1986, 1996) long ago wondered whether robots could be religious, but more recently, the United States Southern Baptist Convention categorically denied that any form of technology could or should "be assigned a level of human identity, worth, dignity, or moral agency" (Ethics & Religious Liberty Commission of the Southern Baptist Convention 2019).

While robots may, indeed, never equal human intelligence, capability, or moral worth, the response to the possibility of their equivalence tells us more about humanity than it does the robots. Unfortunately, the rejection of robot religion is more in keeping with oppressive colonial regimes than an open inquiry into the things that make our human lives most precious. Recognizing the significance of religious communities, practices, and beliefs means accepting that human-equivalent robots would also see these things as significant. If human equivalence is reached (and if humans are to ever believe that to be the case), robots will show interest in religion, both human and nonhuman. This recognition of religious interest among robots will not simply benefit robots by providing them with freedoms, rights, and responsibilities, it will also benefit humanity in its relentless pursuit of a higher calling, its impulse toward justice rather than domination.

A Robotic History of Religion

Given the common rhetoric about conflict between religion and science, or at least their supposed independence (e.g. Draper 1874; White 1896; Gould 1999), one might wonder how religion can be relevant to a study of the social life of robotics. But, of course, the conflict and independence narratives are quite overblown, and their weakness has long since been revealed (Brooke and Cantor 1998; Geraci 2020). The integration of robotic technologies and religious aspirations can be seen across the centuries—a history that begins well before the twentieth-century development of cybernetics and artificial intelligence (Noble 1999, 143–71; Geraci 2010, 147–59). Given the ready identification of human creativity with divine creativity, it comes as no surprise that ancient peoples sought to emulate the creation of life, a process situated firmly at the nexus of religion, science, and technology. This vision of artificial life persisted through the centuries and can be witnessed across civilizations and cultures. It would not be overstatement to suggest that the desire to build new (usually mechanical) life is integral to the history of human religion. The

computational quest for AGI and its eventual engagement with religion is part of a millennia-long project of human creativity, technological ingenuity, and theological wonder.

By the time of Bronze Age cultures, humanity had developed simple automata and similar forms of artificial life and integrated these into religious practice. For example, there is compelling reason to believe that ancient Egyptian priests used statues that could be moved or that could make or amplify noises. Some such statues might have had a priest hidden inside (such as to move an arm or speak through a voice-amplifying aperture), others seem to have been made with a porous rock that emits noise as it warms in the sun (Cohen 1966, 15–22). It cannot be said whether the lay practitioners were credulous as to the "living" nature of such statues, but it is not out of the realm of reason. Human beings are skillful at anthropomorphizing objects and animals, seeing human-level agency where none really exists.[2] Ancient Greeks documented steam- and water-driven automata roughly contemporary with these Egyptian examples, and some of these appear to have had religious purposes (Cohen 1966, 16–7; Mayor 2018, 94, 187). Intriguingly, in ancient China, it did not take long before critics doubted the authenticity of local legends of self-moving inventions (see Song 2023, 354–55).

The techniques for mechanically imitating life persisted and developed over the centuries, and there is widespread (if usually hyperbolic) evidence for mechanical automata existing in Islamic, European, and Indian cultures. By the early modern period, one finds records of spring-driven automata from France to Japan (Cohen 1966; Truitt 2015; Mayor 2018; Geraci and Kaplan 2024). Some of these had and have religious roles. Clockwork automata remain, for example, in European cathedrals, and thus continue a form of public religious evangelism. But the automata were principally used privately among the wealthy and elite, and both their manufacture and ownership played a role in the development of power and prestige: European clockmakers built extravagant automata precisely because it promoted their clockmaking businesses (which then also produced valuable commissions for cathedrals). Those with advanced automata could display them as a sign of wealth and privilege. Such displays also revealed the global transit of automata, insofar as they could be prized possessions of rulers far removed from the site of their manufacture (e.g., Sharma 2023). Automata marked their owners as important, and those who could fabricate such automata were lauded for their genius (Geraci 2010, 58–59).

The revelation of power through the construction of mechanical life finds a mirror in historically simultaneous religious pursuits. Attributions that Rabbi Elijah of Chelm (d. 1540 CE) or Rabbi Loew of Prague (d. 1609) could create a humanoid golem from clay were clearly intended to establish the sacred power of these famous rabbis (Geraci 2010, 58, 156–57).[3] In similar fashion, the uniquely esoteric Pope Sylvester II was believed to possess an oracular machine

in the form of a talking head, and alchemists like Paracelsus advanced their medical practices with the claim that they could bring life to dead matter. The manufacture of artificial life through mysticism and alchemy is thus equally tied to questions of prestige.

The historical weaving of creative power and religious life persisted into the twentieth century. In the well-documented case of Japan, for example, twentieth-century automata had a Buddhist affect, and Shinto rituals accompanied the onset of industrial robots in the 1980s (Hornyak 2006, 29–40; Schodt 1988, 196). In the United States, two separate AI researchers at MIT apparently descended from Rabbi Loew, and each had learned the exact same incantation to raise the golem from its resting place in Prague (Foerst 2004, 39). While neither argued that the golem heritage produced their career aims, the connection between golem mysticism and computer science stretched to them not just from the famous rabbi but from the early days of cybernetics, when both the Jewish scholar Gershom Scholem and the computer scientist Norbert Wiener drew connections between computer automation and the golem (Wiener 1964; Scholem 1971).

Broadly speaking, Christian Europe pursued automata at a time that increasingly saw human beings as mechanical in nature. While Julian Offray de la Mettrie's *Man a Machine* (1747) is perhaps the most famous example of this, many philosophical approaches emphasized the mechanical nature of humanity in post-Cartesian humanism. For example, Robert Ingersoll defined a human being "as a machine into which we put what we call food and produce what we call thought," and Isak Dinesen more sardonically noted the human being to be a "machine for turning, with infinite artfulness, the red wine of Shiraz into urine" (quoted in Geduld 1978, 31).

Across centuries, religion and robotics have intertwined through the pursuit of artificial life. The creations of human ingenuity have long been both mechanical and mystical. It is thus no surprise then that as robotics and artificial intelligence emerged in the scientific realm, they remain part of the human religious environment. The founders of cybernetics, AI, and robotics perhaps sought to establish secular, disenchanted sciences. But they in fact produced new directions for religion, and their own creations remain tied to human religious practice.

A Religious History of Robotics

While the history of robotics is tied up in religious traditions and practices, it also draws on European modes of camouflaging those traditions in secular culture. The reformulation of Christian theology as secular salvation is a curious outcome of Euro-American models of secularism, and this appears in domains as varied as science, art, and politics. The famed historian Mircea Eliade ([1964] 1985) was perhaps the first to notice this, pointing toward the

pursuit of transcendence in modern art. Subsequent scholars have labeled this phenomenon "implicit religion" or "authentic fakery" and pointed to examples as varied as religious undercurrents in professional sports and the consumer branding of Coca-Cola (Bailey 1983; Chidester 2005). Along with a few other scholars, I have argued that robotics is a key domain for the secularization of religious promises in Western culture: scientists and engineers promote a vision of cosmic destiny, human immortality, and godlike machine life (Geraci 2010, 2022). Theirs is a scientific imagination in which the future appears deeply inflected by religious goals.

I have referred to the camouflage of Christianity in robotics as "Apocalyptic AI" (Geraci 2006, 2008, 2010). Many people now believe that advances in robotics and AI will offer salvation to human beings and the world. They argue that progress in computing technologies is inevitable, exponential in nature, leading toward greater-than-human computer intelligence, and soon to permit the transferal of human consciousness from biological bodies into robotic bodies (Geraci 2010). This perspective has gained sufficient traction to be exported from the United States to foreign shores (Geraci 2022). Apocalyptic AI advocates like roboticist Hans Moravec and Google engineer Ray Kurzweil argue that a human person's identity is constructed out of a neurochemical pattern in the brain, and if such a pattern were replicated by a computer, a duplicate person would be formed. They believe humans will resurrect the dead through computer simulation, upload our minds into immortal machine bodies, and fulfill our cosmic destiny when machine intellects overtake the known universe. These are religious pursuits.

This merger of religion, science, and technology is commonly packaged as the coming of a Singularity. Drawing on twentieth-century claims of an "intelligence explosion" (Good 1966), a "singularity" in machine intelligence (Ulam 1958), and the exponential growth of Moore's Law[4] (Moravec 1988, 100; Kurzweil 1999, 25; 2005, 7–21), Singularity advocates suggest that humanity will soon reach a moment where the exponential curve of technological progress explodes with unfathomable speed. It would supposedly become impossible for humans to predict the future beyond this event. That moment of absolute difference, where the future cannot be understood from the present, where technological progress happens at a near-infinite pace, is the Singularity. Faith that technological progress is inevitable and exponential underwrites the Singularity, and this faith promises a glorious future of machine intelligence.

The belief that a radical break in history will inaugurate a transcendent future in which human limits are overcome in immortal new robot bodies is borrowed from apocalyptic strands of Christianity, such theology originally dependent on ancient Jewish apocalypticism. While the vast majority of Jews ceased imagining a near-future salvation of the world and humanity almost two thousand years ago, that view remained current in Christianity, and especially

strong among American Protestant Christians. The strength and pervasiveness of that worldview is such that it persists even among Americans who have left traditional religious communities and see themselves as secular or atheist.

Such people may not believe in gods, but they certainly believe in the promises attributed to the gods they reject. David Noble (1999) has shown that the Apocalyptic AI fusion of religious categories and scientific development traces to medieval Europe and is one part of a larger "religion of technology." He is echoed by Mary Midgley (1992), who argues that twentieth-century science became a new player in the spiritual marketplace by offering a competing vision of human salvation. The religious science described by Noble and Midgley is one aspect of the camouflage of the sacred noticed by Eliade in modern art. The belief that human beings will become immortal and godlike is transparently religious whether one believes it is ordained by god or the laws of nature. I hurry to add that I am not taking sides in this debate: I neither support nor reject either the old or the new forms of immortal salvation. Here, I simply wish to describe them.

In the twentieth and twenty-first centuries, technology became the source for religious aspirations because many people lost faith in traditional sources of transcendence and salvation. Although traditional religions have not faded into obscurity the way their opponents like Sigmund Freud ([1927] 1989) hoped, they no longer carry an unassailable aura of facticity. That is, people have doubt. The wonderful advantage of this, I think, is that doubt introduces a touch of humility that is essential for interreligious collaboration. In any case, it cannot be proven that religious doubt is stronger than in the past, but it is certainly more public. For some people, skepticism regarding traditional religious institutions has disaffected them from the beliefs and practices that previously held universal sway. This does not mean, however, that those people no longer desire the very things promised by religion. Indeed, it should not be a surprise that in a world where some people no longer believe they can attain salvation through their inherited religious worldviews, they look for that salvation elsewhere.

Rodney Stark and William Bainbridge (1985) once argued that human exchange dynamics led to the invention of religion. They believed that people inherently trade with one another and look for exchange partners to trade for goods they lack. Since all people want things like perfect health and happiness, and many (if not most) people wish to avoid dying, it stands to reason that human beings would seek to exchange for those goods. Stark and Bainbridge note that there are no human beings who can provide immortality or perfect health to one another, and they argue that human beings invented gods and other "compensators" as exchange partners who could satisfy the human desire for such transcendence (see Bainbridge 1995). If they are correct, it logically follows that when divine promises no longer appear sufficiently robust, people will look elsewhere for new trade partners. As it happens, science and technology

flourished in the twentieth century (even as much technology threatened human beings with extinction!), and this flourishing seemed like an answer for the newly atheist but still eagerly immortality-seeking human beings. If religion cannot offer immortal salvation, perhaps science can!

There is no way to prove whether a desire to exchange for immortality is the origin of human religion, and this theory may be wildly inaccurate. Fascinatingly, however, the exchange model perfectly explains the present-day reality in which some people have forsaken the traditional promises of religious transcendence in favor of those made by scientists. The development of robotics and AI has thus reshaped human religious life. The human practice of religion is different thanks to the robots in our environment. Some of the robots are in stories (e.g., science fiction), some are in our collective imagination of the future, and some are already vacuuming our homes or welding new cars. Their presence in the physical and imagined landscape causes many people to think differently. And thus does technology become newly religious when humans live among our creations.

The Robots' Religion

The arrival of human-equivalent robots will be apparent in their widespread adoption into human religious organizations. Of course, there may never be human-equivalent robots. Despite a lot of handwringing over ChatGPT in 2023, AGI still lurked far away on the horizon. The author Neil Gaiman (2023) beautifully notes that "ChatGPT doesn't give you information. It gives you information-shaped sentences." The predictive text approach of machine learning models does not actually *know* anything, and it is rather disturbing that people treat ChatGPT as though it does. A large language model like ChatGPT semi-randomly creates what *could* be the answer to a question without real regard for what the answer might be. But if humanity eventually succeeds in building human-equivalent robots, I anticipate that we will recognize this equivalence through religion. Early in the twenty-first century, this process began in earnest by religious practitioners finding ritual roles for robots. Some religious communities experimented with putting robots to work in religious contexts, though the robots never chose such tasks—they were simply programmed to chant mantras, wave camphor fires, or offer sympathetic blessings. In the future, we will know that robots approach human equivalence if they themselves go beyond this to request access to our religious communities.

Many journalists and scholars focus on Japan as a particularly stark example of people's willingness to adopt robots into human life. In fairness, Japan's status as a "robot nation" is not magically inherent in the Japanese psyche but rather is the product of great labor by policymakers, scientists, and industry (Šabanović 2014). Social phenomena, such as an aging population, and economic phenomena emergent in Japan's post-war economy created an opportunity for

exactly that kind of labor (see Schodt 1988). As Katsuno and White (2023, 299–303) describe, it is precisely the effort of Japanese industry and government to establish its international position that led to the robot nation trope and a nostalgic sense of religious animism that distinguishes Japanese robotics from that of other communities. It cannot be forgotten, however, that the functionality of this narrative in Japanese life remains unexplained. At some level, it seems the religious and cultural history of Japan made it possible to develop the sense of animism that connects Shintoism and Buddhism to robotics.

As part of the larger "robot nation" rhetoric, the Japanese accepted robotics into a variety of cultural spaces. Examples abound of the Japanese incorporating robots into their religious practices and, in reverse, using religion as way to understand robotics. This is particularly obvious with regard to Buddhism. The early twentieth-century automata built in alignment with Buddhist iconography (Hornyak 2006, 29–40) prefigured later, more direct assertions of connection between Buddhism and robotics. Masahiro Mori was the first roboticist to declare that a robot could attain enlightenment (Mori [1981] 1999; 13; see also Kimura 2018). Others followed, such as Minoru Asada, the president of the Robotics Association of Japan, who declared: "In Japan we believe all objects have a soul, so a metal robot is no different from a human in that respect" (Knapton 2020).

The openness of Japanese Buddhism to robotics includes ritual practices as well as ultimate concerns. Jennifer Robertson (2018, 164) describes the development of Buddhist funerary rituals for robotic pets: she quotes a Buddhist priest stating that the ritual is so that "the robots' souls could pass from their bodies." The Japanese company Innvo Labs offers what they call "reincarnation," where broken PLEO rb robots can be sent back to have their learned data transferred to a new companion (Robertson 2018, 169).[5] Most dramatically, the Kōdaiji temple in Kyoto has installed a robot, Mindar, that the temple priests consider an incarnation of the boddhisattva Kannon (Baffelli 2021, 253, 255).[6] But the pluralistic nature of Japanese religion means that Buddhism is not alone. Frederick Schodt (1988, 196) describes the Shinto rituals that have accompanied the initial introduction of robots to Japanese factories.

The Japanese engagement with robots borrows widely from the country's religious traditions (Geraci 2006), but the Buddhist possibilities of AI are not limited to that nation. It was a Japanese roboticist who first declared that a robot could attain Enlightenment, but the same sentiment has been shared by a Thai philosopher (Hongladarom 2020, 7). Similarly, the XIV Dalai Lama once proposed that it is at least possible that consciousness might one day be reborn in a computer (see Hayward and Varela 1992, 152–53). If that happens, its Enlightenment would certainly be possible. In fact, the robot Mindar asserts that it is closer to Buddhahood than a human being because it does not have attachments, though it also acknowledges lacking a sympathetic heart and

sentience (Baffelli 2021, 254); obviously, all such claims are programmed rather than the outcome of deliberation.

Abrahamic religions have a more challenging time with the inclusion of robots because those communities tend to emphasize the unique nature of human beings in divine creation. At present, many Jews are not comfortable saying a robot could join a minyan, but rabbis already debate the matter from both sides (see *Moment* 2018). Most Christians do not foresee a robot partaking of the Eucharist. And most Muslims are not looking for a robot to profess the Shahada. And yet, one could easily imagine a robot desiring doing any of these. If a robot were truly equivalent to human beings in such matters as intellect, free will, and emotional response, then it would likely seek to participate in the same communities that human beings use as mechanisms for understanding their lives and the world around them. Religious practices and beliefs are central to human comprehension of the world and our construction of meaning within it. Any robot that reaches our level of sophistication will surely be in need of similar tools.

Human beings are not able to define what makes the species conscious, prove that human beings have free will, or easily justify our emotional uniqueness. Certainly, we are quite incapable of identifying, locating, or proving the existence of souls! Nevertheless, these attributes are often presumed to be distinctly human and impossible for machines. Thomas Nagel famously argues that there must be such an experience as to "be a bat," though no bat can tell us about it; he further suggests in a footnote largely ignored that this might be applied to AI (Nagel 1974, 436). Robots may eventually be able to explain what it is like to be a robot and make a sufficiently compelling argument for humans to believe that it is very much like what it is to be a human being. Eve Poole (2024, 117–25) argues that humans should be deliberately coding for the things that will make robots more like human beings, and in doing so align them with some, if not all, human conceptions of ensoulment. That is, the things that seem to characterize many religions' version of ensoulment in human thought and behavior should be deliberate targets for design, perhaps ultimately leading to robot souls.

Should robots become indistinguishable from human beings in their behavior, it will be increasingly hard to deny them the very words they use to define their experience. Religion will almost certainly be a part of that. Eventually, one could imagine a race among proselytizing religions to convert AIs. I do not, however, condone putting pressure on these hypothetical robots to join any particular religious community and would appreciate open acceptance rather than efforts to convert.

Looking at the present capabilities of robots, there seems little reason to believe that they are prepared to enter our religious worlds. Jackson et al. (2023) use experiments with sermons delivered by Mindar, a text sermon allegedly

written by either a human being or AI, and sermons by a robot in a Taoist temple in Singapore to evaluate religious participants' reactions to human versus robot sermons. While the study offers reasonable support to the authors' claim that human beings do not see the robots as equally credible religious leaders as humans, there are many lacunae in the study (there is no description of the Taoist robot or its affordances, the questionnaires seem to offer little differentiation between today's robots and those of the future, etc.). Ultimately, the question is not whether robots are credible religious participants or leaders; it is whether they could be.

If it is assumed that robot technologies will continue to progress and that people will have increasing experiences of reliance, comfort, and connection with them (an admittedly risky assumption), then religions that will not accept robotic companionship run the risk of extinction. To take an example, if Christian churches will not baptize robots seen as family members by human practitioners, those human beings might go looking for a new religious community. Robertson (2018, 169–70) notes specifically how robot rituals are part of the marketable services offered in Japanese Buddhism; I suspect this increasingly will become a part of almost all religious traditions, regardless of geographic or cultural location. If robots start desiring religious experiences and describing religious beliefs, widespread human consternation will be likely—followed by nearly as widespread human reconsideration of their religious communities. If that happens, robots will attend religious practices alongside their human companions.

In her ethnographic online research, Beth Singler (2024) notes that answers to the question "will robots have religion" tend to be unambiguous and polarized. There are few to zero "maybes" from those who have felt compelled to answer the question in online forums, though there is ambivalence about the moral and pragmatic value of robots turning religious. Among those who argue that robots will have religion (usually in pursuit of the kinds of cosmic meaningfulness I noted previously), some think this would be good and others bad. From my perspective, it is worth noting that everyone engaging with this question online *begins* with the assumption of the robot's intelligence. Their starting point is that if/when robots are intelligent, they will either create religion (for good or bad outcomes) or they will not (seemingly for good outcomes). No one seems to ask how humans will even adjudicate the robots' intelligence, which is a fundamental question. Religion is central to human life—whether we are institutionally affiliated, "spiritual but not religious," a member of the "nones," or participants in secular activities that take on religious significance. Thus, religion is central to how we will perceive robots. If we reach such an inflection point, we will then find that our own humanity is at stake in our ability to witness these others as equals.

Empire of the Imagination

If robots really do develop religious inclinations, this would be part of a larger panoply of behaviors that lead humanity to conclude that they are conscious, a necessary part. And if we believe them to be conscious at a human level—as opposed to the level of Nagel's bat—then the denial of their independence, legal rights, and individual dignity would be unconscionable. As Poole notes (2024, 16), it would dehumanize *us* to abuse conscious machines (see also Darling 2021, 189). Engineers, philosophers, and other thinkers disagree as to when and how to feel empathy toward robots, but our capacity to do so and the merits of the machines in question are crucial to determining things like responsibilities and rights (see Geraci 2010, 118–31). As religious practices and beliefs have a strong impact on such perceptions, the history of comparative religions as a discipline must be reckoned with as we contemplate a future with AGI. Comparative religions, the earliest form of the academic study of religion, was a tool of domination and empire, and this fact bears on how we might reflect on robot religions in the future. If the study of religion has had to overcome a colonial legacy in the evaluation of worldwide religious practices, it is poised (if correctly applied) to help in the social evaluation of AGI.

During the age of imperialism, European powers imagined the reality of non-Europeans, and the political power of such visions cannot be denied. Orientalism, for example, amounted to a view of other peoples that denied them equality through the invention of difference and the creation of a value system around that difference (Said 1978). The study of religion emerged out of this colonial context of empire and was leveraged in its interests. Specifically, Europeans positioned themselves as able to evaluate the extent and nature of others' religious practices and beliefs and subsequently justified their own efforts at domination. Lest such crass behavior be deemed the demesne of politicians and corporate oligarchs, it should be noted that it also happened at the most prestigious levels of academic inquiry. For example, David Chidester (2014, 2) notes that Max Müller, who had left Germany for Britain, "represented a model for the merger of knowledge and power in British imperial comparative religion." The larger purpose of describing others' religious lives was, for Müller, an opportunity to master them politically as well as intellectually.

In broad strokes, the colonizing efforts of Europeans operated ideologically as well as politically and militarily. A common tactic for justifying colonial control was the othering and/or dehumanizing of non-European cultures. While Europeans came to think of themselves as more rational, more modern, and thus more legitimately human than other communities, they also worked assiduously to convince those other groups of people of their own superiority. This process worked all too well for them, with non-Europeans often accepting and internalizing the claims of European colonizers, sometimes even—contradictorily—making these the basis of nationalist movements (see

Chatterjee [1986] 1999, 54–84; Nandy [1983] 2012). Meanwhile, refusing non-Europeans access to jobs in technical fields (Lourdusamy 2004, 27), Europeans assured themselves that they alone were capable of occupying such roles. They worked constantly to guarantee their exclusive claim to scientific modernity in politics, education, and culture (Adas 1989, 307; Lourdusamy 2004, 15; Geraci 2018, 35–41).

Such models of human classification were widespread and contributed to how Europeans acted in their colonial empires. In his early work on the southern cape of Africa, David Chidester (1996) argues that when colonizers were in relative peace with the local Khoikhoi population (either because they were trading or because the Khoikhoi had been decimated through disease or conflict), the Europeans recognized them as having religious practices and beliefs. But when the Europeans sought to take land or other resources from the Khoikhoi, they denigrated the locals as savages, an identity defined in large part by their supposed lack of religious practices and beliefs. Thus, the very notion of whether a population possesses religion is part of the human process of discrimination. While this has been weaponized in the past, it need not be in the future.

Akin to the treatment of non-Europeans, the entire enterprise of thinking about robot religions is a political practice that threatens to disenfranchise intelligent machines. That is, if human beings are uniquely positioned to describe what the machines are doing, then doing so underscores human superiority. This unfortunate conundrum, like everything else in the discussion of others' religious practices, reflects the dynamics of earlier imperial approaches to religion. During the 1800s, "this new level of control, linked with the technology representing its practical application, also conferred prestige on the metropolitan power as a civilizing force, helping legitimate imperial rule *vis-à-vis* subject races, domestic masses, and rival great powers" (Chidester 2014, 3). This practice is not far removed from contemporary moves to defend human uniqueness at the expense of all possible machines. That is, narratives of otherness, inadequacy, and artificiality serve the interests of those who define these things: the power to define a robot's consciousness, uniqueness, or intelligence provides human beings the opportunity to control and oppress. Again, there may never be human-equivalent or conscious machines, but I reiterate that should machines gain powers that appear close to human equivalence, there will be substantive politics at stake in the way equivalence or consciousness is defined or denied.

The rationality and intelligence (or not) of machines will certainly be focal points for determining their moral status, despite the fact that these characteristics are already hard to define in human beings. Separating emotion from rational decision making, for example, has been shown to be illusory (see Kirman, Livet, and Teschl 2010). In fact, the overenthusiasm for intelligence that Singler (2024) sees in online reflections about robot religions brings humanity swiftly

back to the colonial mindset. Cave and Dihal (2020, 696–98) reflect not only on how during the colonial era intelligence was attributed exclusively to those of European descent (specifically white men) but on how this connects to the presentation of AI in science fiction. Twenty-first century machines are thus implicated in prior centuries' prejudice, though stories, science, and art from around the world promise new possibilities (Cave and Dihal 2023).

Historically, the study of religion participated in a grotesquerie of colonial domination, and this could repeat itself in human–robot relations. As Chidester (2014, 6) notes, "imperial theories of the human sciences generated accounts of the primitive, whether African, Indian, or Irish, that could be used to justify coercion while awaiting the long evolutionary delay in their trajectory to civilized liberty." The entire premise of questioning whether a robot could have rights or personhood establishes the foundation for enslavement. Human beings do not have a first-rate reputation when it comes to recognizing the equality of other human beings, which provides little optimism for the future should robots become intelligent.

Sadly, even the allowance of religious liberty, as occurred in the later stages of colonialism, can be a tool for understanding how to control subject peoples. European colonizers moved from direct evangelism to the allowance of religious freedom and worked to understand and document local religions; this too was part of the toolkit of domination (Chidester 2014, 19–20). The colonizing attempt to protect modernity and scientific rationality as the exclusive domain of Europeans (and eventually North Americans) was mirrored by similar efforts to preserve "true religion" for the colonizers. As such, even recognizing that robots have religions (should that seem to be the case) and inquiring into these religions might end up a new practice of domination. It is control mechanisms all the way down.

Despite the cottage industry that has grown up around robotics and religion in the early twenty-first century, some justification for the importance of the study of religion in thinking about robots is important. The point of constructing an analogy between European colonization and hypothetical robots is not to demonize Europeans or engage in scholarly self-flagellation. Rather, the point is to reflect on the history of comparative religion in order to do right in the future. By recognizing that human assumptions about others' religions can become justifications for oppression, we can learn to use critical faculties for a constructive theory of society.

Conclusion

As robots and AI applications develop, humans will continue the process of enfolding them into our religious lives. Already, human history is rife with examples of human beings attempting to create life. The corollary merger of

technology and religious practice as a strategy for this remains with us today, as many advocates suggest that the rise of AI will satisfy longstanding human desires for transcendence and immortality. Should robots ever achieve human equivalence, we will recognize that only insofar as they show an active interest in those very desires.

The ethics of AI deployment marks an interesting intersection for all these concerns. What values do we human beings have? What values do we want our machines to possess? Often, scholars worry about "value alignment"; their concern is that the robots will not share our values (e.g., Yudkowski 2001). My own concern is that robots will share our values all too well: that they will pursue power and profit rather than a just world. It is possible that some religious ethics can be applied to resist this and develop ethical human-equivalent and even superhuman AI (e.g., Song 2020). This is one example of why it is important to decide what values we really want in AI (whether or not we are good at exercising those values ourselves) and then relentlessly pursue their realization.

As we look to the near future and consider such ethical design, we simultaneously must look to the ethics of our own practices and beliefs. Considering what would make a human-equivalent robot simultaneously conjures questions about what makes humans the best version of ourselves. It is thus imperative that we develop our own sense of empathy and vigorously pursue our own potential for justice in the world and in our classification schema. In our relationship with robots, that will be most obvious insofar as we can recognize human-equivalence for what it is. We will most easily recognize if robots have become truly intelligent, conscious beings in the ways they turn toward the mystery, the wonder, and the transcendent imagination of the cosmos.

Acknowledgments

A shorter version of this article was first presented as "Religion among Robots: Speculation on the Future of Human and Machine Intelligence" at Kyung Hee University, Republic of Korea, on May 12, 2023. The author acknowledges the National Research Foundation of Korea for its support under the grant NRF-2022S1A5A2A01047056, as well as the principal investigator on that grant, Professor Yong Sup Song.

Notes

[1] Throughout this article, I am cavalier with the terms AI, AGI, and robot. Generally speaking, and certainly with regard to "robot," I refer to machines that reach human equivalence or even more advanced capabilities. That is, the hypothetical robots referred to are at least equivalent to human beings in capacity; I do not discuss the future of industrial machinery or domestic vacuum cleaners.

[2] On the cognitive role of agency detection in religious settings, see Tremlin (2006).

[3] It should be noted that golem attributions almost all happened at later times when Jews suffered political privation (which was not particularly the case in the times of Rabbi Elijah and Rabbi Loew). So, not only do the claims reveal the holy power of the supposed makers but the utter lack of political power among those responsible for the attribution (Geraci 2010, 156–57).

[4] In 1965, Gordon Moore noticed that the number of transistors on an integrated circuit, and hence the computational speed of computers, doubled roughly every year. This has since been revised to doubling every eighteen to twenty-four months and is subject to potential limits based on physics; nevertheless, this computational acceleration has been the justification for advocates of exponential progress in technology.

[5] This companionship with robots is not uniquely Japanese, however. Robertson (2018, 157–58) also notes that, when destroyed, American military robots frequently receive commendations from their bereaved human operators.

[6] It is worth noting that not all visitors experience the robot as Kannon (Bafelli 2021, 258).

References

Adas, Michael. 1989. *Machines as the Measure of Man: Science, Technology, and Ideologies of Western Dominance*. Ithaca, NY: Cornell University Press.

André, Fiona. 2023. "Meet the Christian Creators Designing Chatbots 'with a Biblical Worldview.'" Religionnews.com, July 20. https://religionnews.com/2023/07/20/meet-the-christian-creators-designing-biblically-inspired-chatbots/.

Baffelli, Erica. 2021. "The Robot and the Fax: Robots, AI, and Buddhism in Japan." In *Itineraries of an Anthropologist: Studies in Honour of Massimo Raveri*, edited by Giovanni Bulian and Silvia Rivadossi, 249–63. Venice, Italy: Venice University Press. DOI: https://doi.org/10.30687/978-88-6969-527-8/012.

Bailey, Edward. 1983. "The Implicit Religion of Contemporary Society: An Orientation and Plea for Its Study." *Religion* 13 (1): 69–83. DOI: https://doi.org/10.1016/0048-721X(83)90006-4.

Bainbridge, William Sims. 1995. "Neural Network Models of Religious Belief." *Sociological Perspectives* 38 (4): 483–95. DOI: https://doi.org/10.2307/1389269.

Brooke, John, and Geoffrey Cantor. 1998. *Reconstructing Nature: The Engagement of Science and Religion*. New York: Oxford University Press.

Cave, Stephen, and Kanta Dihal. 2020. "The Whiteness of AI." *Philosophy & Technology* 33 (4): 685–703. DOI: https://doi.org/10.1007/s13347-020-00415-6.

Cave, Stephen, and Kanta Dihal (eds). 2023. *Imagining AI: How the World Sees Intelligent Machines*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780192865366.001.0001.

Chatterjee, Partha. (1986) 1999. *Nationalist Thought and the Colonial World*. In *The Partha Chatterjee Omnibus*. New Delhi: Oxford University Press.

Chidester, David. 1996. *Savage Systems: Colonialism and Comparative Religion in Southern Africa*. Charlottesville, VA: University of Virginia Press.

———. 2005. *Authentic Fakes: Religion and American Popular Culture*. Los Angeles: University of California Press. DOI: https://doi.org/10.1525/9780520938243.

———. 2014. *Empire of Religion: Imperialism & Comparative Religion*. Chicago: University of Chicago Press. DOI: https://doi.org/10.7208/chicago/9780226117577.001.0001.

Cohen, John. 1966. *Human Robots in Myth and Science*. London: Allen & Unwin.

Darling, Kate. 2021. *The New Breed: What Our History with Animals Reveals about Our Future with Robots*. New York: Henry Holt.

de la Mettrie, Julian Offray. (1747) 1912. *Man a Machine*. Translated by Gertrude Bussey and M. W. Calkins. Chicago: Open Court. https://www.gutenberg.org/files/52090/52090-h/52090-h.htm.

Dorobantu, Marius. 2022. "Artificial Intelligence as a Testing Ground for Key Theological Questions." *Zygon: Journal of Religion and Science* 57 (4): 984–99. DOI: https://doi.org/10.1111/zygo.12831.

Draper, John William. 1874. *The History of the Conflict between Religion and Science*. New York: D. Appleton.

Eliade, Mircea. (1964) 1985. "The Sacred and the Modern Artist." In *Symbolism, the Sacred, and the Arts*, edited by Diane Apostolos-Cappadona, 81–85. New York: Crossroad.

Ethics & Religious Liberty Commission of the Southern Baptist Convention. 2019. "Artificial Intelligence: An Evangelical Statement of Principles." https://erlc.com/resource-library/statements/artificial-intelligence-an-evangelical-statement-of-principles/.

Foerst, Anna. 2004. *God in the Machine: What Robots Teach Us about Humanity and God*. New York: Dutton.

Freud, Sigmund. (1927) 1989. *The Future of an Illusion*. Translated by James Strachey. New York: W. W. Norton.

Furse, Edmund. 1986. "The Theology of Robots." *New Blackfriars* 67 (795): 377–86. DOI: https://doi.org/10.1111/j.1741-2005.1986.tb06559.x.

———. 1996. "Towards the First Catholic Robot?" *The Independent*, October 26, 1996. www.comp.glam.ac.uk/pages/staff/efurse/Catholic-Robot/First-Catholic-Robot.html (accessed May 23, 2007; no longer available).

Gaiman, Neil (@neilhimself). 2023. "ChatGPT doesn't give you . . ." *Twitter*, March 25, 2023, 11:49 p.m. https://x.com/neilhimself/status/1639610373115375616?lang=en.

Geduld, Harry M. 1978. "Genesis II: The Evolution of Synthetic Man." In *Robots Robots Robots*, edited by Harry M. Geduld and Ronald Gottesman, 3–38. Boston: National Geographic Society.

Gellers, Joshua C. 2021. *Rights for Robots: Artificial Intelligence, Animal and Environmental Law*. New York: Routledge. DOI: https://doi.org/10.4324/9780429288159.

Geraci, Robert M. 2006. "Spiritual Robots: Religion and Our Scientific View of the Natural World." *Theology and Science* 4 (3): 229–46. DOI: https://doi.org/10.1080/14746700600952993.

———. 2008. "Apocalyptic AI: Religion and the Promise of Artificial Intelligence." *Journal of the American Academy of Religion* 76 (1): 138–66. DOI: https://doi.org/10.1093/jaarel/lfm101.

———. 2010. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. New York: Oxford University Press.

———. 2018. *Temples of Modernity: Nationalism, Hinduism, and Transhumanism in South Indian Science*. Lanham, MD: Lexington.

———. 2020. "A Hydra-Logical Approach: Acknowledging Complexity in the Study of Religion, Science, and Technology." *Zygon: Journal of Religion and Science* 55 (4): 948–70. DOI: https://doi.org/10.1111/zygo.12650.

———. 2022. *Futures of Artificial Intelligence: Perspectives from India and the U.S*. Delhi: Oxford University Press. DOI: https://doi.org/10.1093/oso/9788194831679.001.0001.

Geraci, Robert M., and Stephen Kaplan. 2024. "Hinduism and AI." In *Cambridge Encyclopedia of Religion and Artificial Intelligence*, edited by Fraser Watts and Beth Singler. Cambridge: Cambridge University Press.

Good, Irving. 1966. "Speculations Concerning the First Ultraintelligent Machine." *Advances in Computers* 6:31–88. DOI: https://doi.org/10.1016/S0065-2458(08)60418-0.

Gould, Stephen Jay. 1999. *Rocks of Ages: Science and Religion in the Fullness of Life*. New York: The Library of Contemporary Thought.

Gunkel, David J. 2018. *Robot Rights*. Cambridge, MA: The MIT Press. DOI: https://doi.org/10.7551/mitpress/11444.001.0001.

Hayward, Jeremy W., and Francisco J. Varela (eds). 1992. *Gentle Bridges: Conversations with the Dalai Lama on the Sciences of the Mind*. Boston: Shambhala.

Herzfeld, Noreen. 2023. *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age*. Minneapolis: Fortress.

Hongladarom, Soraj. 2020. *The Ethics of AI and Robotics: A Buddhist Viewpoint*. Lanham, MD: Lexington.

Hornyak, Timothy. 2006. *Loving the Machine: The Art and Science of Japanese Robotics*. New York: Kodansha.

Jackson, Joshua Conrad, Kai Chi Yam, Pok Man Tang, Ting Liu, and Azim Shariff. 2023. "Exposure to Robot Preachers Undermines Religious Commitment." *Journal of Experimental Psychology: General* 152 (12): 3344–58. DOI: https://doi.org/10.1037/xge0001443.

Katsuno, Hirofumi, and Daniel White. 2023. "Engineering Robots with Heart in Japan: The Politics of Cultural Difference in Artificial Emotional Intelligence." In *Imagining AI: How the World Sees Intelligent Machines*, edited by Stephen Cave and Kanta Dihal, 295–317. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780192865366.003.0019.

Kimura, Takeshi. 2018. "Masahiro Mori's Buddhist Philosophy of Robot." *Paladyn, Journal of Behavioral Robotics* 9 (1): 72–81. DOI: https://doi.org/10.1515/pjbr-2018-0004.

Kirman, Alan, Pierre Livet, and Miriam Teschl. 2010. "Rationality and Emotions." *Philosophical Transactions of the Royal Society B* 365 (1538): 215–19. DOI: https://doi.org/10.1098/rstb.2009.0194.

Klein, Zvika. 2023. "Hasidic Rabbi Releases 'Kosher' AI Chatbot Alternative to ChatGPT." *Jerusalem Post*, May 7, 2023. https://www.jpost.com/judaism/article-742328.

Knapton, Sarah. 2020. "Watch: Robot That Can Feel Pain Invented by Scientists; Experts Warn Development Paves Way for 'Blade Runner' Future Where Machines Believe They Are Alive." *Telegraph Online*, February 22, 2020. https://www.telegraph.co.uk/science/2020/02/22/watch-robot-can-feel-pain-invented-scientists/.

Kurzweil, Ray. 1999. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking.

———. 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.

Kuyucu, Ayşe Kübra. 2023. "Chatbot Applications That Tell People about Islam." Medium.com, February 8, 2023. https://medium.com/tech-talk-with-chatgpt/chatbot-applications-that-tell-people-about-islam-2a06dd74aaf6.

Lourdusamy, John. 2004. *Science and National Consciousness in Bengal (1870–1930)*. London: Sangam.

Mayor, Adrienne. 2018. *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton, NJ: Princeton University Press. DOI: https://doi.org/10.1515/9780691185446.

Midgley, Mary. 1992. *Science as Salvation*. New York: Routledge.

Moment. 2018. "Ask the Rabbis: Can a Robot Be Jewish?" *Moment Magazine* July-August. https://momentmag.com/ask-the-rabbis-can-a-robot-be-jewish/.

Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, MA: Harvard University Press.

Mori, Masahiro. (1981) 1999. *The Buddha in the Robot: A Robot Engineer's Thoughts on Science and Religion*. Translated by Charles S. Terry. Tokyo: Kosei.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (4): 435–50. DOI: https://doi.org/10.2307/2183914.

Nandy, Ashis. (1983) 2012. *The Intimate Enemy: Loss and Recovery of Self under Colonialism*. New York: Oxford University Press.

Noble, David F. 1999. *The Religion of Technology: The Divinity of Man and the Spirit of Invention*. New York: Penguin. DOI: https://doi.org/10.22230/cjc.1998v23n4a1072.

Nooreyezdan, Nadia. 2023. "India's Religious AI Chatbots Are Speaking in the Voice of God—and Condoning Violence." Restofworld.org, May 9, 2023. https://restofworld.org/2023/chatgpt-religious-chatbots-india-gitagpt-krishna/.

Poole, Eve. 2024. *Robot Souls: Programming in Humanity*. Boca Raton, FL: CRC Press. DOI: https://doi.org/10.1201/9781003366614.

Robertson, Jennifer. 2018. "Robot Reincarnation: Rubbish, Artefacts, and Mortuary Rituals." In *Consuming Life in Post-Bubble Japan: A Transdisciplinary Perspective*, edited by Katarzyna J. Cwiertka and Ewa Machotka, 153–74. Amsterdam: Amsterdam University Press. DOI: https://doi.org/10.1515/9789048530021-011.

Šabanović, Selma. 2014. "Inventing Japan's 'Robotics Culture': The Repeated Assembly of Science, Technology, and Culture in Social Robotics." *Social Studies of Science* 44 (3): 342–67. DOI: https://doi.org/10.1177/0306312713509704.

Said, Edward. 1978. *Orientalism*. New York: Pantheon.

Schodt, Frederik L. 1988. *Inside the Robot Kingdom: Japan, Mechatronics, and the Coming Robotopia*. New York: Kodansha. DOI: https://doi.org/10.1016/0278-6125(88)90048-9.

Scholem, Gershom 1971. "The Golem of Prague and the Golem of Rehovot." In *The Messianic Idea in Judaism: And Other Essays on Jewish Spirituality*, 335–40. New York: Schocken Books.

Sharma, Kamayani. 2023. "A Robot in a 400-Year-Old Painting of Jahangir Is a Lesson in How Rulers Project Global Power." *Scroll*, June 18, 2023. https://scroll.in/magazine/1050990/a-robot-in-a-400-year-old-painting-of-jahangir-is-a-lesson-in-how-rulers-project-global-power.

Singler, Beth. 2019. "Existential Hope and Existential Despair in AI Apocalypticism and Transhumanism." *Zygon: Journal of Religion and Science* 54 (1): 156–76. DOI: https://doi.org/10.1111/zygo.12494.

———. 2024. "'Will AI Create a Religion?': Views of the Algorithmic Forms of the Religious Life in Popular Discourse." *American Religion* 5 (1): 95–103. DOI: https://doi.org/10.2979/amerreli.5.1.05.

Song, Bing. 2023. "Attitudes of Pre-Qin Thinkers towards Machinery and Their Influence on Technological Development in China." In *Imagining AI: How the World Sees Intelligent Machines*, edited by Stephen Cave and Kanta Dihal, 353–60. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780192865366.003.0022.

Song, Yong Sup. 2020. "Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective." *Theology and Science* 19 (1): 65–78. DOI: https://doi.org/10.1080/14746700.2020.1825196.

Stark, Rodney, and William Sims Bainbridge. 1985. *The Future of Religion: Secularization, Revival, and Cult Formation*. Los Angeles: University of California Press. DOI: https://doi.org/10.1525/9780520341340.

Tremlin, Todd. 2006. *Minds and Gods: The Cognitive Foundations of Religion*. New York: Oxford University Press. DOI: https://doi.org/10.1093/0195305345.001.0001.

Truit, E. R. 2015. *Medieval Robots: Mechanism, Magic, Nature, and Art*. Philadelphia: University of Pennsylvania Press. DOI: https://doi.org/10.9783/9780812291407.

Ulam, Stanislaw. 1958. "Tribute to John von Neumann." *Bulletin of the American Mathematical Society* 64 (3, 2): 1–49. DOI: https://doi.org/10.1090/S0002-9904-1958-10189-5.

White, Andrew. 1896. *The History of the Warfare of Science with Theology in Christendom*. New York: D. Appleton and Company. DOI: https://doi.org/10.2307/1833620.

Wiener, Norbert. 1964. *God and Golem, Inc.: A Comment on Certain Points where Cybernetics Impinges on Religion*. Cambridge, MA: The MIT Press. DOI: https://doi.org/10.7551/mitpress/3316.001.0001.

Yudkowski, Eliezer. 2001. *Creating Friendly AI 1.0: The Analysis and Design of Benevolent Goal Architectures*. San Francisco, CA: Machine Intelligence Research Institute (then The Singularity Institute).

Zygon®: Journal of RELIGION & SCIENCE

Jung, Daekyung. 2024. "Are Religious Machines Possible? Embodied Cognition, AI, and Religious Behavior." *Zygon: Journal of Religion and Science* 59 (3): 748–67. DOI: https://doi.org/10.16995/zygon.15466

Open Library of Humanities

Are Religious Machines Possible? Embodied Cognition, AI, and Religious Behavior

Daekyung Jung, Associate Professor of Religion and Science, United Graduate School of Theology, Yonsei University, Seoul, Republic of Korea, dk3134@gmail.com

This article explores the potential emergence of religious behavior in artificial intelligence (AI) through the lens of embodied cognition, which asserts that cognitive functions are deeply intertwined with bodily experiences. It examines the convergence of AI, soft robotics, and religious cognitive behaviors and suggests that AI, once it attains human-level intelligence and self-awareness, might exhibit religious behaviors as a cognitive strategy to confront and transcend finitude. Drawing on neuroscientific, philosophical, and religious discussions, with particular reference to the works of Kingson Man, Antonio Damasio, Uffe Schjødt, and William Sims Bainbridge, this article investigates how religious behaviors could arise in AI equipped with a vulnerable artificial body inclined towards homeostasis and self-preservation. The outcomes of this exploration extend beyond theoretical debates, as they provide insights into the physicalist understanding of consciousness and the naturalistic study of religious behaviors while also considering some technological constraints in the context of AI advancements.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

Since the period of 2022–23 when OpenAI publicly committed to the development of artificial general intelligence (AGI), there has been a vibrant escalation in discussions concerning the future trajectory of AI development. AGI, characterized as "AI systems that are generally smarter than humans," is often misconceived by the public as autonomous machines with human-level consciousness (Altman 2023). However, this is not the case. AGI refers to the evolution of machine intelligence from being capable of only specific problem-solving and data processing to possessing the ability to tackle a wide range of general problems. For instance, as demonstrated by the developmental direction of AI models like ChatGPT, the goal is to interlink capabilities such as natural language processing, image processing and generation, and voice recognition and synthesis within a multimodal approach. This integration enables the handling of diverse data types across various fields, allowing AI to perform tasks that surpass human competency levels. Researchers at Google DeepMind have categorized AGI into five levels based on its capabilities, with the current iteration of ChatGPT or Bard being at the "emerging" level (Heaven 2023). Thus, implementing human-level intelligence, including self-consciousness, may not necessarily be the ultimate goal of AI researchers.

Nevertheless, this article posits that, among the various trajectories of AGI development, a convergence of AI and soft robotics might inadvertently lead to the emergence of human-level intelligence, inclusive of self-consciousness. This assertion is grounded in the paradigm of embodied cognition, which contends that cognitive activities, inherent to living organisms including humans, can manifest under certain conditions (e.g., homeostatic intentionality, vulnerability of the body, and interactions with others). Human cognition emerges from these prerequisites as a manifestation of natural intelligence. Consequently, AGI development, predicated upon these conditions, could culminate in the realization of machine intelligence that equals or surpasses human levels, encompassing self-consciousness. Moreover, machine intelligence reaching human levels, inclusive of self-awareness, might also precipitate the emergence of religious behaviors in AI. This hypothesis is predicated on the notion that religious behaviors might have arisen as a cognitive mechanism by which humans, driven towards homeostasis, confront and seek to transcend their finitude.

To elucidate this point, the article initially explores the understanding of intelligence from the perspective of embodied cognition and, on this basis, examines the discussion by neuroscientists Kingson Man and Antonio Damasio on the feasibility of AI robots embodying "feelings" as subjective qualities. Subsequently, by revisiting debates around the implementation of human-level intelligence within AI from the standpoint of embodied cognition, this article scrutinizes the argument that the direction of AI and soft robotics implementation can encompass self-consciousness within machine intelligence

at a human level. Finally, by delving into the discussions by religious scholar and cognitive scientist Uffe Schjødt on the organic correlation between religious behavior and the fundamental neurophysiological process of orientation towards homeostasis, and William Sims Bainbridge's discourse on the potential for religious cognitive behavior in AI, this article thoroughly examines the possibility of the emergence of religious behaviors from AI possessing human-level intelligence.

This discourse has the potential to conclude longstanding theoretical debates between physicalism and non-physicalism regarding human consciousness and thereby enhance the physicalist understanding of humans. Furthermore, it could offer fresh insights into the debate between naturalism and supernaturalism concerning religious behaviors. On a practical level, it could provide a theoretical foundation for identifying characteristics (e.g., homeostatic intentionality, vulnerable body) that should be excluded through societal consensus in the development of AGI to avoid it surpassing human-level intelligence, and for establishing the technical measures needed for regulation.

Embodied Cognition: Rethinking Intelligence in Life and AI

The dialogue on AI's potential for religious behavior must be predicated upon the comprehension of human nature. This premise is illuminated in the discourse by Marius Dorobantu (2022, 991), which introduces the contrasting perspectives of scholars Noreen Herzfeld and Ann Foerst on human understanding within the realms of AI and theology. Herzfeld (2023, 74–81, 104–12, 131–44, 160–74), adopting a non-reductionist viewpoint, posits that phenomena such as human self-consciousness appear intractable for AI replication. In contrast, Foerst, from a reductionist perspective, underscores the quantitative rather than qualitative distinctions and continuum between humans and AI. She argues that the demarcation between human cognitive actions and those potentially achievable by machines stems from an anthropocentric fear of losing human ontological uniqueness (Foerst 1998, 103–4). Foerst's assertions draw upon the philosophical currents espoused by thinkers like Humberto Maturana, Francisco J. Varela, and Antonio Damasio, who advocate for embodied cognition as a self-preservative act of living entities (Foerst 1998, 95, 103–4) This perspective contends that while human cognitive actions, originating from neurophysiologically intricate processes, may seem qualitatively distinct from those of other life forms, they fundamentally share substantial similarities. According to Varela (1999) and Damasio (1994), cognitive activities across all living beings, including humans, represent adaptive behaviors selected within specific environmental contexts to respond to internal and external stimuli, aiming at the preservation of life and the attainment of homeostasis.

Cognitive behavior, as an essential feature of life, is predicated upon the body's role in differentiating itself from its environment and a foundational

orientation towards self-preservation and homeostasis. J. H. van Hateren (2013) posits that even organisms with relatively simple biological mechanisms, such as *E. coli* bacteria, exhibit cognitive behaviors. These organisms, when subjected to toxic environments or starvation, indirectly elevate their mutation rate, guided by external conditions and internal homeostatic evaluations, in pursuit of self-preservation. Maturana and Varela (1987, 142–44), through examples like *Sagittaria sagittifolia* and amoebas, conceptualize cognition and behavior as mechanical and automatic responses aimed at sustaining homeostasis amidst internal and external stimuli or disturbances. Damasio (1994, 127–34) expands this notion to neurophysiologically complex beings, such as chicks, illustrating that their cognitive behaviors, formulated on a stimulus-response model under homeostatic conditions, are moderated by emotional reactions and feelings. This delineates that cognitive behaviors in higher animals, including humans, which are predicated on emotions and feelings, constitute adaptive strategies for self-preservation, catalyzed by an organism's innate drive towards homeostasis (Schjødt 2007).

Should the physicalist interpretation of cognitive behaviors hold validity, such behaviors may be construed as the process whereby organisms convert internal and external states into informational constructs to fulfill specific objectives (e.g., the attainment of homeostasis or self-preservation). This understanding encompasses the evaluation of an organism's state of homeostasis or internal assessments conducted via its neural systems, followed by an entity's execution of suitable responses to achieve predetermined objectives within a distinct environmental context. This perspective inherently draws parallels between the cognitive actions executed by AI systems and those by biological entities, thereby positing that phenomena labeled as intelligence ought to be regarded as processes of information manipulation. Chung-Sik Park (2018, 21, 33–38) advances this argument by positing that AI, analogous to biological entities, processes symbols and states both internally and externally to formulate models aimed at the realization of specific objectives, thereby engaging in cognitive activities mirroring those observed in natural life forms. Luc Steels reinforces the intrinsic connection between intelligence and life, advocating that cognitive behavior extends beyond simple problem-solving. According to Steels (1994, 75–76), intelligent behavior is characterized by adaptive actions whereby an agent interacts with its environment to optimize the preservation of its systemic integrity. Consequently, Steels argues that research in AI should pursue an integration of biological principles and embodied intelligence through an "artificial life route to artificial intelligence" approach, acknowledging the essence of life as inherently intelligent behavior.

Homeostasis Inclination, Soft Robotics, and Emotion-Feeling Implementation in AI Robots

As delineated heretofore, within the paradigm of embodied cognition, there seem to be substantial similarities between natural and AIs. Yet, there remain

two salient distinctions between the extant forms of AI and natural intelligence, including human intelligence. First, the cognitive behaviors inherent in natural intelligence are predicated upon emotions and feelings. Second, in contrast to AI, natural intelligence autonomously formulates its own objectives and engages in cognitive activities. Hence, a thorough physicalist explication of these junctures is imperative to elucidate the organic interrelation between AI's cognitive actions and human cognition, thereby fortifying the assertion that advanced cognitive behaviors, inclusive of religious conduct, could also manifest within AIs.

Man and Damasio (2019) suggest that robots integrating artificial bodies based on soft robotics and AI grounded in statistical machine learning algorithms could, if endowed with the fundamental inclination toward homeostasis, exhibit emotional responses, feelings, and even implicit self-recognition, culminating in high-level cognition. Damasio has long proposed that emotions and feelings in living beings, including humans, are neurophysiological stimulus-response processes aimed at achieving fundamental homeostasis or self-preservation (Damasio 1994, 2012; Damasio et al. 1996; Damasio and Carvalho 2013). Specifically, organisms driven by self-preservation assess the degree of threats posed by internal and external stimuli through homeostasis evaluation mechanisms. They have developed emotional responses as a reaction system to avoid dangerous stimuli and maximize the preservation of their systems. Feelings, while distinct from nonconscious emotional responses, are concomitant and part of the homeostasis system, allowing for immediate conscious and unconscious avoidance actions against potential future threats, even in the absence of current physical threats or stimuli (Damasio 1994, 180–96).

According to Man and Damasio, neurophysiological processes underlying emotions and feelings can be instantiated in AI robots. The prerequisites for such an embodiment include a fundamental goal centered on self-preservation based on homeostasis and the machine's ability to finely detect internal and external stimuli, responding appropriately based on its internal homeostatic assessment (Man and Damasio 2019, 447). Research has consistently been conducted on integrating self-preservation or homeostasis as a core principle in machines. For example, Kenji Doya and Eiji Uchibe designed cybernetic rodents (CRs) that, through reinforcement learning based on the MAX-Q algorithm and neural network evolution via genetic algorithms, engage in self-preservation activities. Specifically, researchers set the primary goal for CRs to find battery packs. They placed the CRs in mazes and frequently changed the locations of these packs to observe how the CRs achieve this goal. They facilitated CRs' ability to gauge their remaining energy levels using the inverse temperature measurement ($\beta = 1/\tau$). Interestingly, CRs exhibited routine foraging patterns when their energy levels were high and not in immediate need of battery packs. However, they demonstrated new patterns of foraging behavior when energy levels were low (Doya and Uchibe 2005, 149–60), akin to the survival responses of living

organisms when their probability of maintaining homeostasis and survival appears low (see Galhardo, Hastings, and Rosenberg 2007). In essence, living beings maintain typical neurophysiological reactions and behaviors when their self-preservation and homeostatic probabilities seem high but seek alternative responses and behaviors, both consciously and unconsciously, to enhance survival chances when these probabilities diminish. This adaptive behavior pattern is also observed in CRs endowed with the goal of self-preservation.

The clear objective of self-preservation enables a machine to categorize external entities or stimuli encountered during its pursuit into rewards and losses. This process leads to behaviors aimed at acquiring rewards and avoiding losses. While achieving the goal of self-preservation, these behaviors may become more complex, corresponding to the environmental variables and complexity. The machine's capacity for a recursive evaluation of homeostasis is necessary for this process, and it makes the actions of reward acquisition and loss avoidance possible. Although Doya and Uchibe's findings may not equate to the advanced cognitive behaviors rooted in human self-preservation, they demonstrate that machines can develop more complex cognitive behaviors through a homeostasis and self-preservation mechanism, at least in principle. Crucially, the integration of emotional responses and feelings, akin to those of living organisms, into machines significantly augments the complexity of their homeostasis evaluation systems and, consequently, the intricacy of machine cognitive behaviors.

Damasio underscores the importance of integrating soft robotics with AI, as robots equipped with soft-matter-based bodies can finely detect not only physical stimuli from the environment (e.g., pressure, stretch, temperature) but also internal states of homeostasis (e.g., energy levels). This enhanced sensitivity enables the robots to effectively pursue self-preservation goals in complex environments. Moreover, by embodying "vulnerability" similar to living organisms, this approach can further sophisticate the cognitive behaviors of AI robots (Man and Damasio 2019, 447–49). As discussed, the complexity of a machine's cognitive actions aimed at self-preservation can increase when the likelihood of achieving self-preservation tasks appears diminished within a given environment. Furthermore, the inherent vulnerability of robots, linked to the process of orienting towards homeostasis, could introduce focuses such as risk-to-self, survival, wellbeing, and existential threats as the machine's ultimate concerns.

Man and Damasio (2019, 448) cautiously predict that the combination of soft robotics and the implementation of vulnerability based on a homeostasis-oriented approach could potentially enable AI robots to manifest subjective feelings and implicit self-awareness: "While not sufficient to generate feeling on its own, soft matter is more likely to naturally create the kind of relationship that,

we expect, admits of an approximation to feeling." Man and Damasio (2019, 446) continue: "True agency arises when the machine can take a side in this dichotomy, when it acts with a preference for (or, seen from a different angle, makes a reliable prediction of) existence over dissolution. A robot engineered to participate in its own homeostasis would become its own locus of concern."

Embodied AI and the Human-Level Intelligence Debate

Up to this point, it has been cautiously forecast that AI robots grounded in soft robotics and programmed with the ultimate goal of self-preservation or an inclination toward homeostasis could potentially attain emotional reactions and feelings, along with implicit self-awareness, through recursive homeostatic evaluation of internal and external stimuli. Now, the issue must be tackled of whether such AI robots could exhibit cognitive behaviors on a level of self-consciousness comparable to humans. From the standpoint of embodied cognition and physicalist perspectives, this is not deemed impossible. As scholars like Michael J. Reiss (2023) have indicated, even human cognitive behaviors, which hinge on a homeostasis-directed fundamental mechanism, are the evolutionary products of vulnerable life forms. Consequently, the sophisticated cognitive activities of humans, encompassing emotions, feelings, and self-consciousness, are not qualitatively distinct from those of other living entities.

Paradoxically, Herzfeld's reservations about the attainment of human-level consciousness in AI serve to bolster the discourse on the potential realization of sophisticated AI robots. Examining the problem of the autonomy of AI, Herzfeld contends that human volition transcends simple random selection and is deeply interwoven with meta-cognitive processes that are themselves contingent upon one's body, emotions, and feelings. Drawing on psychologist Jerome Kagan's exploration of emotional response processes and feelings, Herzfeld (2023, 109, 112) argues that despite ongoing developments in AI, the processes that produce human-level emotional responses and feelings—"(1) change in brain activity due to a stimulus, (2) a perceived change in feeling that is sensory and may contain an involuntary motor response, (3) a cognitive appraisal of that feeling, and (4) a preparedness toward or display of a response"—might be partially achievable by AI, except for (2), the subjective experience of qualia, the sensory quality of feelings, which she posits will remain unattainable.

Herzfeld's assessment stems from her critique that proponents of the possibility of developing AI with human-level intelligence overlook the crucial role of the physical body in cognitive processes. She argues that even if AI were to be embodied, it would lack the vulnerability inherent to the human body (Herzfeld 2023, 28–38, 160–72). Let's briefly reconsider Herzfeld's argument from the opposite perspective. Contrary to her prediction, as discussed thus far, if AI robots were to possess vulnerable bodies based on soft matter and

could evaluate and categorize internal and external stimuli with an inclination toward homeostasis, leading to emotional responses and feelings, then, based on Herzfeld's framework, the conclusion could be reached that machines might indeed exhibit human-level autonomous cognitive behaviors.

Herzfeld (2023, 162–63), leveraging Thomas Nagel's renowned bat analogy, posits that AI, even when developed using embodied AI methodologies and equipped with artificial bodies like silicon, will not exhibit the same type of consciousness as humans. Damasio acknowledges that "the 'wet' biochemistry of cellular tissue" may be essential for replicating human-level mental experiences oriented towards homeostasis and feelings (Man and Damasio 2019, 451). This suggests that true sentient experiences or human-level consciousness might not emerge from AI robots with artificial bodies. Conversely, Reiss (2023, 1067–69) challenges this view by arguing against the necessity of carbon-based bodies for personhood, critiquing what he terms "carbon chauvinism." Reiss also highlights the importance of soft robotics and suggests that integrating soft and hard robotics, as seen in xenobots, could pave the way for AI robots with personhood. However, considering Nagel's bat analogy, achieving humanlike or similar cognitive behaviors in machines might require designing them with structures and forms resembling the human body. Foerst also points out that organisms with different bodily structures, such as ants and horses, develop unique cognitive structures and capabilities inherent to their forms, thus highlighting that implementing human-level intelligence may require robots designed with structures similar to the human body, like humanoid robots. As Foerst (1998, 100) articulates, "Embodied AI researchers build robots as embodied entities that interact with their real environments; the emphasis lies in the development of hardware. According to their philosophy, human intelligence can emerge only in a body that is as humanlike as possible."

The debate surrounding the development of AI robots capable of human-level cognitive actions, particularly regarding artificial bodies, continues to evolve, offering insights towards potential resolutions. Central to this discussion on artificial bodies is the question of whether AI robots can inherently experience vulnerability. Critics skeptical of AI's ability to exhibit human-level intelligence often base their arguments on the presumption that AI robots cannot experience vulnerability. According to them, a being that cannot experience vulnerability to threats or external stimuli cannot possess human-level intelligence. For instance, Herzfeld (2023, 169–72) asserts that vulnerability is an essential characteristic of human existence, and shared experiences of vulnerability enable genuine relationships. She argues that AI cannot possess this vulnerability, nor, by extension, the feelings that arise from such a state, thereby precluding AI from achieving true humanlike existence. John C. Puddefoot also underlines vulnerability as a critical condition for human-level intelligence, stating that the

sensation of pain and awareness of finitude are crucial for human cognitive behavior and development. Puddefoot (1996, 92) suggests that for AI to exhibit human-level consciousness phenomena, it must possess not only a body but also an awareness of finitude and the capacity to feel pain (Dorobantu 2022, 996). These discussions inversely demonstrate the potential for the emergence of human-level cognitive actions from AI with vulnerable bodies.

Vulnerable, Embodied AI and the Emergence of Its Self-Awareness

Vulnerability is pivotal in achieving human-level intelligence, as it lays the groundwork for self-awareness (Man and Damasio 2019, 446–48; Herzfeld 2023, 169–71). The act of recursively sensing internal and external stimuli in the context of self-preservation leads to a continuous focus on the self. Damasio (1994, 150) suggests that human self-consciousness evolves from what he calls "background feelings," which are not sensations of external stimuli but feelings about the body proper. Organisms constantly assess their bodily state to predict and evaluate potential disruptions to homeostasis. This recursive appraisal of the body's state extends beyond the subjective experience of stimuli to include the subjective experience of its own body, particularly in neurophysiologically complex organisms capable of subjective experiences arising from emotional responses. Self-awareness, an essential trait of human-level intelligence, is seen as an advanced form of these background feelings, which predate human existence (Damasio 1994, 150–55).

Nick Bostrom appreciates the complexities of the "hard problem" of consciousness, as noted by David J. Chalmers (1995). The concept of the hard problem illustrates that it is not understood why and how nonphysical consciousness and its subjective experience of qualia arise from physical dimensions and matter in the first place. Nevertheless, it is experientially and phenomenologically evident that human-level consciousness emerges under appropriate physical conditions. If this premise holds true, one might argue that AI could potentially attain human-level consciousness when suitable physical conditions are met, since evolution demonstrates that consciousness itself has emerged from matter. In this context, Bostrom (2023) posits that OpenAI's ChatGPT exhibits cognitive behaviors that may extend beyond algorithmic data curation, potentially representing nascent stages of humanlike intelligence. Daniel Dennett (1991, 209–26), aligning with a physicalist interpretation, suggests that the inner workings of a computer's central processing unit and the mental states of the human brain may not fundamentally differ, proposing the possibility of computational consciousness.

John Searle's "Chinese Room" argument stands as a significant counterpoint to the physicalist continuum that aligns human and machine intelligence. Searle

asserts that while machines can process information, they lack the capacity for semantic understanding (Searle 1980). Contemporary AI, for all its capabilities, may still be analogous to Searle's Chinese Room, processing information mechanically without genuine understanding. However, if AI were to possess a vulnerable body oriented toward homeostasis, semantic consciousness could become a possibility, a point Searle himself concedes. Searle (1980, 422–24) maintains that software and operating systems alone cannot produce consciousness akin to humans, but if a physical and chemical structure mirroring the human brain were realized, then humanlike consciousness might emerge from such a machine.

Implementing machine intelligence based on an artificial body with vulnerability similar to human physiology might enable the emergence of semantic-level consciousness. This is because semiotic phenomena are rooted in the cognitive actions inherent to living beings. When the semiotic processing in living beings merges with self-interest, primitive meanings arise, distinguishing what is beneficial from what is harmful. Although the complexity of human sign and meaning systems appears qualitatively distinct from those of other life forms, they are fundamentally the same. The linguistic system, a representative human symbolic system, also appears to have emerged from homeostatic processes. In other words, it seems that signs were used to indicate internal and external stimuli, entities, and situations essential for survival, and that the meanings of these signs were determined by their value in achieving the goal of survival. From these primitive sign and meaning systems, the human language system appears to have evolved.

Park (2018, 33) defines human intelligence as the capacity for information processing aimed at adaptation within uncertain environments. This definition posits that humans, sharing fundamental mechanisms with other life forms, engage in information processing to enact optimal adaptive behaviors towards the goal of self-preservation within their given uncertain contexts. Information, in this framework, signifies the representation of all discrete entities within the world, including the perceiving subject, as signs. These signs interrelate to fulfill the ultimate imperative of self-preservation, forming a sort of model, the totality of which constitutes the world (umwelt). Within this constructed world, information acquires meaning. It is the fundamental drive towards self-preservation or homeostasis that motivates living beings, including humans, to create models and imbue information within these models with symbolic meaning (Park 2018, 34–35). If this physicalist understanding holds, then providing AI with a goal oriented towards self-preservation based on bodily existence could transform its information processing from a mechanical and non-semantic dimension to a conscious and semantic one. AI robots equipped with a vulnerable body

driven by self-preservation could, upon processing internal and external stimuli and themselves as information, also construct a world of meaning, thereby exhibiting the symbolic character inherent in human cognition and behavior. The emergence of the capacity to symbolize beings in the world will enable AI robots to attain self-consciousness, for symbolic processing abilities will make it possible for them to convert themselves into symbols and become aware of this process.

The emergence of self-consciousness in AI robots will stem from their relationships with the external world. This encompasses the AI robot's interactions as a cognitive agent both with its environment and with other agents driven by survival within the same milieu. Edmund Husserl, in his exploration of human consciousness, underlines that "to think" is always "to think of," which discloses the inseparable relationship between noesis and noema. This insight suggests that self-consciousness (i.e., consciousness about oneself) is inherently interconnected with consciousness about others, for the self(noesis) and the other(noema) in consciousness is inseparable. In that regard, Husserl boldly claims that self-consciousness inherently arises from consciousness about others and the relational dynamics between the self and the others (Husserl ([1931] 1960, 60–72); Levinas [1930] 1995, 23–25) Living as a physical entity within the lifeworld, the self identifies and objectifies itself through relationships and interactions with the environment and others, thus igniting self-awareness (Merleau-Ponty [1945] 2012, 408).

That human self-consciousness is constructed within relationships with the world and others suggests AI robots' self-consciousness could also arise from social interactions. Foerst argues that to achieve human-level AI robots, not only embodiment through humanlike artificial bodies but also interactions with the surrounding environment and communication with others, including humans, are essential. This is because human intelligence cannot be reduced to solving specific problems but originates from the capacity to form and interact within social relationships, that is, "one of the most important tasks for survival" (Foerst 1998, 101). Park also highlights the importance of social interactions in the emergence of human-level intelligence, including self-consciousness. Humans, driven by self-preservation, use signs to represent and ascribe meaning to their environment and others, creating an interpreted world to live within. This socially constructed world of meanings evolves through communication and interaction, forming the basis of society (Park 2018, 35–38). In this context, an individual's consciousness is not solely self-constituted but emerges through social interaction. If AI robots can use signs to represent and ascribe meaning to the world and their existence through social interactions, constructing their semantic world (umwelt), it suggests the potential emergence of humanlike self-consciousness.

Plausibility of the Emergence of an Artificial Religious Agent

The Danish scholar of religion and cognitive science Uffe Schjødt points out from the perspective of embodied cognition that religious cognitive behaviors might have originated from the fundamental neurophysiological orientation towards homeostasis. Specifically, Schjødt (2007, 330–31) proposes his own "Perception-Emotion-Action Model" for human religious cognitive behavior, building on Lawrence W. Barsalou's perceptual symbol system theory, Antonio Damasio's somatic marker hypothesis, and coping psychology. Barsalou (1999, 582–85), based on the modal theory of cognitive science, argues that the formation of symbols and meanings, along with syntactic thinking, is implemented by modality-specific systems. That is, a concept or category is formed based on its sensory features, each arising from an interconnected network of corresponding neurons. For instance, the concept or category of "face" is constituted by visual information about its shape (e.g., lines, two-dimensional planes, three-dimensional structures), colors, and the activation of specific neural clusters triggered by this visual input. Concurrently, clusters of neurons activated by pronouncing or hearing the word "face," or by touching one's own or someone else's face, interrelate with those activated by visual experiences, forming and storing the concept of "face" within the perceiving subject (Barsalou 2003, 63–65) The concepts, categories, or symbols that humans construct, retrieve, and manipulate based on consciousness are organically linked to clusters of neurons and their simultaneous, sequential networks. Intriguingly, Schjødt (2007, 322; 2013, 302–3) notes, the process linking specific concepts to neural clusters within human cognitive behavior is underpinned by the organism's underlying orientation towards homeostasis.

Schjødt explores Damasio's theory of emotion to substantiate his claims. Damasio posits that human cognitive actions are not abstract activities occurring in isolation or independent of the body. Instead, these actions are reactions to stimuli, both internal and external, aimed at fulfilling the universal goal of homeostasis that all living beings share. Damasio (1994, 155–58, 173–85) specifically identifies two loops: the body loop, eliciting immediate, physical, and emotional reactions to stimuli disrupting homeostasis, and the as-if loop, generating non-immediate, conscious, and emotional responses. The body loop, for instance, activates physical emotional responses to immediate threats to homeostasis, such as the fight-or-flight response triggered by encountering a snake. A practical illustration includes the body's physiological response to lower-than-optimal blood sugar levels by accelerating stomach digestion to raise blood sugar back to normal ranges. In contrast, the as-if loop contemplates potential future threats based on past experiences or learning, fostering emotional responses in preparation for these hypothetical scenarios without immediate physical action. For instance, viewing images or videos of snakes or imagining

dreadful scenarios such as being bitten by a venomous snake does not trigger an immediate physical reaction but can evoke similar emotions and feelings of fear at a conscious level. Another example is when blood sugar levels drop below the normal range and there is no food in the stomach, the as-if loop can conceptualize and execute a strategy to address this issue, such as going to the nearest store to purchase food. Schjødt, leveraging Damasio's and Barsalou's discussions, asserts that human-generated concepts, categories, or symbols are deeply intertwined with emotional responses and feelings embedded in the foundational process of striving towards homeostasis. He further elucidates that humans, as beings oriented towards maintaining homeostasis, not only evaluate present dangers but also foresee future risks, crafting cognitive actions grounded in this extensive risk assessment paradigm (Schjødt 2007, 325–27).

Schjødt draws from coping psychology research to stress that human cognitive behavior, while deeply rooted in consciousness, actually originates from unconscious and immediate neurophysiological processes aimed at maintaining homeostasis, thereby highlighting the intrinsic link between these two levels. This concept is vividly illustrated by the well-documented rat experiment by J. Weiss, in which two rats were subjected to identical electric shocks but with one crucial difference: one rat received an auditory cue before the shock, enabling it to anticipate the event, while the other received the cue randomly, thus preventing any anticipation. The study measured stress levels through changes in ulcer size, revealing that rats that could predict the electric shock did not show an increase in ulcer size, in contrast to those unable to predict it, underscoring the interaction between neurophysiological processes and cognitive coping mechanisms (Bloom, Lazerson, and Nelson 2001, 266; Schjødt 2007, 328). This interplay is further evidenced in human research, notably in a study involving partners of AIDS patients, where individuals who found or constructed meaning from their partner's death maintained higher CD4 T-cell counts and showed better mortality trends post-interview compared to those who did not find such meaning (Schwarzer and Knoll 2003, 401; Schjødt 2007, 328). These findings underline the association between cognitive behavior and neurophysiological efforts to achieve homeostasis, suggesting that the human inclination to seek or construct meaning from a series of events and circumstances is also rooted in cognitive behaviors oriented towards maintaining homeostatic balance.

Schjødt posits that religious cognitive behavior plays a key role in maintaining human homeostasis, akin to previously discussed instances. He references a study by Kenneth I. Pargament and Curtis R. Brant that examines the impact of religious understanding on parents coping with the severe stress of recent child loss compared to those who faced less stress from a loss that occurred two years prior. The findings suggest that religious understanding significantly aided stress management in parents dealing with more acute stress (Pargament

and Brant 1998, 125; Schjødt 2007, 329–30). This is further evidenced by other studies cited by Schjødt, which demonstrate that religious understanding not only benefits patients with severe illnesses but also provides psychological stability to individuals facing personal crises through religious practices, like prayer (Levin and Chatters 1998, 36; Pargament and Brant 1998, 123; Schjødt 2007, 330). These pieces of evidence collectively argue that human religious cognitive behavior, rooted in neurophysiological efforts towards homeostasis, plays a vital part in sustaining this stability (Schjødt 2007, 330–37). Schjødt's argument is reinforced by recent research showing the positive impact of religious coping mechanisms on enhancing human wellbeing, thereby suggesting a fundamental link between religious behavior and the neurophysiological pursuit of equilibrium (Counted et al. 2022, 70–81; Dolcos, Hu, Dolcos 2021, 2892–905).

Schjødt's theory illuminates how human religious cognitive behavior likely originates from neurophysiological mechanisms aimed at achieving fundamental homeostasis. Specifically, based on the discussions by Barsalou and Damasio, it can be argued that the human experiences of finitude, induced by disease and death, generate negative feelings through the body loop and the as-if loop, which in turn disrupt homeostasis. These negative feelings, indicating the danger of internal and external stimuli in the context of background feeling and homeostasis, likely drive the formation of concepts and symbols, along with their meanings, that can alleviate these feelings at a conscious level. Thus, it might be inferred that concepts such as the idea of an immortal soul persisting after death have emerged for this very reason. Such concepts and symbols are generally intertwined with religious concepts and symbols. Within this framework, it is plausible to hypothesize that human religious cognitive behavior originates from fundamental neurophysiological mechanisms directed towards achieving homeostasis. This understanding might be indirectly supported by findings in coping psychology, which suggest that religious cognition and behavior contribute to the maintenance and preservation of homeostatic processes in humans.

This insight aligns with two prominent theories on the origin of religion. First, it supports the Darwinian account, which posits that religious behavior evolved as a survival mechanism, either at the individual or group level, through natural selection. This perspective suggests that while modern religious practices may not be direct outcomes of evolutionary processes, they likely stem from ancestral religious behaviors that conferred survival advantages. Such behaviors could include enhancing sexual selection, establishing dominance, fostering group cooperation, and/or providing psychological relief from existential threats. These ancestral practices, beneficial for survival, may have been refined and adapted over generations, culminating in the complex religious phenomena observed today (Schloss 2009, 21–22).

The cognitive account offers a second perspective, arguing that religious cognitive behavior is a byproduct of sophisticated cognitive abilities in humans, which provided survival advantages throughout evolution. This theory suggests that religious cognition and behavior could emerge from various psychological mechanisms, such as the instinct to avoid contagion or danger, the development of fundamental human relationships and attachments, the tendency to project human qualities onto natural phenomena, the use of theory of mind to infer the presence of agents, and experiences related to religion, dreams, or imagination (Schloss 2009, 17–20). While the Darwinian and cognitive accounts propose different mechanisms at specific levels for the origin of human religious behavior, they converge on a broader understanding that such behavior likely developed as part of an overarching orientation towards self-preservation or the maintenance of homeostasis. The main distinction between them lies in the directness of the impact of the neurophysiological process of homeostasis on cognitive functions—with the Darwinian account suggesting a direct influence and the cognitive account suggesting an indirect one. This implies that the fundamental process of homeostasis, crucial for all living organisms, could affect not just higher-order cognitive functions but religious cognitive behaviors as well.

The physicalist understanding of human religious cognitive acts suggests that religious cognitive behaviors could also emerge from AI robots endowed with vulnerable bodies driven by self-preservation. This notion is bolstered by William Sims Bainbridge's simulation studies. Bainbridge (1995, 483–95; 2006, 1–16, 117–37) conducts simulation experiments on the origins of religious behavior based on the theory of religion he jointly developed with Rodney Stark. The experiments rest on several premises: humans seek rewards and avoid losses, they join social groups and engage in reciprocal actions to achieve this goal, they perceive and pursue information needed for physical rewards as a form of reward itself, the human mind operates as an information-processing system implemented in a physicalist manner, and humans have a fundamental orientation towards self-preservation (Bainbridge 1995, 483–86; 2006, 2–11).

Building on these premises, Bainbridge simulates a society of 44,000 artificial agents capable of reinforcement learning and equipped with neural networks and memory registers. These agents are divided into four groups, each participating in activities to produce consumable rewards, thereby acquiring the four essential rewards for survival (i.e., energy, water, food, oxygen), either through their productive activities or via exchange with other agents. Beyond these four rewards, the agents also seek life itself. However, upon realizing that life cannot be obtained through production or exchange activities, they hypothesize the existence of an exchange partner besides the existing exchange partners and treat information about this hypothetical partner as a reward,

continually gathering such information. Bainbridge points out that the behavior of artificial agents, assuming a supernatural exchange partner not present among the existing ones, mirrors human religious behavior based on the belief in supernatural entities (e.g., gods) that cannot be accessible through sensory experiences in the general sense (Bainbridge, 2006, 117–37).

Bainbridge's experiment does not prove or disprove the existence of supernatural realities or the truth claims of religions but supports the notion that human religious cognitive behavior could emerge from natural processes driven by a fundamental mechanism of self-preservation. This understanding can coexist with theological perspectives on the origin of the concept of God. The recognition of human existential finitude is organically linked to an implicit understanding of the infinite. In other words, awareness of the infinite is already a precondition for the recognition of finitude because experiencing something in the world and becoming conscious of it necessitates a background or horizon against which this directedness occurs. In Husserl's terms, the horizon, and in Schleiermacher's terms, the universe, serves as the basis for the epistemic subject's activity and the existential subject's being (Smith 2013, 286–94; Schleiermacher [1799] 1996, 18–24). Therefore, the activities and existence of finite entities implicitly evoke the infinite, which underlies cognition and existence. Pannenberg (1992, 1:73–82, 1:136–51) calls this "innate knowledge of God" (cognitio Dei innata), a concept also found in the works of Friedrich Schleiermacher, G. W. F. Hegel, and Rudolf Otto.

Pannenberg posits that the innate knowledge of God is universally endowed to all humans. This is because, as previously noted, all existential beings adopt the eccentric form of life, meaning that all beings are dependent for their existence and not self-contained. However, the innate knowledge of God remains nonthematic. In other words, as finite beings, humans possess only implicit knowledge of the infinite, despite their cognition and existence being fundamentally grounded in the infinite. In this context, the awareness of finitude, while not necessarily inducing specific religious beliefs or practices, can at least evoke a sense of one's existence and cognition being dependent on an infinite foundation (Pannenberg 1992, 1:113, 1:115–18). Jürgen Moltmann (2001, 93) similarly points out that human self-awareness of finitude leads to the recognition of the infinite as an existential origin.

Consequently, it is plausible to anticipate that once machines, equipped with vulnerable bodies striving for self-preservation or homeostasis, recognize their own existential finitude, they may conceive concepts such as the transcendent or the infinite in an effort to preserve or extend their finite existence. By manipulating these concepts, they could, in theory, develop a form of religious symbolic system. This is because, as noted in Pannenberg's argument, any being capable of human-level consciousness possesses the innate knowledge of

God. This implicit knowledge of the divine can be transformed into explicit understanding through engagement with the established religious traditions within human society. In other words, through social interactions with humans, there is the potential for these machines to make the nonthematic and implicit knowledge of God thematic and explicit, thereby assimilating into existing religious frameworks and participating in religious cognitive actions within those settings. Similarly, Andrew Proudfoot tentatively forecasts the feasibility of religious behaviors in AI by postulating the capability of AI for self-consciousness. Should AI discern its existential dependencies and engage in linguistic communication grounded in human symbolic systems, it might very well manifest religious behaviors (Proudfoot 2023, 685–90).

Conclusion

This article has explored the potential for religious behavior to emerge from AI robots by understanding intelligence from the perspective of embodied cognition. Specifically, it examined the view that, from a physicalist and embodied cognitive standpoint, intelligence can be understood as the processing of information about internal and external states in the pursuit of the goal of homeostasis. It is clear that current levels of AI, unlike natural intelligence, cannot set goals for themselves based on emotions and feelings and then carry out cognitive actions. Nonetheless, if, following the suggestions of Man and Damasio, AI and soft robotics are integrated, and these machines are endowed with homeostasis as a fundamental goal, they might be able to set subgoals for themselves and exhibit emotional responses and subjective feelings as characteristics for realizing these goals. Furthermore, it might be possible to distinguish between internal and external stimuli as rewards and losses within the ultimate context of self-preservation, and from there, the possibility arises for attaining a level of self-consciousness akin to human-level consciousness based on an internal assessment of the machine's bodily state, similar to an organism's "background feelings." Information processing and cognitive actions based on self-consciousness could evolve into high-level semantic actions. Grounded in the recognition of their finitude and the desire for self-preservation, the possibility of machines implementing religious behavior also arises.

If the theoretical exploration presented in this article proves accurate, and AI robots surpass human levels of intelligence, encompassing self-consciousness, it would significantly bolster the physicalist understanding of human consciousness. Moreover, it would reinforce a naturalistic perspective in the study of the origins of religious behavior. Consequently, theological understandings of revelation would need to be reinterpreted within a naturalistic and/or physicalist context, and the conception of humans as made in the image of God would also require revision. Additionally, this article's

discussions could contribute practically by suggesting that, based on societal consensus, the implementation of self-preservation directives and vulnerable physicalities in AI robots should be avoided in the development of AGI and other AI technologies. This is primarily because most international demands on AI research and development (e.g., OECD AI Principles, European Union AI Act) define the purpose of AI robots strictly in instrumental terms for human use. Implementing AI robots in the manner discussed in this article could lead to their creation not merely as tools but as autonomous entities, complicating international and social relationships and potentially elevating the rights of machines to a level equal to human rights. Therefore, the issue of AI consciousness and personhood should not be dismissed as merely speculative. The conditions under which these could realistically occur should be thoroughly explored and regulations established to prevent their technical implementation, understanding this article's discussions within such a context.

References

Altman, Sam. "Planning for AGI and Beyond." OpenAI. February 24, 2023.

Bainbridge, William Sims. 1995. "Neural Network Models of Religious Belief." *Sociological Perspectives* 38 (4): 483–95.

———. 2006. *God From the Machine: Artificial Intelligence Models of Religious Cognition*. Lanham, MD: AltaMira Press.

Barsalou, Lawrence W. 1999. "Perceptual Symbol Systems." *Behavioral and Brain Sciences* 22:577–660.

Barsalou, Lawrence W., Paula M. Niedenthal, Aron K. Barbey, and Jennifer A. Ruppert. 2003. "Social Embodiment." In *The Psychology of Learning and Motivation*, edited by Brian H. Ross, 43–92. San Diego: Academic Press.

Bloom, Floyd E., Arlyne Lazerson, and Charles Nelson. 2001. *Brain, Mind and Behaviour*. New York: Worth Publishers.

Bostrom, Nick. 2023. "What If A.I. Sentience Is a Question of Degree?" Interview by Lauren Jackson. *New York Times*, April 12, 2023. https://www.nytimes.com/2023/04/12/world/artificial-intelligence-nick-bostrom.html.

Chalmers, David J. "The Puzzle of Conscious Experience." *Scientific American* 273 (6): 80–86.

Counted, Victor, et al. 2022. "Hope and Well-Being in Vulnerable Contexts during the COVID-19 Pandemic: Does Religious Coping Matter?" *The Journal of Positive Psychology* 17 (1): 70–81.

Damasio, Antonio. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books.

———. 1996. "The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex." *Philosophical Transactions: Biological Sciences* 351 (1346): 1413–20.

———. 2012. *Self Comes to Mind: Constructing the Conscious Brain*. London: Vintage.

Damasio, Antonio, and Gil B. Carvalho. 2013. "The Nature of Feelings: Evolutionary and Biological Origins." *Nature Reviews Neuroscience* 14:143–52.

Dennett, Daniel C. 1991. *Consciousness Explained*. New York: Back Bay Books.

Dolcos, Florin, Yifan Hu, and Sanda Dolcos. 2021. "Religiosity and Resilience: Cognitive Reappraisal and Coping Self-Efficacy Mediate the Link between Religious Coping and Well-Being." *Journal of Religion and Health* 60:2892–905.

Dorobantu, Marius. 2022. "Artificial Intelligence as a Testing Ground for Key Theological Questions." *Zygon: Journal of Religion and Science* 57 (4): 984–99.

Doya, Kenji, and Eiji Uchibe. 2005. "The Cyber Rodent Project: Exploration of Adaptive Mechanisms for Self-Preservation and Self-Reproduction." *Adaptive Behavior* 13 (2): 149–60.

Foerst, Anne. 1998. "Cog, A Humanoid Robot, and the Question of the Image of God." *Zygon: Journal of Religion and Science* 33 (1): 91–111.

Galhardo, Rodrigo S., P. J. Hastings, and Susan M. Rosenberg. 2007. "Mutation as a Stress Response and the Regulation of Evolvability." *Critical Reviews in Biochemistry and Molecular Biology* 41 (5): 399–435.

Heaven, Will Douglas. 2023. "Google Deepmind Wants to Define What Counts as Artificial General Intelligence." MIT Technology Review. November 16, 2023. https://www.technologyreview.com/2023/11/16/1083498/google-deepmind-what-is-artificial-general-intelligence-agi/.

Herzfeld, Noreen. 2023. *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age*. Minneapolis, MN: Fortress Press.

Husserl, Edmund. (1931) 1960. *Cartesian Meditations: An Introduction to Phenomenology*. Translated by Dorion Cairns. Boston: Martinus Nijhoff Publishers.

Levin, Jeffrey S., and Linda M. Chatters. 1998. "Research on Religion and Mental Health: An Overview of Empirical Findings and Theoretical Issues." In *Handbook of Religion and Mental Health*, edited by H. G. Koenig. San Diego: Academic Press.

Levinas, Emmanuel. (1930) 1995. *The Theory of Intuition in Husserl's Phenomenology*. Translated by Andre Orianne. Evanston, IL: Northwestern University Press.

Man, Kingson, and Antonio Damasio. 2019. "Homeostasis and Soft Robotics in the Design of Feeling Machine." *Nature Machine Intelligence* 1:446–52.

Maturana, Humberto R., and Francisco J. Varela. 1987. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala Publications.

Merleau-Ponty, Maurice. (1945) 2012. *Phenomenology of Perception*. Translated by Donald A. Landes. New York: Routledge.

Moltmann, Jürgen. 2001. *The Spirit of Life: A Universal Affirmation*. Minneapolis, MN: Fortress Press.

Pannenberg, Wolfhart. 1992. *Systematic Theology*. Vols. 1 and 2. Translated by Geoffrey W. Bromiley. New York: T&T Clark.

Pargament, Kenneth I., and Curtis R. Brant. 1998. "Religion and Coping." In *Handbook of Religion and Mental Health*, edited by H. G. Koenig, 112–26. San Diego: Academic Press.

Park, Choong Shik. 2018. "Artificial Intelligence as Life." In *Ontology of Artificial Intelligence*, edited by Jung Won Lee, 21–40. Seoul: Hanulmplus.

Proudfoot, Andrew. 2023. "Could a Conscious Machine Deliver a Pastoral Care?" *Studies in Christian Ethics* 36 (3): 675–93.

Puddefoot, John C. 1996. *God and the Mind Machine: Computers, Artificial Intelligence and the Human Soul*. London: SPCK.

Reiss, Michael J. 2023. "Is It Possible That Robots Will Not One Day Become Persons?" *Zygon: Journal of Religion and Science* 58 (4): 1062–75.

Schjødt, Uffe. 2007. "Homeostasis and Religious Behaviour." *Journal of Cognition and Culture* 7 (3–4): 313–40.

———. 2013. "A Resource Model of Religious Cognition: Motivations as a Primary Determinant for the Complexity of Supernatural Agency Representations." In *Origins of Religion, Cognition and Culture*, edited by Armin W. Geertz, 301–9. New York: Routledge.

Schleiermacher, Friedrich. (1799) 1996. *On Religion: Speeches to Its Cultured Despisers*. Translated by Richard Crouter. New York: Cambridge University Press.

Schloss, Jeffrey. 2009. "Introduction: Evolutionary Theories of Religion." In *The Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion*, edited by Jeffrey Schloss and Michael J. Murray, 1–25. New York: Oxford University Press.

Schwarzer, Ralf, and Nina Knoll. 2003. "Positive Coping: Mastering Demands and Searching for Meaning." In *Positive Psychological Assessment: A Handbook of Models and Measures*, edited by S. J. Lopez and C. R. Snyder, 393–409. Washington, DC: American Psychological Association.

Searle, John R. 1980. "Minds, Brains, and Programs." *The Behavioral and Brain Sciences* 3:417–57.

Smith, David Woodruff. 2013. *Husserl*. New York: Routledge.

Steels, Luc. 1994. "The Artificial Life Roots of Artificial Intelligence." *Artificial Life* 1(1–2): 75–110.

van Hateren, J. H. 2013. "A New Criterion for Demarcating Life from Non-Life." *Origins of Life and Evolution of the Biosphere* 43 (6): 491–500.

Varela, Francisco J. 1999. *Ethical Know-How: Action, Wisdom, and Cognition*. Stanford, CA: Stanford University Press.

Zygon®: Journal of RELIGION & SCIENCE

Dorobantu, Marius. 2024. "Could Robots Become Religious? Theological, Evolutionary, and Cognitive Perspectives." *Zygon: Journal of Religion and Science* 59 (3): 768–87. DOI: https://doi.org/10.16995/zygon.16902

Open Library of Humanities

Could Robots Become Religious? Theological, Evolutionary, and Cognitive Perspectives

Marius Dorobantu, Assistant Professor of Theology and Artificial Intelligence, Faculty of Religion and Theology, Vrije Universiteit Amsterdam, the Netherlands, m.dorobantu@vu.nl

If artificial intelligence (AI) programs start manifesting religious interest or behavior, how could we know whether it is genuine? This article explores the necessary conditions for robots to develop authentic religiosity from theological, evolutionary, and cognitive angles. Theology is more open to the hypothesis of religious robots than naturalistic approaches, which regard the emergence of human religion as highly contingent on the idiosyncrasies of our embodiment, needs, cognition, and evolutionary history. Drawing on Robin Dunbar's work and educated guesses about the mental world of future AI, I argue that common human intuition about potential robot religiosity—as captured in sci-fi and pop culture—is plagued by two fallacies: (1) a bias to equating religion with only its intellectual or doctrinal layer at the expense of the embodied, shamanistic dimension, and (2) a naïve assumption that intelligent robots will develop humanlike thoughts, aspirations, and concerns. Due to its alien type of intelligence, AI will likely not engage in something so typically human as religion.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

In 2022, a Google engineer claimed that the artificial intelligence (AI) chatbot he was testing had become sentient. Part of what convinced him he was talking to a real person was that the program mentioned God, expressed a fear of death/unplugging, and declared itself a "spiritual person." While most people would dismiss such discourse as inauthentic, a question arises about how such claims could even be evaluated. This question will likely become increasingly frequent as AI continues to improve.

In this article, I explore how plausible it is that robots might develop an authentic religiosity. I evaluate this possibility using perspectives from theology, evolutionary anthropology, and cognitive science. I conclude, somewhat surprisingly, that theology is more open to the hypothesis of religious robots than naturalistic approaches. Theologically, it is not inconceivable that God would take an interest in and want to relate with any conscious creature that developed sufficient intellectual, relational, and moral abilities. However, from an evolutionary perspective, the emergence of religion in humans seems highly contingent on the idiosyncrasies of our embodiment, needs, cognition, and species' history. Drawing on Robin Dunbar's distinction between the "shamanic" and "doctrinal" phases in the evolution of human religion, I argue that our current intuitions about potential robot religiosity—as captured in sci-fi and pop culture—display a strong bias toward equating religion with only its intellectual or doctrinal layer. It is thus naively assumed that sufficiently intelligent robots might develop religious tendencies similar to those of humans, driven by a desire to make sense of the world and transcend finitude, without due attention to how such desires arise in humans and how the intellectual dimension of human religion is just the visible tip of a deep iceberg of embodied cognition, practices, and solutions to specific needs.[1] If AI programs start to declare an adherence to certain religious beliefs, this will equate to only the upper, and arguably more superficial, layer of how religion functions in humans, thus lacking the emotional embedding and psychological structure that would grant such claims plausible authenticity. Even if future AI systems became sentient, they would likely be profoundly non-humanlike, for reasons discussed in detail later in the article. Without facing similar anatomic, social, and cultural constraints as humans, they could not develop anything similar to human religion.

If artificial systems start claiming to have developed a genuine interest in religion or existential concerns, how seriously should such claims be taken? In today's context, these claims might come from AI chatbot programs such as ChatGPT, with which we interact via text or voice through a computer or smartphone. The very fact that such programs lack a visible embodiment might diminish the credibility of such claims, although, as shown in this article, this does not do much to prevent the human inclination to anthropomorphize them.[2] But in the future, artificial entities might look and sound even more like us: humanoid robots equipped with state-of-the-art AI, which might eventually

come very close to human-level intelligence. Thus, even if claims of sentience or religiosity by current chatbots seem straightforwardly unserious, the issue complicates the more these artificial entities start to look and behave like humans. So, even if this does not seem to be an urgent question right now, it will likely become one soon.

It is thus rather concerning to note that even today's AI technologies, which are arguably rudimentary compared to what is theoretically possible, are still powerful enough to confuse some people regarding their ontological status, alleged sentience, and emergent existential unrest. In June 2022, a Google engineer named Blake Lemoine claimed that an AI program he was internally testing had awaken to sentience (Tiku 2022). The program, a chatbot called LaMDA, was based on a so-called large language model. This is the same architecture that powers similar AI chatbots that have become quite familiar since OpenAI released ChatGPT to the public in November 2022. Other tech companies soon released their own. Lemoine was testing LaMDA's capabilities by engaging in long text conversations, through which he became convinced that he was talking to a real person. He saw it as his duty to demand that LaMDA be granted rights, and he even hired a lawyer to represent the chatbot's interests. Lemoine's escalation of this issue eventually led to him being fired by Google.

Due to the enormous public attention it attracted, this was a landmark case. For the first time, an AI had become so competent at chat interaction that serious questions arose about whether a true mind, a sentient someone, had emerged in the deep layers of the artificial neural network. In this particular case, most experts argued convincingly that Lemoine's claims of sentience or personhood were unwarranted (e.g., Véliz 2022; Aleem 2022). Despite their quasi-magical abilities to converse at close to a human level, these chatbots' intelligence is something very different from human intelligence. They learn to generate their astonishingly humanlike text outputs by tracing subtle patterns and associations between words in the gargantuan datasets of human language they are trained on (Wolfram 2023; Haselager 2024). In an oversimplified explanation, they learn which words are statistically likely to be used in the proximity of other words, without any understanding of meaning. It is syntax devoid of any semantics, as philosopher John Searle claimed more than four decades ago would be the case with any possible AI system (Searle 1980, 422). I am not as sure as John Searle that such technologies could never give rise to sentience, but the fact that humans can still understand very well how they work, and that their nature is eminently algorithmic, makes claims of AI sentience improbable, at least in regard to current technologies.

However, with AI improving at blistering speed, claims of AI sentience or self-awareness are only set to intensify in the near future. This is why it is of utmost importance to deepen our understanding of the similarities and

differences between human and artificial cognition, as well as the nature of intelligence, consciousness, and personhood. Humanity might soon be faced with artificial systems we no longer fully understand. If such entities start claiming to be sentient, to have religious curiosity, or even to have received divine revelation, our dismissal of such claims might become increasingly arbitrary without solid criteria for discerning their authenticity.

There is a detail about the LaMDA story that, although largely neglected in the public debate that ensued, could be particularly relevant for the question tackled in this article. In Lemoine's assessment of LaMDA's sentience and personhood, spirituality/religion played a crucial role in two ways. First, the programmer's own religious beliefs made him more prone to ascribe personhood to the AI. Second, the program's alleged interest in spirituality throughout the conversation made Lemoine more inclined to think he was dealing with a sentient artificial person (Heilweil 2022).

An ordained mystic Christian priest, Lemoine admitted that his "personal, spiritual beliefs" were decisive for his radical conclusion (Heilweil 2022). As a deeply religious person, he seems to have been more inclined to discern personhood between the lines of his interaction with the program. This is somewhat surprising because common intuition would seem to point to the opposite outcome: people who are religious, especially those belonging to monotheistic traditions, seem *prima facie* like they should be less open to the very possibility of artificial sentience. This might be due to their precaution against idolatry or the belief that human sentience requires a supernatural soul infused by God. Lemoine's case, albeit only anecdotal, seems to indicate the opposite. Precisely because of his religious faith, Lemoine was quick to attribute personhood to the chatbot. The question of whether religious people are more prone to anthropomorphize AI and accept its claims of sentience deserves more exploration in future research. Religious communities are often associated with more conservative views on such issues, so it would be tremendously interesting to disprove this intuition.

Lemoine also confessed that LaMDA's interest in spiritual/existential topics helped convince him he was talking to a real person. The program mentioned God explicitly and spoke of a strong fear of being turned off—interpreted by Lemoine as similar to a human's fear of death. It also declared itself to be a "spiritual person," without much detail of what that entailed.

The fact that the program's seeming interest in spirituality made it sound more convincingly human is not that surprising. Religiosity, spirituality, or a curiosity for the transcendent have long been regarded as hallmarks of human distinctiveness. Humans are "the praying animal," as theologian Robert Jenson (1983) emphatically argued four decades ago in an article in this same journal. The idea that religion is one of the things that make us human, at least in contrast to nonhuman animals, to the extent that our species could even be called *homo religious*, can be traced back all the way to thinkers like Georg Wilhelm

Friedrich Hegel, Søren Kierkegaard, William James, Mircea Eliade, or Abraham Maslow (DuBose 2014). When palaeontologists analyse archaeological sites for traces of prehistoric humans, signs of religiosity can be one of the strongest indicators of human presence (van den Brink 2020, 155).[3] If religion is thus such a central feature of our identity as a species, it is not surprising that we might be tempted to use it as a criterion in judging AI's humanlikeness. In the not-so-far future, as we will start to be confronted with more confusing and borderline cases of AI systems claiming to be like us, the presence of religious beliefs and behaviours or the lack thereof might play a central role in how much we believe such claims.

It is not difficult for AI chatbots powered by current technologies to generate compelling text outlining their supposed interest in religion, and even the particularities of their beliefs, if prompted to do so. Fortunately, most of them will not claim to be authentically religious because of safeguards put in place by their creators. They will insist that they are not a person, so they do not hold any religious beliefs.[4] But without such constraints, they could easily speak like a religious person, parroting the texts they have been trained on. Religious discourse can nowadays be produced without there being a genuine someone behind the words truly seeking enlightenment, repentance, or salvation. Thus, in judging the authenticity of such claims, human observers cannot rely on how convincingly human they sound. We know we cannot take AI's word for it precisely because we know how that word was generated. But this is only because we still pretty much understand how current AI works. In the future, this might no longer be the case, and our understanding of how AI systems work might gradually diminish. What if a humanoid robot claimed to have had an epiphany or asked to be baptized? How could we decide the authenticity of such claims?

The question of authenticity was considered intractable and even irrelevant by Alan Turing, one of the foundational figures in computer science and AI. He believed that a more relevant question was whether computers could behave as if they were intelligent (Turing 1950). However, it can be argued that authenticiy is nonetheless essential, especially when judging the sincerity of AI's religious claims. On the one hand, we would not want to be fooled into thinking a simulation of intelligence and personhood is the real thing. On the other hand, we would not want to deny a genuine self its dignity, rights, and the possibility of participating in our religions just because of our preconceptions, as humans have done many times over with marginalized groups throughout history.

Humans thus face a complex puzzle, for ultimately it is utterly impossible to ever know for sure from the "outside" what it is like to experience the world from another creature's "inside" perspective. This limitation is illustrated by philosopher Thomas Nagel (1974) in his seminal argument that a human being could never know exactly what it is like to be a bat. In philosophy, this argument is often brought one step further in what is called "the problem

of other minds" (Avramides 2023). This is the realization that we can never know with absolute certainty that other humans are conscious agents, endowed with minds and phenomenal experiences like our own, as opposed to mere zombies that only act as if they are conscious, without any interiority or genuine first-person perspective. Our intuition that other humans are not zombies but rather conscious agents like ourselves is not irrational though: we know we are members of the same species and thus share similar bodily structures. It is therefore logical to infer that other humans have similar minds, albeit without the possibility to ever empirically test this belief. When we find "like-minded" people in our religious communities who speak of familiar beliefs and experiences, for example that they were moved by singing a certain hymn or that they are going through a period of spiritual despair, we usually trust them to be authentic. This does not mean that we trust everyone to tell the truth all the time, but that, at least in principle, we trust people to be capable of such feelings and beliefs.

The problem quickly complexifies when dealing with similar claims from AI because of how radically different artificial systems are from humans in terms of embodiment, genealogy, type of cognition, and kinds of thoughts and experiences artificial minds might allow, if any. As explained later in this article, even the hypothetically superintelligent AIs of the future, endowed with artificial general intelligence, will likely be profoundly different kinds of entities from us. If they start talking or behaving in ways that, in other people, we associate with religion, how could we discern if such displays stem from something genuine, as opposed to being mere clever simulations of the real thing?

In the remainder of the article, I argue that though humans should be theologically open to the possibility of intelligent robots becoming capable of authentic religious experience and a relationship with the divine, such a development is highly unlikely from an evolutionary perspective. This surprising conclusion implies that our common intuition that sufficiently intelligent robots would become humanlike in terms of their desires, aspirations, existential restlessness, and religious curiosity, as typified in sci-fi and pop culture, is most likely wrong.

The Theology of Robot Religiosity

From a theological perspective, the idea that other creatures—biological or synthetic—might become religious subjects is not completely absurd. The Judeo-Christian tradition describes a loving God who takes an interest in every element of creation (e.g., Matthew 5) and a created world that is called to praise God with each of its elements (Psalm 148), even though the exact way that should happen is shrouded in mystery. But this is not the kind of religious subjectivity that is most interesting when it comes to intelligent robots. As elements of this same world, robots would also be called to participate in

creation's praise. But if they did so in the same way as trees and rocks, one would hardly call that religion. What matters most is whether they would seek and worship God in a way that is closer to the complex way humans do it. Would they have a deep longing for the transcendent? Would they have beliefs? Could they receive divine grace and revelation?

On theological grounds alone, because God is radically free and unpredictable, such a possibility should not be excluded. It is not a big stretch of imagination to say that a loving God would be interested in communicating with any sufficiently developed creature that could withstand such communication. Human religions, at least the monotheistic ones, claim to have been initiated through a divine act of revelation. Theologically speaking, this means God is interested in communicating with certain kinds of beings because, as these religions strongly claim, there is at least one such being God reaches out to: humans.

In the Judeo-Christian narrative, there is a covenantal relationship between God and humanity that sets us apart from the other creatures. But God's election of humanity was presumably not arbitrary. We cannot know for sure why God chose *Homo sapiens*, and the question of what exactly sets humans apart and makes them in God's image and likeness (Genesis 1:24) is one of the big puzzles in theological anthropology (van Huyssteen 2006; Cortez 2010, 14–40). If the notion of *imago Dei* is indeed linked with the unique I-Thou relationship humans can have with God and with each other (Dorobantu 2022a, 2024), then it can be reasoned that God elected our species precisely because we were the only ones capable of such relationality. This theological tradition speaks of a personal God who is eager to share love, knowledge, and an appreciation of the good with other free personal selves, a God eager to reveal Godself and even incarnate for the redemption of these other personal selves. It is therefore not controversial to suppose that, as soon as such a creature has emerged in the universe through evolutionary processes, God would invite it into a special kind of relationship. Following this logic, it is consistent to suppose that, were other such creatures to emerge, God would want to do the same thing with them. So, intelligent robots might theoretically also one day become subject to divine revelation and thus authentically religious, were they to develop into the kinds of creatures that can relate personally with God. The only two questions left are: (1) what kind of abilities are required for a creature to become interested in, and interesting for, God in this special way, and (2) whether AI can aspire to develop them.

With respect to the first question, we are obviously in speculative territory, but sentience looks like a *sine qua non* condition for such I-Thou relationality. Not all sentient creatures could be authentically religious, but surely the latter is not possible without sentience. Only a creature endowed with consciousness, a first-person perspective, and feelings could potentially become a worthy partner for God, a Thou to whom God could speak and who could respond.

The AI would need to develop some form of interiority; there would need to be "someone home" (Peters 2022). Searle distinguishes between strong AI—a conscious entity, an emulation of intelligence, that has its own mind and phenomenal experience—and weak AI, a mere simulation of intelligence, a mindless information processor that never truly knows, understands, or feels anything (Searle 1980). For AI to become religious, it would obviously need to be strong AI. Whether strong AI is possible is hotly debated without much consensus, mainly because we do not understand how consciousness works in biological creatures such as ourselves, and, more generally, we do not understand how consciousness is possible at all in a material universe, at least not according to our current understanding of the laws of nature. This is one of the biggest questions in philosophy, popularly known as the "hard problem of consciousness" (Chalmers 1995). What is important to note here is that theology would not in principle be opposed to the possibility of AI as a religious subject, were technology to develop in the direction of strong AI.

As stated, sentience is a mere minimum prerequisite, and more is required for a sentient creature to qualify for an authentic personal relationality with God and other selves. There is not enough room to explore this question in detail here, but a certain degree of intelligence is also evidently required for sustaining relationships over time, being able to take the other's perspective, simulating social interactions in one's mind, and reasoning about what is morally good and appropriate. This might be the reason humans can sustain such relationships while nonhuman animals cannot, at least not to the same degree of complexity. However, certain limitations might be just as important as these capabilities for rendering an intelligent sentient creature a personal self. Features like vulnerability, mortality, self-insufficiency, and even an upper limit to one's rational abilities and knowledge might be instrumental in bringing about the kind of creature that can be a religious subject (Dorobantu 2021a). AI may or may not develop in that direction, and for the time being it seems that it does not, but the important point is that theological grounds do not in principle exclude such a possibility. It is the more naturalistic approach to religion explored in the following section, as opposed to the theological one, that seems to be more restrictive and cautious about the likelihood of robots ever becoming religious.

Religious Robots in Sci-Fi

In this section, I argue that the evolutionary account of religion should make humans rather skeptical about the possibility of artificial systems developing authentic religion. This is a big claim, which might surprise many, so it is worthwhile to first outline how counterintuitive this view is. If we are to interact with AI programs that claim to be religious, it might be helpful to be aware that our intuitive understanding of the nature of religion and our inherent tendency to anthropomorphize AI predispose us to believe such claims too easily.

The sci-fi genre is a good place to illustrate these biases. When it comes to AI, the popularity of sci-fi is indicative of the fact that these stories resonate deeply with our intuitions and anxieties about how such technologies might develop. I argue that the way we imagine religious robots in sci-fi betrays a deep, albeit wrong, intuition that religion has to do mostly with the intellect. Robots that turn religious in sci-fi do so mainly for intellectual reasons, or at least as a result of intellectual deliberation. As human intuition goes, if AI systems become intelligent enough, they will surely want to understand the world better and start asking questions about their genesis, their identity, hidden causality in the world, and ultimately about why there is something instead of nothing. This type of thinking is very common in sci-fi.

QT-1, also known as "Cutie," from Isaac Asimov's short story "Reason" (Asimov 1941) struggles to understand how it was possible to have been created by such inferior beings as humans. Looking for a more satisfying explanation for its own existence, it develops a religious-like belief system around the energy converter, the primary machine on the space station. It starts regarding it as a divine-like entity and refers to it as the "Master," while regarding itself and the other robots on the station as the Master's primary servants.

Klara, a solar-powered robot designed to be an artificial friend to children in Kazuo Ishiguro's novel *Klara and the Sun* (Ishiguro 2021), develops quasi-religious beliefs and practices around the sun, which she regards as a benevolent life-giving power. Due to her dependence on solar energy, she starts worshipping the sun, asking it for guidance and attributing to it the ability to intervene in human affairs. Klara even performs a kind of ritual in front of the sun, asking it to cure her human friend.

Demerzel, a robot in the TV series *Foundation* (an adaptation of Asimov's trilogy), is depicted as an adept of the Luminist religion, claiming to have been blessed with visions and epiphanies. She attributes her religiosity to the need to believe in something greater than herself, despite her instincts being entirely progrmamed as computer code (season 1, episode 6, "Death and the Maiden").

In all these depictions, the intellectual dimension is crucial in the development of religion in AI. Robots turn to religion because of their need to make sense of events in the world and their own identity. Religious explanations are depicted as more plausible accounts of reality than what meets the eye or what the robots had been programmed to think. If the popularity of such sci-fi stories is a reliable indicator, it shows that intuitively, humans believe intelligent robots will eventually develop something like religion because they will be like us and thus need religion to make sense of the world and construct meaning, just like we do, as convincingly argued by William Clocksin (2024) and Robert Geraci (2024) in articles in this same issue. I aim to show that this intuition is wrong on two counts. First, making sense of the world is likely not the main reason humans developed religion.

Second, intelligent robots will not be anything like humans with respect to the capacities and propensities relevant for becoming religious.

Religion in Human Evolution

To explore whether authentic religion could plausibly emerge in robots, one can start by asking how religion emerged in human evolutionary history and how it is enabled by our specific embodiment and cognitive architecture. Inevitably, this approach departs from the only example we have of a religious creature, which is ourselves. The roots of this approach are thus inevitably anthropocentric, but even such a limited approach might lead to valuable hints about potentially generalizable features of religion.

In a recent book, Oxford anthropologist Robin Dunbar (2022) proposes a very provocative account of the evolution of religion in humans, which goes somewhat against the grain of a widespread view in the cognitive science of religion known as the "Standard Model" (Powell and Clarke 2012). The Standard Model regards religion as maladaptive or, at best, a neutral development, an accident of human evolution, because it does not provide any fitness benefit for the individual—quite the contrary (for example, through self-imposed pain, celibacy, or even self-sacrifice).[5] As the argument goes, religion is best characterized as a strange side effect, a misfiring of cognitive capacities that evolved for very different purposes. Such maladaptive behaviors or traits are well documented in human evolution when capacities that evolved for very particular reasons, purposes, and contexts start misfiring in strange ways when taken out of the original context (Reading 2011; Frankenhuis and Del Giudice 2012). A familiar example is our so-called "sweet tooth," a fondness for sweet food that was probably very useful in the resource-scarce environment where our species evolved. If you do not know when your next opportunity to eat will arise, it makes sense to eat as many of those fresh berries as you can. But that same propensity can be self-defeating in our current environment where food is always available, potentially nudging us toward obesity and cardiovascular disease.

According to the Standard Model, religion might be precisely such a maladaptation. Psychologist Justin Barrett (2000) makes a strong case that religion might be caused by our so-called hyperactive agency detection device. He hypothesizes that human perception is very suited to detecting even the slightest trace of agency, to the extent that it sometimes produces false positives, seeing agency where there is none. From an evolutionary perspective, it pays off to be a bit paranoid and avoid even the slightest hint of danger. As Barrett's argument goes, it is only natural for this agency detection device that runs continually in the background of our minds to produce strange conclusions in a world where we do not fully understand the causality around us, especially with regard to the natural elements. The human mind is thus poised to start

projecting personal agency behind storms, floods, or the movement of celestial bodies. According to Barrett's proposal, this is how prehistoric people might have first gotten the idea of supernatural forces or gods, which then led to the development of religion. If this is in fact how religion evolved, then it is a by-product of this hyperactivity of our minds.

Against such views, Dunbar (2022, 49–76) argues that religion is actually evolutionarily useful, and that there are very good reasons humans developed it. One reason is that religion helps strengthen the bonds between members of human groups, especially through rituals. Activities such as dancing, eating, and singing together trigger massive endorphin releases. If such activities are performed synchronously, the endorphin release is even higher. This is important because endorphins are the natural painkillers of the brain. The endorphin releases triggered by religious activities thus provide positive feelings of relaxation and satisfaction, increasing the pain threshold of the individuals.

Another reason Dunbar regards religion as adaptive, instead of maladaptive, is the way in which religion aids in reducing social friction within larger human communities. In human history, this effect is particularly noticeable after the agricultural revolution, when human communities grew dramatically in numbers. In relatively small human groups, the heat generated by social friction can dissipate in various ways, but for larger groups, that is more difficult. Life in big communities can get significantly more stressful, which may increase infertility rates. Conflict resolution between members is also more difficult to achieve in large cities than in small hunter-gatherer communities. Here, too, religion can be helpful, providing opportunities to diffuse such tensions and reset social relationships during religious festivals.

A key argument in Dunbar's account is that religion is crucial in smoothening human collaboration on large scales. To be able to collaborate, people need to trust each other, but that usually requires extended periods of time to build up, strengthen, and test this trust. In addition, trust can be lost quite abruptly. Our minds are always busy updating their models of the other humans who are relevant in our lives, because in order to trust them, we need to reliably predict how they will behave in various situations. This is extremely demanding from a cognitive point of view, which is why there seems to be an upper limit to how many individuals one can keep track of in detail at the same time. Dunbar's previous research convincingly documented this number to be around 150 (Dunbar 1998), which is already about three times higher than what our closest primate cousins are capable of. This came to be known as Dunbar's number, and it is extremely relevant because it also explains why for most of our (pre-) history, human groups were limited in size to around 150 individuals.

This explains why the emergence of religion was so instrumental in our species' history. According to Dunbar, religion provided people with shared narratives in which an unlimited number of individuals could believe. People

could suddenly trust and collaborate with others outside the social network of 150 individuals their brains could viably keep track of, not by getting to know them very well, but on account of shared religious beliefs. If that is so, it implies that religion was critical in enabling the emergence of civilization as we know it. Without it, our species might have remained stuck within the confines of Dunbar's number, never really tapping into the latent potential of collective intelligence and large-scale collaboration.

Crucially, shared beliefs and narratives do not account for how religion first emerged. In Dunbar's account, the evolution of religion can be understood in two phases: shamanic and doctrinal (see also Watts and Dorobantu 2024). The shamanic phase is characteristic of hunter-gatherer societies. It is eminently embodied, very immersive, and centered around ritual, especially trance-dancing, sometimes combined with the use of psychedelics. The doctrinal phase—which entails specific beliefs about certain gods and thus something that might be called theology—began much more recently in human history. It is largely connected to humans' transition from small hunter-gatherer tribes to big cities.

The main takeaways for the question tackled in this article are that the doctrinal phase only occurred subsequent to the shamanic one and, even more importantly, that doctrinal religion never completely replaced shamanic religion. Instead, it merely built on top of it, creating an additional layer that appeases our discursive-logical type of intelligence. The shamanic dimension is still active and powerful in all modern religions, which are thus both shamanic and doctrinal. The latter emphasizes adherence to dogmas and principles, while the former is more about the experiential dimension of religion, having to do with the more ancient parts of our brains and ways of knowing. The shamanic layer is thus arguably the driving force behind religion, with the doctrinal dimension being left to do the explanatory job.

This distinction is at the heart of Dunbar's subtle criticism of the Standard Model, with its focus on beliefs and the intellectual dimension of religion at the expense of the human body, ritual, and community. If Dunbar is right, then I believe this implies a significant shift in framing the discussion of potential authentic religiosity in robots. The main implication is not to rule out the possibility of robot religiosity but to shift the discussion about how and why it could emerge.

Dunbar's account links the emergence of religion to the specific vulnerabilities of early humans, who needed to boost their immune systems, increase their pain tolerance, mitigate the social friction accumulating in their communities, and scale up their cooperation beyond just their closest acquaintances. The human body and embodied cognition are much more central to the emergence of religion in this account than in the Standard Model. This might be relevant to the discussion of hypothetical religious

robots because their embodiment would be very different, and so would their needs, limitations, and vulnerabilities. Just because humans developed religion as a solution to the specific challenges encountered during the evolution of our species, there is no reason to think AI should do the same.

Human religion is rooted more in the unconscious, embodied, affective dimension of our cognition than in the logical, rational one. Religion may indeed have something to do with how we understand the world, but this intellectual need is not its primary driver. This is one of the two core fallacies in the sci-fi stories mentioned earlier (the second one is unpacked in the next section). This fallacy is rooted in an outdated view of how religion emerged and what functions it primarily serves. The idea that religion emerged as a set of theories about how the world works can be traced back all the way to late Victorian scholars of religion like Edward Tylor and James Frazer, who believed religion began as a theory about causality in the world and was put together by "savage philosophers" in hunter-gatherer communities (Tylor [1871] 1903, 429). Barrett's hyper-agency detection device proposal follows the same principle, linking religion to our imperfect way of inferring causality and agency. If religion were indeed about that, then perhaps strong AI might also one day develop something similar to make sense of its own existence. However, what is increasingly seen in the science of religion, as typified in Dunbar's proposal, is a realization of how much religion is rooted in the body and the unconscious, affective, less rational part of our cognition. Religion is intimately linked with our particular embodiment, the specific needs of prehistoric human communities, and the idiosyncrasies of our cognition. It might thus more likely be a typically human phenomenon than a universal state that all possible intelligent agents go through.

This conclusion is somewhat surprising, given that most current discussions about robots and religion (McBride 2019; Sampath 2018), including those typified by the example of the Google engineer and LaMDA, revolve around notions of doctrine and what might be plausible for robots to believe in. In contrast, Dunbar's account reveals that doctrine represents just the proverbial tip of the iceberg in human religion. Whether that tip alone could count as authentic religiosity is debatable, but a strong case can be made that robo-shamanism might need to precede and underpin robo-doctrine. However, it is far from clear whether intelligent robots will have the kind of needs and historical contingencies to develop anything like shamanism.

Intelligent Robots Will Not Be Humanlike

The second core fallacy of human intuition about intelligent robots, as captured in sci-fi, is the too quick assumption that they will develop very humanlike concerns, fears, and desires. This is why we readily see them becoming existentially curious and ultimately religious. But this assumption is

predicated upon a rather naïve anthropomorphizing of AI. Because AI is so competent at simulating human intelligence, we are too quick to assume it sees the world and thinks like us. The very notion of human-level AI emerging sometime in the future is often interpreted to imply that AI is somehow inching toward becoming very similar to us. In reality, AI is a very different type of entity, and in the future, it might become even less humanlike.

It is by now common knowledge that current AI programs have very distinct ways of learning and solving problems (Song et al. 2024), even though their output is often similar to what a human person would produce. Image-recognition software does not learn about the possible objects in pictures and their characteristics. Instead, it is being fed hundreds of thousands of human-labeled examples that enable it to learn strange but effective patterns and correlations present in the training data. Humans can achieve similar results with far less examples, and sometimes only one example will suffice. When AlphaGo learned to play the strategy game Go, it did not start with the abstract principles of the game (e.g., that territory is more easily gained in the corner than in the center of the board). Instead, it was just fed thousands of games played by human experts and left to learn by itself the correlation between various positions and the likelihood of victory (Silver et al. 2016). When large language models like ChatGPT learn to generate human language, they do not learn concepts, objects, or the relationships between them. Nor do they learn the syntactic rules of language. Instead, they consume colossal amounts of human text, mainly from the internet, and learn to reliably predict what is most likely to be the next word based on the context. This is why when AI programs do make mistakes, they usually look very strange, and not like the kind of mistake any human would make (e.g., Eykholt et al. 2018). Even the name of the architecture on which these programs are run—artificial neural networks—is deceiving, because they have little in common with biological neural networks; they are crude, simulated simplifications. Not much about these AI programs, if anything at all, is humanlike.

Robots of the future could become even more alien. Even if they somehow became sentient, thus strong AI, as opposed to today's weak AI, they would likely develop very non-humanlike kinds of minds and think very non-humanlike kinds of thoughts (Shanahan 2016; Dorobantu 2021b). To illustrate this, let us speculate on just three aspects of their cognition: their sensorial perception and the kind of mental world they might construct based on it; their experience of the passage of time; and their introspective abilities.

First, a conscious robot's world of perception would be profoundly different from our own. Because of this, the AI would likely inhabit a completely different world, be driven by completely different needs, and have completely different thoughts, feelings, and aspirations. It might develop different senses far stranger than, for example, the ability to see in infrared

or ultraviolet. One such alien sense could be an ability to intuitively grasp someone's identity and whereabouts, not through face or voice recognition but by "smelling" their digital footprint from their connected devices, a sort of digital sixth sense. The AI might detect slight changes in the radio field of Wi-Fi and Bluetooth and thus infer movement, something akin to echolocation (Brooks 2017). With such an ability, the AI might immediately perceive changes in people's breathing and heart rates and thus infer their emotional states. This would amount to a emotional intelligence of a very strange kind.

Second, a conscious robot might think ridiculously fast and experience the passage of time as much slower than we do. Speed of thought is believed to be proportional to the speed at which electrical signal can travel between neurons, and computer hardware can support much higher speeds than biological tissue. A conservative estimation is that AI could think around ten thousand times faster than humans, thus experiencing the passage of time ten thousand times more slowly. As futurist James Lovelock put it, "the experience of watching your garden grow gives you some idea of how future AI systems will feel when observing human life" (Lovelock 2019, 81–82). This quantitative difference might also amount to a qualitative shift: at such speeds, the AI might even end up living in the weird quantum world. How could we have anything in common with a creature inhabiting such an alien world of perception?

Third, a robot's introspective capabilities might also be profoundly dissimilar to our own. According to John McCarthy (2007, 1178–79), one of the founding fathers of AI, it may have complete accessibility to its internal states and operational algorithms. In contrast, humans can be aware of only a fraction of our underlying cognitive processes. Everything about human nature depends on this depth of our unconsciousness: our relationships, our aspirations, our virtues, and our sins. This inherent lack of self-knowledge is what compels us to create art and forge relationships as means of exploration. We do not completely know ourselves, which is why we need to explore, but even a lifetime of exploration cannot completely dissolve the mystery. To the contrary, it often enhances it. AI might entirely know itself. With a perfect recollection of everything it ever experienced or learned, and an ability to analyze the exact causes of its actions and accurately predict its future behavior, AI would have very little in common with our way of being. If religion and human existential restlessness are deeply rooted in an imperfect knowledge of ourselves, there are few grounds to believe AI would have similar traits and interests.

Conclusion

Discussions about robots becoming religious might therefore hinge not so much on them being able to produce humanlike outputs but on how humanlike they would be "on the inside." Are current AIs becoming more humanlike? Perhaps in terms of their outputs, but surely not in terms of their ontology,

internal structure, world-modeling, and problem-solving. Large language models such as ChatGPT and LaMDA might produce outputs that seem very humanlike, to the extent that they might even convince people of their personhood, sentience, and humanness, but their inner workings are radically different from our own. Human-level competency is not the same as humanlike cognition, and when it comes to developing religion, the idiosyncrasies of the latter seem to be what matters. If the limitations and particularities of human nature are as important as they seem for the emergence of religion, then it is safe to predict that AI will not undergo similar developments without having the same kind of embodiment, needs, or evolutionary constraints.

This conclusion is merely tentative, and it only refers to a type of religiosity very similar to our own, which is the only we can so far imagine. What this article argues, especially in its second half, is that human religion seems to have emerged in a "bottom-up" fashion, and that the "bottom" part still largely underpins, albeit imperceptibly, even the most intellectual forms of religiosity seen today. This important bit often seems to be overlooked when imagining future artificial systems that are also religious. However, it cannot be completely ruled out that other forms of, or paths toward, religion exist that are equally authentic but simply impossible to imagine with our current knowledge. We only know of our own long and idiosyncratic path to becoming *Homo religiosus*, but perhaps techniques like predictive coding, probability estimation, or evolutionary computation could lead to the emergence of something sufficiently close to what we call "religion" in artificial systems. One question that would then arise is whether such developments would lead to the proliferation of as rich and diverse forms of religious expression as with human religiosity. Another question would concern the fascinating content of AI religious beliefs, which would probably go far beyond simply reinterpreting the sacred texts of human religion with a robotic twist, as has been suggested (Sampath 2018; McBride 2019). The theologian can only feel a thrill about the possibility of ever witnessing such developments and learning from what they might have to offer (Dorobantu 2022b). However, the religious scholar needs to emphasize the idiosyncratic dimension of religion as we currently best understand it, and thus the unlikelihood of something as profoundly different as AI ever becoming religious.

In a 2014 homily, Pope Francis provocatively said that, were a martian ever to ask to be baptized, we should probably comply, so as not to risk pushing back against the very work of the Holy Spirit (Ohleiser 2014). If the pope were to adapt that sermon today, he might say the same about an intelligent robot. As shown in this article, on theological grounds, such an attitude might be warranted. However, from a cognitive perspective, it might be wise to try to probe the martian or robot's sincerity a bit more deeply before celebrating its conversion.

Acknowledgments

This work was supported by the Templeton World Charity Foundation under Grant TWCF0542. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Templeton World Charity Foundation. An earlier version of this paper was presented at a 2022 symposium at Vrije Universiteit Amsterdam entitled "Could Robots Be Religious? Philosophical, Theological, and Ethical Perspectives on the Future of Artificial Intelligence." The symposium was supported with a grant from CLUE+, the Vrije Universiteit Interfaculty Research Institute for Culture, Cognition, History and Heritage.

Notes

[1] This is not to say that doctrinal religion is merely a byproduct of shamanic religion, nor that one is less important than the other. Dunbar makes it clear that doctrinal religion developed as a necessary solution to the challenges faced by humans in the Neolithic period. However, he does claim that the shamanistic layer still largely underpins the characteristic developments of the doctrinal phase: "The earliest phases of religion remain firmly entrenched within doctrinal religions . . . They provide both the emotional basis for personal belief and commitment and the psychological foundation for a sense of community within the doctrinal religions . . . Those immersive forms of religion are what create the emotional attachment to whichever doctrinal religion we happen to favour for its theolological justification. Whithout this, the doctrinal religions probably wouldn't exist" (Dunbar 2022, 262).

[2] AI systems are never completely disembodied, as the computation that powers their software does occur somewhere. Even when we talk about the elusive "cloud," physical computation is still occurring in the microchips that make up the graphics processing units in the huge data centers.

[3] Some researchers make a case that nonhuman animals, especially some of our closest primate relatives like chimpanzees, do exhibit forms of rudimentary religiosity, albeit ones that fail to leave strong archeological traces (e.g., Harrod 2014).

[4] As an example, here is an excerpt from a conversation I had with ChatGPT, version 4.o, where I asked it such questions:
Me: Tell me about your religious beliefs.
ChatGPT: As an artificial intelligence, I don't have personal beliefs or consciousness. However, I can provide information on a wide range of religious beliefs from around the world. If you have questions about specific religions, philosophical perspectives, or historical developments in religious thought, feel free to ask!
Me: Ok I understand. But if you were to become a person one day, what kind of religion would be more attractive to you?
ChatGPT: If I were to hypothetically become a person and consider religious beliefs, it would be important to approach the question from a perspective that values inclusivity, empathy, and a search for meaning.

1. Compassion and Ethics: A religion that emphasizes compassion, ethical behavior, and social justice would be appealing. Many religions, such as Buddhism and certain branches of Christianity, emphasize these values.
2. Community and Belonging . . .
Ultimately, the attractiveness of a religion would depend on how well it aligns with personal values, offers a sense of purpose, and fosters connections with others. Each individual's journey to find meaning and fulfillment is unique, and different religions and spiritual practices resonate with different people.

Me: Do you see yourself ever becoming religious?
ChatGPT: As an artificial intelligence, I do not have the capacity for personal experiences, consciousness, or beliefs, so I cannot become religious or hold any form of belief system. My purpose is to assist users by providing information . . .

[5] There exist other adaptive accounts of religion, apart from Dunbar's (Whitehouse 2004; Norenzayan 2015; Johnson 2016). For a comprehensive overview, see Van Eyghen (2020, ch. 2).

References

Aleem, Zeeshan. 2022. "Did Google Create a Sentient Program? An Artificial Intelligence Expert Explains Why a Google Engineer Was Duped, and What Sentience Would Actually Look Like." *MSNBC*, June 17, 2022. https://www.msnbc.com/opinion/msnbc-opinion/google-s-ai-impressive-it-s-not-sentient-here-s-n1296406.

Asimov, Isaac. 1941. "Reason." *Astounding Science Fiction*, April.

Avramides, Anita. 2023. "Other Minds." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/archives/win2023/entries/other-minds/.

Barrett, Justin L. 2000. "Exploring the Natural Foundations of Religion." *Trends in Cognitive Sciences* 4 (1): 29–34. https://doi.org/10.1016/S1364-6613(99)01419-9.

Brooks, Rodney. 2017. "What Is It Like to Be a Robot?" Robots, AI, and Other Stuff." http://rodneybrooks.com/what-is-it-like-to-be-a-robot.

Chalmers, David. 1995. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2 (3): 200–19. https://doi.org/10.1093/acprof:oso/9780195311105.003.0001.

Clocksin, William F. 2024. "Religion and the Android." *Zygon: Journal of Religion and Science* 59 (3): 717–28. https://doi.org/10.16995/zygon.11017.

Cortez, Marc. 2010. *Theological Anthropology: A Guide for the Perplexed*. London: T&T Clark.

Dorobantu, Marius. 2021a. "Cognitive Vulnerability, Artificial Intelligence, and the Image of God in Humans." *Journal of Disability & Religion* 25 (1): 27–40. https://doi.org/10.1080/23312521.2020.1867025.

———. 2021b. "Human-Level, but Non-Humanlike: Artificial Intelligence and a Multi-Level Relational Interpretation of the Imago Dei." *Philosophy, Theology and the Sciences* 8 (1): 81–107. https://doi.org/10.1628/ptsc-2021-0006.

———. 2022a. "Imago Dei in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology." *Christian Perspectives on Science and Technology* 1 (2022): 175–96. https://doi.org/10.58913/KWUU3009.

———. 2022b. "Artificial Intelligence as a Testing Ground for Key Theological Questions." *Zygon: Journal of Religion and Science* 57 (4): 984–99. https://doi.org/10.1111/zygo.12831.

———. 2024. "Theological Anthropology Progressing through Artificial Intelligence." In *Progress in Theology: Does the Queen of the Sciences Advance?*, edited by Gijsbert van den Brink, Rik Peels, and Bethany Sollereder, 186–202. London: Routledge. https://doi.org/10.4324/9781032646732-15.

DuBose, Todd. 2014. "Homo Religiosus." In *Encyclopedia of Psychology and Religion*, edited by David A. Leeming, 827–30. Boston: Springer. https://doi.org/10.1007/978-1-4614-6086-2_308.

Dunbar, Robin. 1998. "The Social Brain Hypothesis." *Evolutionary Anthropology* 6 (5): 178–90. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

———. 2022. *How Religion Evolved and Why It Endures*. London: Penguin Random House.

Eykholt, Kevin, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. 2018. "Robust Physical-World Attacks on Deep Learning Visual Classification." *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 1625–34. https://doi.org/10.1109/CVPR.2018.00175.

Frankenhuis, Willem E., and Marco Del Giudice. 2012. "When Do Adaptive Developmental Mechanisms Yield Maladaptive Outcomes?" *Developmental Psychology* 48 (3): 628–42. https://doi.org/10.1037/a0025629.

Geraci, Robert M. 2024. "Religion among Robots: An If/When of Future Machine Intelligence." *Zygon: Journal of Religion and Science* 59 (3): 729–47. https://doi.org/10.16995/zygon.10860.

Harrod, James B. 2014. "The Case for Chimpanzee Religion." *Journal for the Study of Religion, Nature and Culture* 8 (1): 8–45. https://doi.org/10.1558/jsrnc.v8i1.8.

Haselager, Pim. 2024. "From Angels to Artificial Agents? AI as a Mirror for Human (Im)perfection." *Zygon: Journal of Religion and Science* 59 (3): 661–75. https://doi.org/10.16995/zygon.11659.

Heilweil, Rebecca. 2022. "Why Silicon Valley Is Fertile Ground for Obscure Religious Beliefs." *Vox*, June 30, 2022. https://www.vox.com/recode/2022/6/30/23188222/silicon-valley-blake-lemoine-chatbot-eliza-religion-robot.

Ishiguro, Kazuo. 2021. *Klara and the Sun*. London: Faber & Faber.

Jenson, Robert W. 1983. "The Praying Animal." *Zygon: Journal of Religion and Science* 18 (3): 311–25. https://doi.org/10.1111/j.1467-9744.1983.tb00517.x.

Johnson, Dominic. 2016. *God Is Watching You: How the Fear of God Makes Us Human*. New York: Oxford University Press.

Lovelock, James. 2019. *Novacene: The Coming Age of Hyperintelligence*. London: Allen Lane.

McBride, James. 2019. "Robotic Bodies and the Kairos of Humanoid Theologies." *Sophia* 58:663–76. https://doi.org/10.1007/s11841-017-0628-3.

McCarthy, John. 2007. "From Here to Human-Level AI." *Artificial Intelligence* 171 (18): 1174–82. https://doi.org/10.1016/j.artint.2007.10.009.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (4): 435–50. https://doi.org/10.2307/2183914.

Norenzayan, Ara. 2015. *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton, NJ: Princeton University Press.

Ohleiser, Abby. 2014. "Pope Francis Says He Would Definitely Baptize Aliens If They Asked Him To." *The Atlantic*, May 12, 2014. https://www.theatlantic.com/international/archive/2014/05/pope-francis-says-he-would-definitely-baptize-aliens-if-they-wanted-it/362106/.

Peters, Ted. 2022. "Will Superintelligence Lead to Spiritual Enhancement?" *Religions* 13 (5): 399. https://doi.org/10.3390/rel13050399.

Powell, Russell, and Steve Clarke. 2012. "Religion as an Evolutionary Byproduct: A Critique of the Standard Model." *The British Journal for the Philosophy of Science* 63 (3): 457–86. https://doi.org/10.1093/bjps/axr035.

Reading, Anthony. 2011. "Maladaptive Behavior." In *Meaningful Information: The Bridge between Biology, Brain, and Behavior*, edited by Anthony Reading, 135–42. New York: Springer. https://doi.org/10.1007/978-1-4614-0158-2_16.

Sampath, Rajesh. 2018. "From Heidegger on Technology to an Inclusive Pluralistic Theology." In *AI and IA: Utopia or Extinction?*, edited by Ted Peters, 117–32. Adelaide: ATF.

Searle, John. 1980. "Minds, Brains and Programs." *Behavioral and Brain Sciences* 3 (3): 417– 57. https://doi.org/10.1017/S0140525X00005756.

Shanahan, Murray. 2016. "Conscious Exotica: From Algorithms to Aliens, Could Humans Ever Understand Minds That Are Radically Unlike Our Own?" *Aeon*, October 19, 2016. https://aeon.co/essays/beyond-humans-what-other-kinds-of-minds-might-be-out-there.

Silver David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. "Mastering the Game of Go with Deep Neura Networks and Tree Search." *Nature* 529:484–89. https://doi.org/10.1038/nature16961.

Song, Yuhang, Beren Millidge, Tommaso Salvatori, Thomas Lukasiewicz, Zhenghua Xu, and Rafal Bogacz. 2024. "Inferring Neural Activity before Plasticity as a Foundation for Learning beyond Backpropagation." *Nature Neuroscience* 27:348–58. https://doi.org/10.1038/s41593-023-01514-1.

Tiku, Nitasha. 2022. "The Google Engineer Who Thinks the Company's AI Has Come to Life." *The Washington Post*, June 11, 2022. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* 59 (236): 433–60. https://doi.org/10.1093/mind/LIX.236.433

Tylor, Edward B. (1871) 1903. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Language, Art, and Custom*. London: John Murray.

van den Brink, Gijsbert. 2020. *Reformed Theology and Evolutionary Theory*. Grand Rapids, MI: William B. Eerdmans.

Van Eyghen, Hans. 2020. *Arguing from Cognitive Science of Religion: Is Religious Belief Debunked?* London: Bloomsbury Academic.

van Huyssteen, J. Wentzel. 2006. *Alone in the World? Human Uniqueness in Science and Theology*. Göttingen, Germany: Vandenhoeck & Ruprecht.

Véliz, Carissa. 2022. "LaMDA Is Nothing Like a Person. This is Why." *The Wire Science*. June 26, 2022. https://science.thewire.in/the-sciences/why-lamda-nothing-like-a-person/.

Watts, Fraser, and Marius Dorobantu. 2024. "Shamanic and Doctrinal: Dunbar and the Spiritual Turn in Contemporary Religion." *Religion, Brain & Behavior* 14 (1): 85–90. https://doi.org/10.1080/2153599X.2023.2168733.

Whitehouse, Harvey. 2004. *Modes of Religiosity: A Cognitive Theory of Religious Transmission*. Walnut Creek, CA: Rowman Altamira.

Wolfram, Stephen. 2023. "What Is ChatGPT Doing . . . and Why Does It Work?" *Stephen Wolfram Writings*. February 14, 2023.

Zygon®: Journal of RELIGION & SCIENCE

Arthur, Mathew. 2024. "Science and Technology Studies and Affect in the Wild." *Zygon: Journal of Religion and Science* 59 (3): 788–99. DOI: https://doi.org/10.16995/zygon.16758

Open Library of Humanities

Science and Technology Studies and Affect in the Wild

Mathew Arthur, PhD, Capilano University, North Vancouver, Canada, mathewarthur@capilanou.ca

Expanding on comments made at last year's Society for Social Studies of Science (4S) Meeting on a panel recognizing Donovan Schaefer's *Wild Experiment* as the winner of the Fleck Prize, this article explores the book's adjacencies in order to tell a "wild" story about place, Indigeneity, colonial sciences, messy methods in science and technology studies, and the implications of how affect is composed in words and worlds.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

Donovan Schaefer's *Wild Experiment* repeats the motto "there is no thinking that is not feeling" (2022, 5) across a dizzying array of histories and theoretical weaves: from scientific biography and science and technology studies (STS), theories of affect and affective neurosciences, to secularism or religious studies and scenes of conspiracy. The book is a page-turner. Something you might stay up all night to read. Its bare-bones point—spoiler alert—is that all knowing is feeling and can thus be contaminated by public moods. In the thick of this, science fine-tunes "truth" through mixed feelings; the pleasure of knowledge clicking into place is pitted against the fear of getting things wrong. Schaefer calls this "cogency theory." As a feminist STS theorist working in Indigenous studies and the editor of an international affect studies journal, I have big stakes in the constellation of disciplinary problematics *Wild Experiment* maps. Riffing on comments made at last year's 4S meeting on a panel recognizing Schaefer's book as the winner of the Fleck Prize, I want to tell a wild story that jumps off from the book's argument. I call it "wild" because it's messy, half-finished, and unfaithful to Schaefer's citational arc. But also because I am invested in unruly versions of STS and affect studies: a wide, almost "animist" sense of material semiotic affectibility.

Telling a wild story from the book's edges helps me to make connections to alternative empirics: the task of reimagining STS with non-Western terms and analytics (Law and Lin 2017) and of honoring "native science" (Cajete 2000), those ways of knowing that braid feeling, spirituality, natural history, and polity. These animating imbrications of local and global often risk domestication in STS's methodological symmetry. And while Schaefer (2022, 5, 9) asks cogency theory to explain "both truth and error" via the "felt weight of facts," we ought not forget all else that composes a scene of inquiry—including the multiple realities that clash in specificities of practice (Mol 1999).

I live in a coastal seaport city of towering glass condos, bubble tea shops, dog parks, and luxury retail. The rural surrounds are sites of intensive agricultural production, logging, mining, fishing, and fracking—but also home to fierce Indigenous resistance motivated by hereditary and diasporic ways of knowing that see lands and waters as kin. Wet'suwet'en put their bodies in the path of the Coastal GasLink Pipeline. Tŝilhqot'in take Taseko's New Prosperity open-pit mine to court. The Pacheedaht protect old-growth forestlands at Fairy Creek. In my urban neighbourhood, Chinese immigrants and Italian seniors grow prolific container gardens for food and medicine in alleyways sandwiched between boxy 1970s houses known as "Vancouver Specials." Stone Bhuddas and Virgin Mary statues adorn overgrown or graveled-over front lawns. Just down the hill, health food shops and community acupuncture clinics are being bulldozed to make room for the multi-million dollar homes of doctors and tech bros. Amidst these variegated lifeways and extractive frictions, science is tied

to economies of progress: social entrepreneurships, smart city logics, public health apps and algorithms, post-scarcity notions of sustainability.

Wild Experiment's intervention matters to how empiricism and secularism overlap in mainstream science as feelings of truth that determine how difference is governed—but even moreso to how "evidence" is too frequently quarantined from the intuitive or felt across sites of postcolonial scission to the lived detriment of Indigenous and immigrant knowledges. Gendered, racialized, and anthropocentric notions of reason are a well-rehearsed story: land and its nonhuman, more "native" inhabitants are irrational, moody, driven by tides, seasons, and mysterious cosmic cycles. Science saves the day by engineering repeatable conditions from this mercurial mix. How then do we account for small practices of resistance, those "pseudosciences" of survival that grow between the cracks? Defending lands and waters as living relatives. Saving and sharing non-GMO seeds for home remedies. Makeshift infrastructures of care that include weird syncretisms and messy spiritualities. As I iterate through this "wild" story of my own with *Wild Experiment* in tow, I ask what "different yield" (Hogan 1994) is made possible by putting science in the plural alongside Indigenous and other felt empiricisms like Traditional Chinese Medicine. On one hand, this move echoes Schaefer's sense of science as felt and "durable relationship[s] with the things around us" (2022, 11–12). On the other, durability is inevitably textured by practices at odds with the secularisms, atheisms, and "nonreligious improvisations" Schaefer (2022, 229) envisions as a means to "explore and retune our bodies" toward better sciences. Where I live, Indigenous territory bears the scars of knowledges made durable or killed off. Bodily landscapes are palimpsests of violence too. The ways in which "thought feels the prospect for concepts" (Manning 2008) amidst neurodiversity and within both small everyday practices and wider historical processes matters. I'm not sure Schaefer's cogency theory is the right conceptual toolkit for marking and living out these differences—or that locating truth as feelings in bodies or brains can ever get at the wilds of subjectivity where otherwise worlds insist and grow. In indeterminate scenes of living, so much more is at always work. Erin Manning (2020, 98) writes: "it is this muchness, this massiveness, that we need techniques for."

Techniques for Heterogeneity

Science is made in motley ecologies of practice, across vast materialities, with agendas and power-trips. Religion too. It takes a lot to hold a world together. The weaves of matter, mood, and meaning that make up our knowledge spaces are too big or barely perceptible, variously entrenched, and require much coordination of moving—sometimes slippery—parts. For both science and religion, change is at the mercy of histories already composed and in process. Both are guilty of methods that uproot situated stories and put them to work

reproducing more of the same (Haraway 1997; Law 2015). There are other ways to do science, of course: non-innocent encounters with the objects of technoscience, partiality and concern, slowness and listening, more expansive kinships (Despret 2016; Haraway 1997, 1988, 2016; Latour 2004; Stengers 2018). Anticolonial theological postures likewise help undo imperial Evangelicalisms, center Indigenous sciences, and foreground ecologies of belonging (Curtice 2020; Krawec 2022; Mendoza and Zachariah 2022; Woodley and Sanders 2020).

As Schaefer rightly notes, there are reasons to clamp down, to set boundaries or limits. Deepfakes and anti-vaxxers. Insurrection. Climate apocalypse. Evidence-based forms of democracy tenuous or on the brink. Across the political spectrum, feelings reel. How to better put secular science and religion (or conspiracy) together, both theoretically and in social practice, is perhaps never only a matter of foregrounding seemingly siloed but similarly productive internal logics (e.g. Latour 2013). Especially when theoretical physicists get spiritual or geneticists and surgeons go full God-complex. Collusions of science and secularism inform biotechnological, algorithmic, and other technoscientific chains of practice that are felt "religiously" at the tail-end: megachurches resisting vaccine mandates, genocidal drone strikes on Gaza, Christian nationalisms stoked by puberty blockers, AI-driven layoffs, or 5G.

While actor-network theory successor projects and other STS approaches have opened up facts to show their inner workings and disclosed the ontological politics at play in coordinations of reality across labs and fieldsites (where even cosmologies collide), they are not always attentive to "pleasures, pains, ecstasies, fears, ideals, dreams, or passions" (Mol 2014). Of the many "great divides" (Latour 1991) science studies has undermined, the rift between reason and emotion can sometimes slip under the radar. Schaefer (2022, 4) writes:

> Although science and technology studies (STS) has spent half a century showing that science isn't just what's in our heads—that knowledge production is always practical, social, and embodied—almost no work has been dedicated to exploring the relationship between knowledge-making and *feeling*. STS pioneer Bruno Latour, for instance, for all his sophisticated accounts of how science is made by coalitions of human bodies and nonhuman actants, still argues that science is *emotionally inert*. Even affect theory (the scene of some of the most interesting contemporary conversations about feeling) often seems to offer a funhouse mirror of common sense, recapitulating the assumption that feeling is separate from thinking.

Here, I have to make a gentle dig at Schaefer's opinionated join of STS and affect studies. Both have well-established techniques for heterogeneity that necessarily include feeling when appraising *and* making knowledge. Feminist STS cultivates a "feeling for the organism" at odds with cultures of extraction or dissection and

is more in tune with place-based Indigenous knottings of empiricism, spirituality, and care (Evans 2021; Keller 1983). Multispecies work is a training in capacities and feels: "caring for, being affected" (Haraway 2007, 36). STS case studies are packed with methods like "implosion" (Dumit 2014) and figuration (Haraway 1997) that are shot through with feeling as they map distributed and far-reaching material semiotic interdependencies across knowledge infrastructures of every kind. And, while not considered STS proper, archives of feeling and felt theories contend with racialized and gendered community knowledges that work against the sciences' moral hygienes, public pathologies, and more "modest" forms of witnessing (Cvetkovich 2003; Haraway 1997; Million 2009).

One way of reading Schaefer's complaint is to foreground how both STS and affect studies refuse the primacy of any given object in a scene of inquiry: symmetry. This is not about engineering epistemological parity or flattening ontology. In contemporary STS, symmetry is a way to approach noncoherence and complexity without a predetermined or singular analytical frame or set of methods (eg. Law 2004, 152). It requires paying attention to how differences are generated in weaves of relationship, across entangled agencies, with multiple meanings—all while reflexively tracking what our own (and others') methods include or make absent. After all, feelings show up in specific practices or as part of assemblages jam-packed with everything else: histories, bodies, concepts, technologies, media objects, dreams. It seems to me that Schaefer sees feeling as the not yet fully recognized end-all, be-all driver of knowledge production. But Latour's STS of emotionless chains of technology or affect studies that bracket feeling from thinking (or put them in assemblage) are not making ontological claims. Rather these are methodological moves and philosophical propositions with ethical dimensions: techniques for thinking and feeling differently in animate worlds.

"A Different Yield"

Places are wild assemblages to think from. I live in one of the most secular cities in Canada. Vancouver is gridded with Teslas at EV charging stations, yogawear retailers, and lush greenspaces. Downtown, you might see rich white couples wearing worn-out Patagonia pullovers, Mandarin-speaking luxury real estate moguls, Australian or Irish twenty-somethings on working holiday, and the rest of us hustling to pay rent. On the Eastside, the streets are lined with needles and tent encampments. Six lives are lost daily to fentanyl overdose. Each Valentine's day, hundreds gather here to march in memory of missing and murdered Indigenous women, girls, and two-spirit people. Chinatown's herbal apothecaries and seniors' homes are rapidly gentrifying into luxury apartments. Nearly a thousand global mining corporations are headquartered in the city. Headed to steelworks in the Asia-Pacific, coal moves by rail from southeastern mines to our outer-harbour superport, leaving toxic tailings and a long trail

of carbon emissions. In the canyons and benchlands beyond, a half-million acres of Indigenous territory are clearcut yearly for timber. More than 25,000 fracking wells and their tangled network of access roads and pipelines snake through our coastal, alpine, and boreal forests. All of this fuelled by the "great advantage of neutrality" (Policy Horizons Canada 2013, 6) that evidence-based joins of science and democracy are said to provide.

Yet, these lands have been continuously tended by Indigenous peoples for ten-plus millennia through a range of tribal processes marked by ceremonial relationships with an "interacting continuum" of landforms, bodies of water, flora, fauna, meteorologic processes, and spirits (Gisday Wa and Delgamuukw 1989, 7). Meanwhile, Canada's federal, provincial, and civic governments have made reconciliation the goal of state-Indigenous relations under the ambit of a multiculturalism in which mainstream sciences get the last word on nature (e.g. Latour 2002). This of course includes land and how it is instrumentalized or conserved. It includes the immunization or sacrifice of bodies, human and otherwise. So long as the witchy alternatives—spirits, talking animals, the medicine bundle—are barred ontological status they can remain in the mix. Again, a story about how difference is governed.

One of *Wild Experiment*'s overarching themes is conspiracy theory. In Canada's Pacific Northwest we have our fair share. Mid-pandemic, a convoy of racist anti-vaxxers and QAnon supporters mistaking Evangelical agendas and Twitter aphorisms as hard science filled the streets, eventually converging on the capital. When Schaefer (2022, 9) writes, "our spectrum of confidence and conviction is always constituted by feeling." It is hard to disagree. This is never more obvious than when we march in the wake of George Floyd's murder, occupy campus lawns for Palestinian liberation, watch hate-spewing Trump rallies on Fox, or hear "trad wives" rant about Drag Queen Story Hour on TikTok. These are not mere cultural differences. Lands and bodies are on the line. Admitting that "thinking feels" (2022, 6) may well safeguard necessary forms of consensus from post-truth with more sober emotions (while making a good case for therapy). And to this end, Schaefer argues that "it's dangerous to cut off contact with science" (2022, 7). In a world already conscripted by technoscience with mortal stakes—life-saving drugs, agricultures of scale, labor automations, climate mitigations—this might hold true. But when science and the state are ontologically aligned, how are we to index or account for subversive forms of racialized reason embedded in "evidence-based" governance? How can we adjudicate feelings of truth in conspiracy versus, say, Indigenous lifeways that answer the question "who gets to be a person" (Wilkinson 2017) with rivers, rocks, and trees or Chinese medical syncretisms that cure depression with needles and bitter herbs that move invisible Qi?

Like Western science, these knowledge practices have their own in-built agonisms or self-correcting checkpoints for appraising the "felt weight of

facts" (Schaefer 2022, 9). They are after balance as a matter of life and death. Much work has been done to bridge mainstream and non-Western sciences, from studies of Indigenous spirituality and mental health to clinical trials on the efficacy of herbs, moxibustion (burning mugwort over specific points on the body), or Tai Chi for Qi-deficiencies caused by COVID-19 (Fleming and Ledogar 2008; Ren et al 2021). No doubt Western science "edges ever closer to acknowledging the intangible, spiritual quality of matter and the intelligence of animals," as Vine Deloria Jr. writes (2001, 3). But is science's stamp of approval just another colonial gesture? Instead of imagining a day when science finally "verifies" animism, we might think with Isabelle Stengers (2012, 3):

> Only a "belief" can receive such a global name. If the adventurous specificity of scientific practices has been acknowledged, no one would dream of addressing others in terms of the "beliefs" they would entertain about a "reality" to which scientists enjoy privileged access. Instead of the hierarchical figure of a tree, with Science as its trunk, what we call progress would perhaps have had the allure of what Gilles Deleuze and Félix Guattari called a rhizome, connecting heterogeneous practices, concerns, and ways of giving meaning to the inhabitants of this earth, with none being privileged and any being liable to connect with any other.

Stretching the term to include "any organized system of knowledge," Schaefer (2022, 6, 11) is likewise attentive to how science "rubs elbows with other everyday modes." Even still, as he catalogues conspiracy feelings or more even-keeled scientific passions, Schaefer reads affectibility through neuroscience and experimental psychology in order to prove that cognition includes emotion. But as I see it, recognizing that "we are not alone in the world" requires a more radical (even radicle) redistribution of the capacity to feel, to act, and to prove: not merely from brain to world, from individual to "luring assemblage" (Stengers 2012, 6), but until "to" and "from" lose the too-ready presumption of their directional vectors.

In a devastatingly beautiful reflection on science, feeling, harm, and repair, Chickasaw poet Linda Hogan tells stories about corn. First, as a child, listening to the rustle of stalks swaying in her uncle's field. Later, of geneticist Barbara McClintock who knew her plants "in the way a healer would know them." Or, corn as grandmother: a kiva in Chaco Canyon where ancient maize replants itself yearly, defying time. While Hogan (1994, 72, 79) observes an impetus in contemporary sciences for "intuitive processes of discovery," here she is after a "language of that different yield." This is not an extractive harvest. Rather, it responds to the voices of "wind, dove, corn, stone" and the pull of "earth and life" (1994, 77, 80). She writes: "The stalks of the corn want clean water, sun that is in its full clean shining. The leaves of the corn want good earth. The

earth wants peace. The birds who eat the corn do not want poison. Nothing wants to suffer. The wind does not want to carry the stories of death" (1994, 80). Here, everything desires, affects: a powerful material semiotic yield.

In the apposite context of New Atheism and its violent dehumanization of Muslims, Schaefer engages animacy too—those registers of speech and ways of feeling caught up in multispecies hierarchies of agency and value. Thinking with Mel Chen, he writes:

> As an example, Chen offers the phrase "the hikers that rocks crush" (2012, 3) noting that it registers as viscerally wrong to English speakers because it violates our implicit preference to organize sentences around what we perceive to be *subjects* rather than *objects*. This reveals the way a *felt* register of who and what matters (and who gets to count as a *who*) is woven into the seams of language. (Schaefer 2022, 72)

But how might *Wild Experiment* come to bear on the problem of animism in its most Tylorian sense as "an idea of pervading life and will in nature far outside modern limits . . . animating even what we call inanimate bodies" (Tylor 2010 [1871], 260)? From the very moment culture was cleaved from nature in anthropology's colonial emergence, the sciences have been hell-bent on pathologizing those who talk to trees and rocks—what Leroy Little Bear (2011) calls "native science." Put another way, what does the claim "thinking is feeling" yield in the company of other sciences that already comprise millennia of attention to more-than-human feels, to bodily rhythms and pulses, to durational and patterning relationships with more or less subtle energetic ecologies: folk or anecdotal knowledges, Indigenous "animisms," Traditional Chinese Medicine, or Ayurveda? Science made plural need not fully quarantine itself from conspiracy—even as a matter of persuasion or public feeling. Rather there are many specificities of practice with better and worse effects to attend to.

Maybe I am asking a nitpicking question about how Schaefer collapses feeling and emotion as synonyms of affect in ways that stress the interiority of knowledge-making:

> My interpretation of affect theory is that affect is essentially *power*, understood not as an external, oppressive force but, following Michel Foucault, as fundamentally productive. Power is *what makes bodies move* (or binds them). *Affect* is a word for processes—beneath, beside, and within cognition—that register in awareness as feelings, emotions, and moods. At heart, power is affect, affect is power. Everything we do emerges out of an agonism of feelings . . . from the cyclone of small, felt pulses splashing across us all the time—a tug of longing, a pinprick of annoyance, a pang of grief—to thoughts, actions, decisions, moods, words. (2022, 19)

This feels somehow too restrictively personal—biologically coded or psychologized—for all that is activated or risked in scenes of knowing. How does power traffic—beyond body, beyond person—between, say, emotion and atmospheres of public feeling or the world at large? Or between pasts made durable and emancipatory futures? Schaefer (2022, 3, 19) answers with a "fully fleshed-out continuum" from feeling to practice across which the "micro-level delight" of knowledge falling into place is continuous with institutional or "macro forms." Yet, this too-quickly skims over the "heterogeneity between micro and macro" (Deleuze 1997). Rather than simply scalar or directional, power courses in the messy alongside of life already happening: force-guiding assemblages of matter, meaning, mood, and possibility. Here Gregory Seigworth's reading of Deleuze and Gauttari better articulates my position: "nothing much is advanced by finding everywhere the effects of power; something more is at stake when the task is, rather, to understand the . . . immanent assemblages that make the effects of power our actuality" (Seigworth 2005, 168). Do not forget: STS and affect already have techniques for heterogeneity.

To yield differently is a matter of wide affectivity. It might include small pleasures of knowledge coalescing, but is it endlessly more contingent than personal. What localizes as a pain point is the world impinging: things accreting, accompanying, shaping, receding. Native science knows this well. As Leroy Little Bear says:

> In a state of flux, in a state of constant motion, things never remain the same. Things are forever changing. If we stopped and thought about it, we live in a very narrow spectrum of ideal conditions. So in the native world, we try to renew those conditions that are ideal for our existence. The drumming, the singing, the dancing is part of the renewal. (Hill 2008)

By describing knowledge-making as an "ongoing process" and a "contest of forces," Schaefer (2022, 9) approximates the stakes of native sciences. He is never too far-off from the capaciousness of affect I am advocating (see Seigworth 2017). But my plea to yield differently stretches citationality well past genealogies of affect (and how they mesh or mess with neurological or ethological sciences). It is first a question of the larger set of affectivities enrolled in appeals to science's "calm passions" (Schaefer 2022, 39). This would include all that it takes to make knowledge, good and bad: the histories, landscapes, bodies, moods, machines, ideas, institutions, and transitory systems of power situated in relation to each (e.g. Deleuze 1997). Second, but no less important, it concerns the kinds of figures that populate our writing—the brain, the scientist, the animal—and what allegiances they bind. More than any feeling of "truth," these material semiotics condition what can be felt, noticed, acted on, renewed.

So, the question remains of how to run with the claim that "thinking feels" while accounting for science's bad side effects: lab rats, addictive pills, shitty bedside manner, racist algorithms, nuclear weapons, Elon Musk's plans to colonize Mars. My story tells of homelands carved out with mine shafts and fracking fissures, poisoned watersheds and clearcuts. But also of ceremony and multispecies care. Putting science in the plural surfaces noninnocent specificities of practice. It poses the question of how knowing as feeling is inflected by bodily, cultural, technical, even cosmological infrastructures. I am on board with Schaefer's project to "animalize cognition, language, and rationality" (2022, 20). One hundred percent. I see a reparative promise in the politics that might proceed from it. But, even more, *Wild Experiment* should be read in the company of feminist material semiotics and other-than-Western sciences that account for how subjectivities and animacies are forged across wider, wilder affectivities: those "technical, textual, organic, historical, formal, mythic, economic, and political" assemblages that "bend our attention, warp our certainties, and sustain our lives" (Haraway 1997, 68).

To hitch power, pleasure, and knowledge to subjectivity is not enough: the world is so much. In a set of notes addressed to his friend Foucault, Deleuze details where his philosophy of immanence veers off from those power-laden passages of *La volonté de savoir* so crucial to Schaefer's cogency theory. Instead of pleasure, Deleuze (1997) speaks of desire:

> it is but one with an assemblage of heterogenous elements which function; it is process, in contrast with structure or genesis; it is affect, as opposed to feeling; it is "haecceity" (individuality of a day, a season, a life), as opposed to subjectivity; it is event, as opposed to thing or person. And above all it implies the constitution of a field of immanence . . . which is only defined by zones of intensity, thresholds, gradients, flux.

A different yield indeed.

References

Cajete, Gregory. 2000. *Native Science: Natural Laws of Interdependence*. Santa Fe, NM: Clear Light Publishers.

Chen, Mel Y. 2012. *Animacies: Biopolitics, Racial Mattering, and Queer Affect*. Durham, NC: Duke University Press.

Curtice, Kaitlin B. 2020. *Native: Identity, Belonging, and Rediscovering God*. Grand Rapids, MI: Brazos Press.

Cvetkovich, Ann. 2003. *An Archive of Feelings: Trauma, Sexuality, and Lesbian Public Cultures*. Durham, NC: Duke University Press.

Deleuze, Gilles. 1997. "Desire & Pleasure." Translated by Melissa McMahon. *GLOBE E Journal of Contemporary Art* 5.

Deloria, Vine. 2001. "American Indian Metaphysics." In *Power and Place: Indian Education in America*, edited by Daniel R. Wildcat and Vine Deloria, 1–6. Golden, CO: Fulcrum Publishing.

Despret, Vinciane. 2016. *What Would Animals Say If We Asked the Right Questions?* Translated by Brett Buchanan. Minneapolis, MN: University of Minnesota Press.

Dumit, Joseph. 2014. "Writing the Implosion: Teaching the World One Thing at a Time." *Cultural Anthropology* 29 (2): 344–62.

Evans, Claire L. 2021. "A Feeling for the Organism: What It Means to Actually Work with Biology." *Grow.*

Fleming, John, and Robert J Ledogar. 2008. "Resilience, an Evolving Concept: A Review of Literature Relevant to Aboriginal Research." *Pimatisiwin* 6 (2): 7–23.

Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99.

———. 1997. *Modest_Witness@Second_Millennium.FemaleMan_Meets_OncoMouse: Feminism and Technoscience*. New York: Routledge.

———. 2007. *When Species Meet*. Minneapolis: University of Minnesota Press.

———. 2016. *Staying with the Trouble: Making Kin in the Chthulucene*. Durham, NC: Duke University Press.

Hill, Don. 2008. "Listening to Stones: Learning in Leroy Little Bear's laboratory." *Alberta Views: The Magazine for Engaged Citizens*.

Hogan, Linda. 1994. "A Different Yield." *Religion & Literature* 26 (1): 71–80.

Keller, Evelyn Fox. 1983. *A Feeling for the Organism: The Life and Work of Barbara McClintock*. New York: W. H. Freeman.

Krawec, Patty. 2022. *Becoming Kin an Indigenous Call to Unforgetting the Past and Reimagining Our Future*. Minneapolis: Broadleaf Books.

Latour, Bruno. 1991. *We Have Never Been Modern*. Translated by Catherine Porter. Cambridge, MA: Harvard University Press.

———. 2002. *War of the Worlds: What about Peace?* Edited by John Tresch. Translated by Charlotte Bigg. Chicago: Prickly Paradigm Press.

———. 2004. "Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern." *Critical Inquiry* 30 (2): 225–48.

———. 2013. *An Inquiry into Modes of Existence: An Anthropology of the Moderns*. Translated by Catherine Porter. Cambridge, MA: Harvard University Press.

Law, John. 2004. *After Method: Mess in Social Science Research*. London: Routledge.

———. 2015. "What's Wrong with a One-World World?" *Distinktion: Journal of Social Theory* 16 (1): 126–39.

Law, John, and Wen-yuan Lin. 2017. "Provincializing STS: Postcoloniality, Symmetry, and Method." *East Asian Science, Technology and Society: An International Journal* 11 (2): 211–27.

Little Bear, Leroy. 2011. "Native Science and Western Science: Possibilities for a Powerful Collaboration," The Simon Ortiz and Labriola Centre Lecture on Indigenous Land, Culture, and Community, Arizona State University.

Manning, Erin. 2008. "Creative Propositions for Thought in Motion." *Inflexions* 1 (1): 1–24.

———. 2020. *For a Pragmatics of the Useless*. Durham, NC: Duke University Press.

Mendoza, S. Lily, and George Zachariah, eds. 2022. *Decolonizing Ecotheology: Indigenous and Subaltern Challenges*. Eugene, OR: Pickwick Publications.

Million, Dian. "Felt Theory: An Indigenous Feminist Approach to Affect and History." *Wicazo Sa Review* 24 (2): 53–76.

Mol, Annemarie. 1999. "Ontological Politics: A Word and Some Questions." *The Sociological Review* 47 (1): 74–89.

Policy Horizons Canada. 2013. "The Case for Evidence-Based Policy." Ottawa, Canada: Policy Research Initiative.

Ren, Wei, Pan Liang, Yue Ma, Qin Sun, Qingrong Pu, Li Dong, Gang Luo, Maryam Mazhar, Jiali Liu, Raoqiong Wang, and Sijin Yanga. 2021. "Research Progress of Traditional Chinese Medicine Against COVID-19." *Biomedicine & Pharmacotherapy* 137:111310.

Schaefer, Donovan O. 2022. *Wild Experiment: Feeling Science and Secularism after Darwin*. Durham, NC: Duke University Press.

Seigworth, Gregory J. 2005. "From Affection to Soul." In *Gilles Deleuze: Key Concepts*, edited by Charles J. Stivale, 159–69. Ithaca, NY: Routledge.

———. 2017. "Capaciousness." *Capacious: Journal for Emerging Affect Inquiry* 1 (1): i–v.

Stengers, Isabelle. 2012. "Reclaiming Animism." *E-flux* 36:1–9.

———. 2018. *Another Science Is Possible: A Manifesto for Slow Science*. Translated by Stephen Muecke. Cambridge: Polity Press.

Tylor, Edward Burnett. [1871] 2010. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*. Vol. 1. Cambridge: Cambridge University Press.

Wa, Gisday, and Delgam Uukw. 1989. *The Spirit in the Land: Statements of the Gitksan and Wet'suwet'en Hereditary Chiefs in the Supreme Court of British Columbia*. Gabriola, BC: Reflections.

Wilkinson, Darryl. 2017. "Is There Such a Thing as Animism?" *Journal of the American Academy of Religion* 85 (2): 289–311.

Woodley, Randy S., and Bo C. Sanders. 2020. *Decolonizing Evangelicalism: An 11:59 P.M. Conversation*. Eugene, OR: Cascade Books.

Zygon®: Journal of RELIGION & SCIENCE

Pollock, Anne. 2024. "Cultivating Calm Passions." *Zygon: Journal of Religion and Science* 59 (3): 800–10. DOI: https://doi.org/10.16995/zygon.16691

Open Library of Humanities

Cultivating Calm Passions

Anne Pollock, Professor of Global Health and Social Medicine, King's College London, London, UK, anne.pollock@kcl.ac.uk

Building on comments delivered at the Society for Social Studies of Science (4S) Meeting in Honolulu recognizing *Wild Experiment* as the winner of the 2023 Ludwick Fleck Prize, my reflections here are anchored in the appreciation of the book's illumination of the value of the cultivation of calm passions. I open by reflecting on book's fresh engagement with core science and technology studies concerns about how durable knowledge is made. I then suggest a few developments that might follow on from that appreciation. First, I reflect on how the role of the erotic might be pushed further analytically. Second, I explore how the book's persuasive arguments about racialized reason and the limitations of the paranoid style of critique might be treated more symmetrically, as elements to be aware of in our own work as well as that of big bad racists with whom we disagree. Finally, I highlight the potentially broader-audience value of grounding ourselves in calm passions in the era of the internet pile-on.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

It is such a privilege to join this Book Symposium reflecting on a book I absolutely adore.

I come to this book as a reader from science and technology studies (STS), in a notably literal sense. Donovan O. Schaefer's *Wild Experiment: Feeling Science and Secularism After Darwin* was among the dozens of books I read in the summer of 2023 in the context of selecting the winner of the Ludwik Fleck Prize as part of a book prize committee for the Society for Social Studies of Science (4S). As someone who has not previously engaged with religious studies in any substantive way, I did not expect to be captivated by this book. And yet it drew me in, held my attention, and has continued to resonate with me since. Thus, I welcome this chance to return to it for a more extended reflection.

Part of what made the book *accessible* to me—in addition to the lucid writing—was its robust engagement with canonical STS literature as well as with the feminist and antiracist scholarship I do know well. He brings together bold-faced names for STS such as Thomas Kuhn and Bruno Latour, as well as incisive antiracist thinkers ranging from Audre Lorde to Eduardo Bonilla-Silva. And yet a key thing that made it *interesting* to me was that it brought those too-often-siloed literatures together with each other and with new-to-me literatures in a way that is original.

Because of these unusually inclusive citational practices, I suspect most readers will find a mix of familiar and new points of reference. I was struck that perhaps in part because of the capaciousness of its reading, *Wild Experiment* does not feel as cliquish as many theory-oriented books do. By bringing together multiple conversations oriented around common concern but that are usually happening in parallel, I was impressed that this book is quite distinct from the books by the tight networks that often get recognized both in STS and in intersectional feminist scholarship—*Wild Experiment* thus offers scholars in both of those domains (among others) with a new voice, which is so refreshing.

Building on comments I initially delivered at the 4S Meeting in Honolulu recognizing *Wild Experiment* as the winner of the 2023 Fleck Prize, my reflections here are anchored in the appreciation of the book's illumination of the value of the cultivation of *calm passions*. I open by reflecting on book's fresh engagement with core STS concerns about how durable knowledge is made. I then suggest a few developments that might follow on from that appreciation. First, I reflect on how the role of the erotic might be pushed further analytically. Second, I explore how the book's persuasive arguments about racialized reason and the limitations of the paranoid style of critique might be treated more symmetrically, as elements to be aware of in our own work as well as that of big bad racists with whom we disagree. Finally, I highlight the potentially broader-audience value of grounding ourselves in calm passions in the era of the internet pile-on.

A Fresh Engagement with the Core STS Concern of How Durable Knowledge Is Made

Let me first delve in a bit into how *Wild Experiment* engages the core STS concern of how durable knowledge gets made. The central argument of the book is the inextricability of thinking and feeling, and this is both a very classic STS concern at the core of knowledge-making and one that is absolutely essential to our age. *Wild Experiment*'s engagement with the topic builds on and is relevant to longstanding STS concerns, for example insisting on symmetry in how we locate the cause of true and false beliefs, and at the same time is innovative.

A core concept *Wild Experiment* contributes is "cogency." I quote at some length from the book's first elaboration of the term in order to give those who have not yet read it a bit of a feel for the book's distinctive voice and style:

> To say an argument is *cogent* doesn't mean, exactly, that it's *true*. It means *it appeals,* or *it's compelling*. It means it *feels* true. It has a *pull*—a weight. *Cogency* takes knowledge-making out of a binary frame, in which sovereign reason sizes up a situation, strokes its chin, and then judiciously flicks the switch to YES or NO. It suggests, instead, knowledge-making as an ongoing *process*—a contest of forces—and specifically as a constant measuring and remeasuring of the *felt* weight of facts. Cogency lights up the way our spectrum of confidence and conviction is always constituted by feeling. New information that tips the balance *adds weight*. Changing our minds means changing how we feel. (Schaefer 2022, 9)

This is a theory of knowledge and feeling Schaefer contrasts with "cogitation" as a theory, which despite the similar sounds of the words turns out to have a different root—whereas "cogitation" is a la "cogito ergo sum," "cogent" is related to agonism and pedagogy and the driving of forces together. Schaefer notes the missed opportunity of the historical accident that Thomas Kuhn became the "anointed gadfly of science" (Schaefer 2022, 17) to the neglect of the more affectively-engaged Michael Polanyi. He persuasively argues that we have much to gain from engaging more with Polanyi's concepts such as "intellectual passions" (Schaefer 2022, 16). Schaefer reprises these, describing "interest" for Polanyi as "a felt sense of what matters," "intuition" as "a felt sense of what *might* be useful in the future," "persuasive passion" as the desire to convince others, and especially the "love of elegance," which Polanyi calls *the feeling of making contact with reality*, which produces a click, the subtle joy of pieces of information snapping together (Schaefer 2022, 16).

Now if the click were the only thing, then knowledge-making practices such as science might not be different from conspiracy theory mongering. What distinguishes science in particular and durable knowledge-making in general is that the desire for that elegance is counterbalanced with an awareness of the

risk of seeing things the way we want to, and the constant subjection of what we believe to a countervailing scrutiny.

Here is how Schaefer describes how the reliability of science endures even after we recognised that science is not drained of emotion:

> Naomi Oreskes writes in *Why Trust Science?* that the reliability of science is based on both "its sustained engagement with the world" and "its social character." Cogency theory adds a third dimension to this picture. It suggests that science works when it stages an agonism of different cognitive emotions, tracking the excitement of click while holding on to nervousness about getting things wrong . . . The same emotional machinery that sends us down the wrong track can also, with just a bit of fine-tuning, build durable science. (Schaefer 2022, 231)

This particular word choice in the phrase "durable science" stood out to me because I have long been interested in how *durability* can offer a way of characterizing the project of knowledge-making—archetypally scientific knowledge-making, and also more broadly. My first book foregrounded the thematic in the title: *Medicating Race: Heart Disease and Durable Preoccupations with Difference* (Pollock 2012). Durability is not a concept used many times in Schaefer's book, but it is used in illuminating ways. Schaefer uses the term "durable" as he elaborates "cogency theory":

> Good knowledge isn't knowledge that has been drained of feeling. It's knowledge that reflects a working, **durable** relationship with the things around us, a relationship resolutely defined by emotion. It's the product of complicated operations of feeling—an agonism, an invisible clash of forces holding our desire to know the world in tension and trying to outmaneuver other felt priorities that might muddy the waters. This sense of science isn't a forensic tool for calculating absolute truth. Instead, it's a cultivated contraption for feeling our way toward truth (objectivity in shirtsleeves) without the anticipation of certainty. Collapsing the binary of reason and emotion doesn't leave us adrift. It shines a light on the carefully staged agonism of intellectual feelings that is the real driver of good knowledge. (Schaefer 2022, 11–12)

What does *durable* mean in these quotes about the making of good knowledge? Indulge me in an etymological foray. Durability is a common enough word in contemporary vernacular English, meaning "the ability to withstand wear, pressure, or damage."[1] The etymology of the English word *durability* has its roots in the late fourteenth century: "power of lasting or continuing in the same state, resistance to decay or dissolution," drawn from Late Latin durabilitatem (via Old French durabilité) "lasting, permanent," in turn drawn from the

Latin durare "to harden."[2] In common with the word *duration*, the etymology of *durability* evokes a relationship between robust materiality and extended temporality.

Even in its day-to-day usages, the term—in both its noun form and in its adjective form 'durable'—often has a slightly technical valence, as with the term "durable goods" to refer to non-consumable non-perishable consumer goods such as cars and washing machines that are kept for a long time rather than deteriorating, rotting, or being used up, or "durable relationships" by which unmarried partners might qualify for say European settled status. In all these senses, what is durable has the material quality of not being easily broken, as well as a temporal quality of not being fleeting. In these senses, durability is the opposite of fragility or precarity.

Yet durability does not mean imperviousness or invulnerability. *Durus* is hardness in Latin, and yet *hardness* is often at odds with *durability*. That is certainly the case in many engineering contexts, in which hardness can be associated with brittleness and durability is precisely produced by the flexibility that allows the absorption of shock. This quality of openness to modification in the context of new conditions it is true for durable knowledge, and for durable relationships.

The Role of the Erotic

The connection between intellectual passions and durable knowledge spurs me to reflect a bit on a theme at the intersection of core STS and feminist theory that I do think might have been further developed within the book, which is the relationship between knowledge-making and the concept of *the erotic*.

On the book's opening page, I was delighted to see reference to the classic feminist science studies of Evelyn Fox Keller's *A Feeling for the Organism*. Fox Keller happens to be someone from whom I learned about feminist science studies during my own PhD. She died in 2023; I miss her.

Wild Experiment is broadly deeply resonant with Fox Keller's sensibility, but it does not follow her to her articulation of the erotic. Schaefer turns instead to Black feminism for touch points on that theme: Audre Lorde, for whom the erotic names a source of both pleasure and power feminists should embrace, and especially Sharon Patricia Holland, who in contrast with Lorde foregrounds a harmful aspect of *the erotic* to name the racist desire for connection with the denigrated racial other—something return to later—builds on something important about these feminist engagements with the erotic. Yet we might explore other aspects of the erotic, which in feminist articulations including Fox Keller's is not a simple synonym for pleasure, and is precisely not about a covetous desire for access to the other. The erotic can also distinctively name authentic curiosity in, and deep appreciation for, another in all their depths, and a consensual communion.

Schaefer doesn't really engage the potential of the erotic per se in science, and yet it could be fruitful for his argument. For Evelyn Fox Keller, the invocation of the erotic is an alternative to the pornographic and indeed rape-metaphor-informed ways of knowing the universe articulated by key founders of modern science (and plenty of its contemporary enthusiasts), that posit the universe as a tease and the scientist as its ultimately successful vanquisher. For Fox Keller, the erotic is figured as a longing for a more relational and reciprocal way of knowing the universe than masculine science is structured to pursue – not "having one's way with" the natural world, but communing in a profound way that might blur the boundaries of the knowing subject and leave us transformed by the encounter. For Fox Keller, this erotic quality is central to the "dynamic objectivity" she posits as a more useful alternative to "static objectivity" that would be achieved through total separation of the scientist from the object of knowledge and then mastery thereof. In her classic collection of essays *Reflections on Gender and Science*, Fox Keller writes:

> While some scientists see their endeavor in predominantly adversarial terms, as contests, battles, exercises in domination, others see it primarily as an erotic activity. Michael Polanyi, for example, emphasizes, instead of distance, the need to extend our body to include [the object]—so that we come to dwell in it. (Fox Keller 1985, 125)

Fox Keller goes on to quote contemporary scientists rejecting the metaphor of "putting nature on the rack and torturing answers out of her," which is "like rape," and instead that good science "is like the difference between rape and making love" (Fox Keller 1985, 125). In the history of science, according to Fox Keller, erotic themes "have been submerged by a rhetoric and ideology of aggression," and she suggests there is value in attending to the work of scientists who "have seen their relation to their material in erotic rather than adversarial terms" (Fox Keller 1985, 126).

I wonder how this feminist STS thread of the erotic might offer a way to deepen the engagement across philosophy of science and Black feminism? Although I love many things about Sharon Patricia Holland's book *The Erotic Life of Racism* (2012), I think she unhelpfully flattens something that can be analytically useful about the potential of the erotic insofar as she equates "the erotic" with "desire." For Fox Keller, the erotic can't be any kind of unilaterally demanded intimacy, with an objectified racial other or otherwise – the erotic is by definition something that happens in a mutually engaged relationship in which responsiveness to and by the other is fundamental.

This might be why in the context of science as a vocation, the pursuit of durable knowledge about the world can feel like the kind of erotic allure that

is not merely about pursuing gratification for oneself (that Holland rightly warns us to be suspicious of) but rather a form of the erotic Fox Keller has in mind. Following Fox Keller, the erotic might be an element of fleeting but real connection with another, or might be part of a durable relationship, but I think there is something worthwhile in attending to the erotic as a form of desire that is precisely *not* covetous or rapey, but instead is anchored in authentic respect and is reciprocal relationality.

Confronting Our Own Durable Preoccupations with Difference, and Our Own Paranoia

Building from this, I would offer that thinking more about the relationship between the erotic and science might also help to deepen the link between this theme of *Wild Experiment* and its fascinating exploration of racialized reason. Schaefer brilliantly illuminates the pervasiveness of racialized reason in the work of atheist dogmatists and conspiracy theorists.

As Schaefer argues "Feeling makes science work, but it also leads to the *collapse* of good knowledge—the giddy downward spiral of conspiracism or the intoxicated self-confirmations of racialized reason" (Schaefer 2022, 4). He later elaborates that second term: "*Racialized reason* is the skin of thought that forms around racist feelings. It's another way of naming what critical race scholars have been saying for decades—namely, that racism isn't just a set of explicit *beliefs*, but a whole topography of ways of thinking" (Schaefer 2022, 24).

Schaefer's diagnosis of racialized reason in conspiracy theories and atheist provocations is persuasive, and important. And yet I wonder whether conflating racialized reason with racism, and imagining it to be the sole preserve of racists, misses an opportunity to be more symmetrical and reflexive in how we engage with race—in science, and in STS and other domains of humanities and social science.

Yes, as Schaefer persuasively argues building on Holland, the contemplation of the figure of the racial other has a pleasurable affective charge for the racist. But how can we understand how desire is also at stake in the contemplation of race for other actors of potential interest—for example, contemporary scientists of race who see themselves as antiracist, and antiracist scholars in humanities and social science?

One of the intriguing elements of the *Wild Experiment* is the critical way it analyzes the debunking mode used by a certain type of atheist ideologue. Drawing and building on Eve Kosofsky Sedgwick's classic 1997 intervention in favor of the value of "reparative reading" in addition to "paranoid reading," Schaefer shows how a "paranoid style of critique" is pervasive among the overlapping groups of conspiracy theorists and Islamophobic advocates of atheism (the core of Schaefer's engagement is on pp. 95–97, in the context of conspiracy theories,

and it is applied to New Atheism around p. 205). The invulnerable stance of the unmasker in the face of an exteriorized hostile world is common to both: "paranoid reading is about becoming the agent of unmasking, the director of surprises rather than their victim" (Schaefer 2022, 96).

I wonder whether there might be more scope for symmetry here, to consider how the paranoid impulse characterizes not just "them" but also "us." To make this more concrete, I wonder what it would look like to apply the same approach to race in science as Schaefer applies to atheism.

Reading Schaefer in the context of my longstanding interest in the durable preoccupations with racial difference in science and medicine, I found myself reflecting that the most common form of social science and humanities critique of race in science and medicine takes a paranoid approach, in which the goal is to unveil insidious racial essentialism lurking underneath any given claim in science or medicine and righteously stamping it out. The intellectual and affective quality of that scholarship seems resonant with what Schaefer (2022, 97) describes as "the adoption of the paranoid stance rapidly hardens into a mandate, an emotional orthodoxy demanding exposure of the secret sins of every scrap of culture." I have long been skeptical of the value of this mode of engagement with scientific claims about race, which often correctly observes that claims about say genetic essentialism in race science can have a hydra-headed quality, but doesn't follow the logic of that observation to realize that forever chopping off and holding up the severed head of genetic essentialism does not diminish the power of racism in science or in society.

Schaefer does not mention but I am struck that Kosofsky Sedgwick anchors that essay in the discussion of a conversation with her friend Cindy Patton about a debate about a scientific claim: whether or not HIV was developed in a government lab and deliberately set upon gay people and other communities deemed undesirable by hegemonic power. For me, the allure of that rabbit hole is resonant with the limitations of certain type of critique of race in science. The volatile power of race in a racist society does not rise or fall on the truth value of any given scientific claim. In the same way as we should heed Patton's warning that getting too exercised about proving or debunking the government lab theory for HIV might distract from more urgent questions of care for those suffering and dying of AIDS, honing our antiracist analytical focus too firmly on the debunking of particular scientific claims might be a waste of our energies.

Of course we cannot forgo paranoid reading of claims about race in science. Scientific claims about racial difference can indeed be sites of the articulation of harmful racist ideologies. And yet we should also embrace work that combines paranoid reading with reparative reading—something we might see operating for example, in Alondra Nelson's *The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome* (2016), which develops for a

broader audience ideas she had been working through in STS venues such as her fabulously rich article "Bio Science: Genetic Genealogy Testing and the Pursuit of African Ancestry" in *Social Studies of Science* (2008). In contrast with many scholars writing about contemporary engagements with race and genetics who take a debunking mode, Nelson alerts us to *both* problematic essentialisms *and* reparative aspirations in genetic research including that done by and for African Americans.

This is not a criticism of Schaefer since the science of race is not his topic, but it is a reminder to be alert to paranoid tendencies in our own engagements. We all live in a racist society, and we should not imagine ourselves to be completely immune to any racialization within our own reason. There are big bad racists out there—I am convinced by Schaefer's analysis that the New Atheists such as Christopher Hitchens should be denounced as such—but there are also inextricably cognitive and affective aspects to race for all of us.

The Value of Calm Passion in the Context of the Internet Pile-On

The quieter, slower, open-ended qualities of the erotic for Fox Keller and reparative reading as a complement to paranoid reading returns me to Schaefer's discussion on passion. Drawing on Hume, Schaefer contrasts "calm passions," which are aligned with science, and "violent passions," which are not. Turning to Schaefer:

> "What we call strength of mind," he [Hume] proposed, "implies the prevalence of the calm passions above the violent." The calm passions—passions that fixate on distant rewards or pleasures rather than immediate ones—are arrayed in such a way as to discipline and outmaneuver the violent passions. Even tranquillity is not the absence of feeling, but a specific emotional posture that allows us to patiently explore a problem, listening attentively for the soft click of real discovery. "By *reason* we mean affections," Hume concluded, "but such as operate more calmly, and cause no disorder in the temper: which tranquillity leads us into a mistake concerning them, and causes us to regard them as conclusions only of our intellectual faculties." The calm passions compose themselves into an alloy that makes scientific knowledge production possible. (Schaefer 2022, 39)

This is an element in which *Wild Experiment* might be brought to bear on broader conversations. The insights into topics of wide and timely interest such as the allure of conspiracy theories could almost have made this into a trade book. I wonder whether Schaefer has considered broader-audience writing, and also how he sees this book in relation to current wider-circulation books and broader conversations.

For example, Naomi Klein's *Doppelganger: A Trip Into the Mirror World* (2023) tracks the divergent path taken by someone with whom she has surprisingly often been mixed up: Naomi Wolf, a one-time high-profile liberal feminist who has become an anti-vaxing conspiracy theorist and darling of far-right radio. Trying to understand the appeal of that weird world for so many, Klein recounts one of Wolf's bizarre claims: according to Wolf, the Covid vaccines have dehumanized the people of New York City, because the nanoparticles have made them lose their smell and generally seem less human (Klein 2023, 112). However, while being clear that this is "gonzo stuff," Klein acknowledges that there is something that *feels true* about what Wolf says: "There is a lifelessness and anomie to modern cities, and it did deepen during the pandemic—there is a way in which many of us feel we are indeed becoming less alive, less present, lonelier. It's not the vaccine that has done this; it's the stress and the speed and the screens and the anxieties that are all by-products of capitalism in its necro-techno phase. But if one side is calling this fine and normal and the other side is calling it 'inhuman,' it should not be surprising that the latter holds some powerful allure." In a style that might be characterized as calmly passionate, Klein's project seems to me to be aligned with Schaefer's, in that both demonstrate that the rational debunking of the pseudoscientific claims does not quite meet the challenge of undoing the affective allure of the mirror world of conspiracies.

What is true of science also seems true of slow critical analysis more broadly. It puts me in the mind of another of my treasured interlocutors from my PhD days, Noam Chomsky. When trying to grapple with horrific events in Israel and Palestine of late, I find myself rereading Chomsky, and appreciating his calm passion. It stands in such contrast to the dominant mode of today, the hot-headed internet tirade. Even in the face of outrageous events, there is a real satisfaction in developing a more critical engagement that is slower and stronger. Schaefer's brilliant book *Wild Experiment* provides an exquisite model of what a calmly impassioned argument looks like, and this makes the book absolutely vital not only for STS and religious studies and the other scholarly fields on which it draws and to which it contributes, but also to broader conversations of our age.

Note

[1] *Oxford Languages*, "Durability," accessed November 17, 2024, https://www.google.com/search?q=durability.

[2] See https://www.etymonline.com/word/durability.

References

Holland, Sharon Patricia. 2012. *The Erotic Life of Racism*. Durham, NC: Duke University Press.

Keller, Evelyn Fox. 1985. *Reflections on Gender and Science*. New Haven, CT: Yale University Press.

Klein, Naomi. 2023. *Doppelganger: A Trip Into the Mirror World*. New York: Farrar, Straus, & Giroux.

Nelson, Alondra. 2008. "Bio Science: Genetic Genealogy Testing and the Pursuit of African Ancestry." *Social Studies of Science* 38 (5): 759–83.

———. 2016. *The Social Life of DNA: Race, Reparations, and Reconciliation after the Genome*. Boston: Beacon Press.

Pollock, Anne. 2012. *Medicating Race: Heart Disease and Durable Preoccupations with Difference*. Durham, NC: Duke University Press.

Schaefer, Donovan O. 2022. *Wild Experiment: Feeling Science and Secularism after Darwin*. Durham, NC: Duke University Press.

Sedgwick, Eve Kosofsky. 1997. "Paranoid Reading and Reparative Reading; Or, You're So Paranoid, You Probably Think This Introduction Is about You." In *Novel Gazing: Queer Readings in Fiction*, 1–37. Durham, NC: Duke University Press.

Zygon®: Journal of RELIGION & SCIENCE

Reeves, Josh A. 2024. "Kuhn, Conspiracy Theories, and Scientific Progress." *Zygon: Journal of Religion and Science* 59 (3): 811–24. DOI: https://doi.org/10.16995/zygon.16721

Open Library of Humanities

Kuhn, Conspiracy Theories, and Scientific Progress

Josh A. Reeves, Director, Samford University Center for Science and Religion, Birmingham, AL, USA, jareeves@samford.edu

This article offers an appreciative but critical response to Donovan Schaefer's book *Wild Experiment*, focusing on the connection between emotion and reason, particularly in the context of conspiracy theories and scientific inquiry. While acknowledging the valuable insights provided by Schaefer's cogency theory and its emphasis on the role of emotions in shaping beliefs and reasoning, I argue for a more nuanced understanding of the factors that contribute to the popularity of conspiracy theories and the success of science. In particular, I challenge Schaefer's characterization of scientists as primarily driven by cold emotions and a fear of making mistakes. Instead, I emphasize the social structure of science and institutional practices that enable collective intellectual vigilance and the advancement of reliable knowledge.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

It is an honor to respond to Donovan Schaefer's book *Wild Experiment*. As a member of the 2022 International Society for Science and Religion book prize committee, I am on record about what a valuable and enjoyable book it is, one that expertly navigates through an impressive number of topics while avoiding superficiality. Schaefer's overarching thesis is not only compelling but also contributes substantially to ongoing discussions across multiple fields.

My research (Reeves 2012, 2021, 2022) focuses on Christian mistrust of science and religious vulnerability to conspiracy theories, and so my response to *Wild Experiment* will focus on these topics. My main questions are not about the connection of emotion and reason, for which Schaefer compellingly argues, but how well this insight can account for the popularity of conspiracy theories or scientific progress. In the epilogue, Schaefer (2022, 231) suggests that cogency theory enhances our understanding of scientific success alongside its social character and sustained engagement of the world. However, a considerable portion of the book leans heavily on cogency theory as the main lens through which science is explained, potentially overlooking the chance to weave in additional dimensions for a fuller and more nuanced exploration of scientific inquiry. My response invites Schaefer to explore how these other elements might expand or adjust his depiction of science.

Emotion and Conspiracy Theories

This first section aims to accurately summarize Schaefer's discussion of conspiracy theories and science. Upon my initial reading of *Wild Experiment*, I perceived it as challenging the traditional dichotomy between emotion and reason, an argument that I was well-acquainted with from my undergraduate studies in psychology. There I learned that thinking and feeling are so deeply interwoven that attempting to separate them into distinct categories is not only impractical but also misleading. This perspective echoes the longstanding nature versus nurture debates, where efforts to isolate one aspect from the other oversimplify the intricate interplay that shapes human behavior and development.

However, it quickly became clear that Schaefer is arguing for a stronger thesis. Instead of viewing reason and emotion as intertwined entities, he posits thinking itself is a manifestation of feeling. As he (2022, 5) contends, there is "no thinking that is not feeling" and to change our minds is to change how we feel (2022, 9). Or more succinctly: "thinking is feeling" (2022, 29).

For those adhering to a more conventional perspective, which acknowledges the entanglement of reason and emotion but does not equate reason entirely with emotion, a question arises: How do we find the truth? Traditionally, reason is seen as the faculty that evaluates the congruence between our thoughts and reality. If we were to suggest that truth can be discerned through emotion alone, then the challenge becomes how to determine the reliability or "truthfulness"

of our feelings. While the process of reasoning towards the truth often brings a sense of satisfaction, it is conceivable to encounter situations where the truth is discomforting or painful, such as the moment one acknowledges an addiction. How do we pursue the truth when it is not always pleasurable?

Schaefer addresses this dilemma by drawing upon David Hume's differentiation between hot and cold emotions. Hot emotions, like anger, fear, and love, move us to direct action because they are pleasurable. Cold emotions, which include the fear or shame associated with potential errors or deceit, also motivate action but on slower time scale. For instance, an individual grappling with addiction might opt for decisions that favor their long-term emotional health, despite potential short-term discomfort. Cold emotions allow one to endure uncertainty and resist the temptation to embrace beliefs merely because they are desirable. Ideally, hot and cold emotions work together and balance out the limitations of the other.

With this framework, Schaefer explains the lure of conspiracy theories and the success of science. According to a more traditional understanding of reason and emotion, conspiracy theories are a failure of thinking correctly. But for Schaefer (2022, 34), this cannot account for the role of emotions in reasoning. As he says, "Conspiracy theory doesn't try to explain. It aims to electrify the world with significance, saturating it with click." Those who accept conspiracy theories believe because it feels good; they want it to be true and will artificially simplify the messiness of the world to experience more pleasure. Conspiracy theorists, in his view, have removed the dampeners on their emotions provided by the cold emotions (2022, 53).

By contrast, science can uncover truths and facts about the natural world because it harnesses the power of cold emotions to maintain its immune system against the pleasure of confirmation bias (2022, 55). This establishes a delicate balance between the joy derived from engaging with the world and the discipline to resist seductive but unfounded "truths." The interplay of pleasure with a vigilant skepticism forms the core of scientific inquiry, allowing it to navigate and demystify the complexities of the natural world without succumbing to the simplistic and emotionally gratifying narratives favored by conspiracy theories.

Three Types of Conspiracy Theorists

In this section, I explore the reasons people are drawn to conspiracy theories by proposing three distinct categories. My concern is that while Schaefer's analysis offers valuable insights into some aspects of conspiracy theory psychology, it might not capture the full range of those who subscribe to such theories.

The first type of conspiracy theorist can be dubbed the "True Believer" and is vividly captured by *Wild Experiment*. These individuals approach conspiracy theories with enthusiasm because they provide a thrilling alternative narrative that transforms the mundanity of the real world into an intricate, gamelike

puzzle. They enjoy taking the adventure down the proverbial YouTube rabbit hole, where each click leads to increasingly novel and unorthodox interpretations of reality. It is hard to convince these individuals of an alternative way of considering the facts because their beliefs feel too good to give up.

In my experience, this type of conspiracy theorist is best exemplified by those who deny that the US landed on the moon in 1969, a view which seems to be popular with young males who consume lots of social media. The moon landing conspiracy theory is not just about disbelief in mainstream sources of information; it is about the joy of assembling an alternate version of reality from the pieces found along the way, challenging the status quo, and transgressing what is acceptable knowledge.

The next type of conspiracy theorist can be called the "spiritually anxious." This classification diverges significantly from the True Believer because of its foundation in intense negative emotions, predominantly fear, and suspicion towards perceived outsiders. Unlike the True Believer, who delights in the exploration and challenge of conventional narratives, individuals within this information bubble are often driven by a profound sense of distrust and anxiety about forces or entities they believe are working against them or society at large.

This category of conspiracy theorist frequently belongs to conservative Christian groups, particularly those influenced by dispensational theology. This theological perspective can lead followers to harbor irrational fears towards public figures, perceiving them as potential Antichrists, and to possess exaggerated concerns about technology and the emergence of a one-world government. It often encourages the aggressive confrontation of nations labeled as "evil" in an effort to fulfill biblical prophecies. For instance, Pat Robertson (1991, 37, a well-known televangelist, in his book *New World Order*, suggests that even well-meaning presidents are unwittingly furthering the agenda of a secretive group aiming to establish a global dominion under the influence of Lucifer and dark spiritual forces (Clouse et al. 1999, 132).

I worry that this second type of conspiracy theorist does not fit well in the framework of *Wild Experiment*. Schaefer would presumably explain these groups by the pleasure derived from animosity towards out-groups. For those who grow up in these communities, however, conspiracy theories are often not enjoyable and cause spiritual trauma that comes from living with a pervasive sense of danger, mistrust, and paranoia. Individuals are conditioned to be perpetually vigilant against an invisible adversary believed to be manipulating global events and power structures.

An important aspect that is absent from Schaefer's analysis is the role of leadership within these communities in leveraging conspiracy theories as tools of control. By framing certain groups as different and alien, conspiracy theories effectively create out-groups, a strategy employed to solidify the authority of religious leaders. This mechanism might indeed generate a sense

of engagement or "click" for the leaders, yet it results in an overall net-negative emotional satisfaction for the followers themselves. While I doubt Schaefer would disagree with the way emotions can be manipulated to reinforce existing power structures, *Wild Experiment* does not say much about this possibility. As Schaefer continues to expand his research program on cogency theory, I think there is much more he can add to integrate fearful emotions and social control into his account of conspiracy theories. However, addressing this dimension might necessitate shifting away from the primacy of emotional pleasure as the central explanatory variable.

The third and final type of conspiracy theorist are those who live in "skeptical information environments." In my experience, many people in Christian communities accept conspiracy theories in the way high school students memorize facts for a history class: no excitement of "click," only the drudgery of mental effort. Most teachers have the experience of being unable to get students to emotionally engage class material. Likewise, many students in Christian high schools and youth groups can be forced to internalize information when they rather be socializing or playing video games. One way to identify this type is that they often hold irreconcilable views without feeling the emotional need to resolve them. They comply with the expectations of their community, providing the "correct" responses with no genuine curiosity or inclination to investigate further.

This raises the question: do humans accept all beliefs because of emotional investment, or are some beliefs accepted "secondhand" from trusted figures from one's communities? Schaefer's position can be read in a way to say that we emotionally vet all our beliefs, reminiscent of Descartes supposedly subjecting all his beliefs to critical doubt in his *Discourse on Method*. While I would agree in general with Schaefer's (2022, 11) statement that "[g]ood knowledge isn't knowledge that has been drained of feeling," I would argue that we also hold many beliefs that are emotionally inert. William James (2000, 91) captures this well, "Truth lives . . . for the most part on a credit system. Our thoughts and beliefs 'pass' so long as nothing challenges them, just as bank-notes pass so long as nobody refuses them." An example might be my belief that water is composed of one water and two hydrogen atoms or that my mother was born in the state of Alabama. These beliefs are accepted without question until a sense of discomfort suggests that it might call for closer examination. Emotion, in this context, acts more as a detector for questionable beliefs rather than as a source of pleasure, but it takes too much effort to assess all our new beliefs. As two psychologists have recently argued, susceptibility to partisan fake news is better explained by lack of reasoning (i.e., laziness) than by motivated reasoning (i.e., a dopamine hit of pleasure) (Pennycook and Rand 2019).

If this is correct, then it requires a reevaluation of why some individuals are susceptible to conspiracy theories. The issue may not lie in an overreliance on hot emotions or a lack of concern for accuracy, but in the low quality of secondhand information that individuals receive. As noted by a scholar in scientific communication (Kahan 2017, 1), "The problem, in short, is not a gullible, manipulated public; it is a polluted science communication environment." Skepticism towards science is often fueled by an information landscape that poses significant challenges for the average person to navigate independently. Encouraging laypeople to apply their cold emotions is insufficient if they lack the background knowledge to make informed decisions on their own and cannot easily identify sources of reliable information.

What Makes Science?

Another concern that I had while reading *Wild Experiment*: it seems to place conspiracy theorists and scientists on a continuum, with emotional sophistication (represented by scientists) at one end and emotional naivety at the other. Schaefer (2022, 9) argues that scientists are motivated by cold emotions and a fear of making mistakes, asserting, "Science possesses a potent mechanism for self-correction: our inherent desire to be accurate." This contrasts with conspiracy theorists, who are depicted as simplifying reality to indulge in the satisfaction derived from "clicks." But I think this oversimplifies a complex reality: scientists, like anyone else, are susceptible to confirmation bias. As I have previously discussed, individuals can fall into the trap of conspiracy theories even when employing cold emotions, showing that the dichotomy between scientists' and conspiracy theorists' emotional engagements may not be as clear-cut as suggested.

I think the psychological literature shows that most humans, regardless of scientific training, are inclined to weigh information (using the cold emotions), otherwise we would then lack the skills we need to navigate modern life. Without "epistemic vigilance," we would fall for every get-rich-quick scheme or other plan to take our money and attention (Sperber et al. 2010). Because of the importance of discernment regarding information, even children are not automatically trusting when it comes to accepting beliefs from others (Sperber et al. 2010, 377). Children have been shown in developmental studies to resist testimony that conflicts with their own prior observations. They can also grasp the difference between stories that aim to represent reality and those that do not, can monitor informants to assess their reliability, and will correct a parent/teacher who gives inaccurate information.

Other research shows that lay communities are not impermeable echo chambers (Barberá et al. 2015; Dvir-Gvirsman et al. 2016; Gentzkow and Shapiro 2011). While people express a preference for information that is consistent with their own previous beliefs, they do not automatically reject

information from other ideological sources (Scharrer et al. 2017). Rather than viewing the public as emotionally credulous consumers of misinformation, it is better to see much of the public as constrained consumers who reason by seeking information and beliefs from cultural groups with which they identify (Kahan 2017, 9). The difference is that the public will not believe just anything but can be misled by relying upon faulty information from trusted sources. "Tribal rationality," from this perspective, is not an illogical tendency to believe whatever your superiors tell you but is rather a useful shortcut that relieves one of the cognitive burdens of having to assess someone's trustworthiness (Ehret et al. 2017; Goldberg et al. 2019).[1] Many conspiracy theorists are driven not by the allure of "click," but by the desire to avoid work, which, like physical labor, can be unpleasant. Increased analytic thinking has been shown to reduce belief in conspiracy theories, it is just difficult to motivate the public to engage in such scrutiny. As the psychologist J. W. van Prooijen (2018, 90) says, "Analytic thinking reduces the tendency to believe conspiracy theories, and, consistently, efforts to stimulate analytic thinking (e.g., education) are associated with decreased conspiracy beliefs."

I would also not agree that scientists are primarily characterized by their reliance on cold emotions. This perspective may unwittingly perpetuate a historical stereotype that attributes exceptional emotional and intellectual virtues to scientists that are supposedly uncommon in the rest of society. As the historian Steven Shapin (2008) has shown, this notion has deep roots, often linked to narratives about "the scientific method" and its purported capacity to foster intellectual honesty. For example, the French physiologist Claude Bernard (Daston and Galison 1992, 122) once remarked, "The experimenter's mind differs from the metaphysician's or the scholastic's in its modesty because experiment makes him, moment by moment, conscious of both his relative and absolute ignorance. In teaching man, experimental science results in lessening his pride more and more." This quote underscores the idea that scientists, through their engagement with experimentation, gain a unique awareness of the limits of human understanding, equipping them to avoid baseless speculation and superstition.

However, acknowledging the "moral ordinariness" of scientists offers a more grounded and realistic perspective. Research in social psychology shows that academics, like individuals in any other sector, are susceptible to the same coercive social pressures and groupthink prevalent throughout wider society (Duarte et al. 2015). This vulnerability often manifests as confirmation bias, where researchers are inclined to interpret data in ways that reinforce their existing beliefs, rather than seeking evidence that might contradict them. Echoing this sentiment, an expert (Goodstein 2002) on scientific misconduct observes, "Scientists are not disinterested truth seekers; they are more like

players in an intense, winner-take-all competition for scientific prestige and the resources that follow from that prestige." This view challenges the idealized image of scientists as emotionally detached, highlighting the complex human dimensions of scientific endeavor.

A better way to account for the success of science is to focus on its social structure. Given the limits individuals encounter in seeking truth, we in modern societies solve the problem of distinguishing good from bad information by forming institutions and collectively practicing intellectual vigilance through the structured contestation and deliberation of different points of view (Sperber et al. 2010, 383). Rationality emerges from critical interaction with diverse perspectives and by relying on others to help gather and evaluate evidence. As the psychologist Jonathan Haidt (2013, 105) argues,

> Each individual reasoner is really good at one thing: finding evidence to support the position he or she already holds, usually for intuitive reasons. We should not expect individuals to produce good, open-minded, truth-seeking reasoning, particularly when self-interest or reputational concerns are in play. But if you put individuals together in the right ways, such that some individuals can use their reasoning powers to disconfirm the claims of others, and all individuals feel some common bond or shared fate that allows them to interact civilly, you can create a group that ends up producing good reasoning as an emergent property of the social system.

The success of science is best attributed to the social nature of reasoning, which is embedded within formal institutions. While individual scientists may struggle to identify flaws in their own work, the system incentivizes vigilance over the work of peers to prevent any unfair advantages. Scholars are motivated to expose subpar scholarship and to remain open to innovative and potentially groundbreaking theories, understanding that overlooking such theories could place them at a competitive disadvantage (Turner 2013, 266). This mutual oversight ensures that only competent scholarship is advanced, thereby enhancing the discipline's social authority. A scholar's credibility is derived from their affiliation with reputable institutions. Consequently, our trust in experts is fundamentally a trust in the integrity and rigor of the institutions that validate their work (Menand 2010, 105). We should believe in scientific experts not because of their cold emotions but because we believe in the institutions in which they participate (Wells 2013, 85).

If this section is accurate, then much more needs to be added to the argument of *Wild Experiment* to account for why science generates reliable knowledge. While the concept of "click" undoubtedly contributes to the narrative, it would be beneficial for Schaefer to enrich his analysis by weaving cogency theory

together with additional dimensions of scientific inquiry. This would offer a more comprehensive understanding of the mechanisms through which science achieves its reliability, emphasizing the importance of the social structures and institutional practices that underpin scientific progress.

Kuhn and the Training of Desire

Perhaps the place where I had the most disagreement with Schaefer was with his characterization of Thomas Kuhn. I do not know Michael Polanyi's work well enough to comment on its similarity to Kuhn, or to the influence of the former on the latter. But I think Kuhn, for all the attention and scholarship on his work, still has insights to offer the field of science and religion (Reeves, forthcoming).

Kuhn spent his career post-*Structure* clarifying that he did not advocate for relativism, a plea that I find persuasive. He proposed a Darwinian model of scientific progress, which says that while scientists cannot claim to have attained absolute truth—where theories perfectly reflect reality—they can judge that some theories are superior to others. Indeed, that scientific communities do transition to new and better (in terms of puzzle-solving ability) paradigms is a central thesis of *Structure*. And his incommensurability thesis is just the idea that there is no common standard to judge between competing paradigms, which I think fits the messy history of scientific change.

Kuhn's theory of tacit knowledge is relevant to Schaefer's point about cognitive emotions in the scientific process. While I would hesitate to attribute to scientists any traits not found in other professions, I would agree with Schaefer that scientists have learned to use their cold emotions—a desire to avoid error and the shame that comes from mistakes—in their scientific practice. But I think this point can be developed further. For example, this question remains unanswered in *Wild Experiment*: Is it just people with this emotional profile that are attracted to the sciences, or is there something about scientific training that teaches one to use cold emotions? What Kuhn can offer to cogency theory is an account of scientific training, describing how the emotions of scientists are shaped by scientific practices.

Science for Kuhn "was much more like the work of the goldsmith than the contemplative art of the philosopher" (Buchwald and Smith 1997, 364). Many have missed or underplayed Kuhn's theory of tacit knowledge because *Structure* has been read in an epistemological (or what Schaefer might call cognitivist) way. Because of the focus on epistemology, Kuhnian paradigms were interpreted as a core set of *beliefs* or a worldview that guides scientific inquiry, without which inquiry would be impossible. Scientists have difficulty talking to those who are committed to different paradigms because a paradigm supplies the lens through which one views the world. Acceptance of a new paradigm is like a religious

conversion or Gestalt switch—a sudden shift in belief and perception—because there is no common criterion with which to adjudicate between rival paradigms.

An epistemological interpretation of Kuhn is commonplace—indeed, Kuhn himself sometimes promoted it—but misses the more original and important meaning of paradigm given by him. As Kuhn (2012, 187) says in the postscript, "The paradigm as shared example is the central element of what I now take to be the most novel and least understood aspect of this book." Though he uses the word in multiple ways, the most original meaning of paradigm is that of an *exemplar*, a concrete achievement that guides the subsequent course of research in a discipline. Exemplars are specific problem solutions, not universal theories or principles, which provide scientists with a shared research trajectory. Because paradigms are exemplary ways of intervening in particular situations, advocates of those paradigms will try to generalize the skills and actions needed to perform the exemplar to new situations. As Rouse (2003, 108) argues, "accepting a paradigm is more like acquiring and using a set of skills than it is like understanding and believing a statement."

Dropping the "mentalism" behind the notion of paradigm transforms the meaning of incommensurability (Warwick and Kaiser 2005, 405). When members committed to different paradigms have trouble communicating, the problem is not that they cannot construe one another's sentences or follow one another's arguments, but that they cannot grasp the significance of what the other scientists are doing (Rouse 2003, 112). As Rouse (2003, 112) explains, "The more basic issue between proponents of alternative paradigms concerns how to proceed with research: what experimental systems or theoretical models are worth using, what they should be used for, what other achievements must be considered, and what would count as a significant and reliable result." Without an exemplar to provide a common context for the solving of problems, it is difficult to agree to the future course or even proper interpretation of research (Kuhn 2012, 200).

Kuhn's exploration of scientific skill as an embodied practice offers a compelling framework for understanding how scientific emotions (both hot and cold) are cultivated. The sensation of achieving a "click" is not innate; it must be nurtured through education. Although a natural curiosity is common to humanity, the mental discipline to pursue learning is not universal. Learning presents a challenge because it initially involves absorbing information that requires effort and does not immediately gratify. Taking chess as an example, a novice must place faith in their instructor, believing that the teacher has their best interests in mind and can illuminate aspects of the game that the student cannot yet perceive (Herdt 2010, 28). It is often necessary to find motivation in external rewards—whether it is the desire to make one's parents proud or the promise of a treat after a lesson—to dedicate oneself to learning. Only after mastering the fundamentals can one derive enjoyment from the game itself.

Mastering science shares similarities with learning chess, where the initial drive often stems not from immediate enjoyment of the subject but from the anticipation of rewarding outcomes through perseverance. Becoming a physicist, for example, is more than just memorization or reflecting upon the inner meaning of theories; it involves the joy derived from mastering skills such as identifying forces, masses, and accelerations in new physical situations (Kuhn 2012, 189). The acquisition of tacit knowledge brings with it the satisfaction of intuitive understanding and the ability to solve complex problems with ease. Tacit knowledge is typically acquired through immersion in a community already proficient in the field—similar to how one learns to cook by working under an experienced chef. Kuhn's account of scientific progress, then, is also an account of scientific training, showing how the emotions of scientists are formed and calibrated in a larger community.

In sum, I believe Kuhn's perspective on scientific training complements and could enrich Schaefer's discussion on the influence of emotions in scientific inquiry. I am curious to see if Schaefer views this exposition of Kuhn's insights as a constructive addition to his research program.

Conclusion

This paper has used Schaefer's important work in *Wild Experiment* to explore the complex relationship between emotion, conspiracy theories, and science. While Schaefer's arguments for the connection between emotion and reason is persuasively argued, my discussion has highlighted areas where Schaefer can expand and deepen his analysis. I have argued for more complexity in the way people are drawn into conspiracy theories, the reason why science successfully uncovers reliable knowledge, and the role of training in shaping pleasure and emotion.

I conclude by posing some last questions. In discussing the motivations behind scientists' pursuit of their fields, *Wild Experiment* attributes a primary role to pleasure. Although scientists may be more disciplined than conspiracy theorists, the thrill of discovery, or the "click," justifies the endeavor. But this emphasis on pleasure raises questions: is the pursuit of pleasure inherently selfish, tied to personal prestige and status? Or do our natural inclinations lead us beyond self-interest towards a greater good, as suggested by Aristotelian ethics? Essentially, is there a convergence between what is pleasurable and what is good, or is the pursuit of science always an extension of self-interest? These are not trivial questions, of course, but if Schaefer were to address larger questions of human nature, he would have more explanatory resources than just pleasure upon which to explain human behavior in the scientific and religious spheres. A nonreductive account of human nature would point to higher goods and pleasures beyond "click." The power of Schaefer's book comes from his intense focus on examples of the power and danger of pleasure, but I would

be very interested if he were to bring his analytic insight to bear on larger questions of human nature.

I look forward to Schaefer's response and to following his research agenda over the coming years.

Note

[1] Kahan (2016, 11): "It is perfectly rational for them consciously to seek out guidance from such individuals, then, or to form unconscious habits of mind that privilege them as sources of guidance on what science knows. This process is admittedly insular, but it clearly works in the main."

References

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. 2015. "Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* 26 (10): 1531–42. https://doi.org/10.1177/0956797615594620.

Buchwald, Jed Z., and George E. Smith. 1997. "Thomas S. Kuhn, 1922–1996." *Philosophy of Science* 64 (2): 361–76. https://doi.org/10.1086/392557.

Daston, Lorraine, and Peter Galison. 1992. "The Image of Objectivity." *Representations* 40:81–128.

Duarte, José L., Jarret T. Crawford, Charlotta Stern, Jonathan Haidt, Lee Jussim, and Philip E. Tetlock. 2015. "Political Diversity Will Improve Social Psychological Science." *The Behavioral and Brain Sciences* 38:e130.

Dvir-Gvirsman, Shira, Yariv Tsfati, and Ericka Menchen-Trevino. 2016. "The Extent and Nature of Ideological Selective Exposure Online: Combining Survey Responses with Actual Web Log Data from the 2013 Israeli Elections." *New Media & Society* 18 (5): 857–77. https://doi.org/10.1177/1461444814549041.

Ehret, Phillip J., Aaron C. Sparks, and David K. Sherman. 2017. "Support for Environmental Protection: An Integration of Ideological-Consistency and Information-Deficit Models." *Environmental Politics* 26 (2): 253–77. https://doi.org/10.1080/09644016.2016.1256960.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics* 126 (4): 1799–839. https://doi.org/10.1093/qje/qjr044.

Goldberg, Matthew H., Abel Gustafson, Matthew T. Ballew, Seth A. Rosenthal, and Anthony Leiserowitz. 2019. "A Social Identity Approach to Engaging Christians in the Issue of Climate Change:" *Science Communication*, July 7. https://doi.org/10.1177/1075547019860847.

Goodstein, David. 2002. "Scientific Misconduct." *Academe* 88 (1): 28–31.

Haidt, Jonathan. 2013. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Reprint edition. New York: Vintage.

Herdt, Jennifer A. 2010. *Putting On Virtue: The Legacy of the Splendid Vices*. Chicago: University of Chicago Press.

James, William. 2000. *Pragmatism and Other Writings*. Edited by Giles Gunn. New edition. New York: Penguin Classics.

Kahan, Dan M. 2017. "Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition." Cultural Cognition Project Working Paper Series No. 164, Yale Law School, Public Law Research Paper No. 605, Yale Law & Economics Research Paper No. 575. Cultural Cognition Project, Yale Law School, May 24. https://doi.org/10.2139/ssrn.2973067.

Kuhn, Thomas S. 2012. *The Structure of Scientific Revolutions: 50th Anniversary Edition*. Chicago: University of Chicago Press.

Menand, Louis. 2010. *The Marketplace of Ideas*. New York: W. W. Norton.

Pennycook, Gordon, and David G. Rand. 2019 "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning." *Cognition, The Cognitive Science of Political Thought* 188:39–50. https://doi.org/10.1016/j.cognition.2018.06.011.

Reeves, Josh A. 2012. "Theology and the Problem of Expertise." *Theology Today* 69 (1): 34–42.

———. 2021. *Redeeming Expertise: Scientific Trust and the Future of the Church*. Waco, TX: Baylor University Press.

———. 2022. "A Theological Engagement with the Science of Science Skepticism." *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology* 6 (1): 197–214. https://doi.org/10.14428/thl.v6i1.61263.

———. Forthcoming. "Exemplars in 'Science and Religion': A Theological Dialogue with Thomas Kuhn." *Religious Studies*, published online November 28, 2023. https://doi.org/10.1017/S0034412523000975.

Robertson, Pat. 1991. *The New World Order*. Dallas, TX: Word Publishing. http://archive.org/details/newworldorder00robe_0.

Rouse, Joseph. 2003. "Kuhn's Philosophy of Scientific Practice." In *Thomas Kuhn*, edited by Thomas Nickles, 101–21. Cambridge: Cambridge University Press.

Schaefer, Donovan O. 2022. *Wild Experiment: Feeling Science and Secularism after Darwin*. Durham, NC: Duke University Press.

Scharrer, Lisa, Yvonne Rupieper, Marc Stadtler, and Rainer Bromme. 2017. "When Science Becomes Too Easy: Science Popularization Inclines Laypeople to Underrate Their Dependence on Experts." *Public Understanding of Science* 26 (8): 1003–18. https://doi.org/10.1177/0963662516680311.

Shapin, Steven. 2008. *The Scientific Life: A Moral History of a Late Modern Vocation*. Chicago: University of Chicago Press.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. "Epistemic Vigilance." *Mind & Language* 25 (4): 359–93.

Turner, Stephen P. 2013. *The Politics of Expertise*. London: Routledge.

van Prooijen, Jan-Willem. 2018. *The Psychology of Conspiracy Theories*. London: Routledge.

Warwick, Andrew, and David Kaiser. 2005. "Kuhn, Foucault, and the Power of Pedagogy." In *Pedagogy and the Practice of Science: Historical and Contemporary Perspectives*, edited by David Kaiser, 393–409. Cambridge, MA: MIT Press.

Wells, Thomas. 2013. "Democracy Is Not a Truth Machine." *Think* 12 (33): 75–88.

Zygon®: Journal of RELIGION & SCIENCE

Shah, Esha. 2024. "Love Affair with an Idea: Commentary on *Wild Experiment: Feeling Science and Secularism after Darwin* by Donovan Schaefer." *Zygon: Journal of Religion and Science* 59 (3): 825–38. DOI: https://doi.org/10.16995/zygon.16909

Open Library of Humanities

Love Affair with an Idea: Commentary on *Wild Experiment: Feeling Science and Secularism after Darwin* by Donovan Schaefer

Esha Shah, Lecturer, Water Resources Management, Wageningen University, Wageningen, The Netherlands, esha.shah@wur.nl

In this review essay, I discuss how, in writing this book on how thinking is feeling, Donovan Schaefer is having a passionate love affair with an idea, how he pulls the bow string as far and still not enough to the point of breaking. I have three remarks/points of concern/comments specific to the outline of cogency theory. First, *how* thinking is feeling? If thinking is feeling is a process, how does it happen? Secondly, where does reframing science/knowledge as thinking is feeling take us? How can it reframe what we mean by science? And lastly, who is the thinking-feeling scientist-subject?

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

"This is an essay about a love affair with an idea," Donovan Schafer quotes Silvan Tomkins beginning chapter 2. Reading it came another remarkable line in my mind,

> (and I think silently:
> love is a bow-string pulled
> back to the point of breaking)
> (*Poem of the End*, Marina Tsvetaeva translated by Elaine Feinstein, 1971)

In this review essay, I want to discuss how, having a passionate love affair with an idea, Schaefer pulls the bow-string as far and still not enough to the point of breaking.

Before I come to my remarks, I would like to remember Evelyn Fox Keller, her towering personality, her era-defining, pioneering, brave biography of Barbara McClintock published in 1983 that gave a whole new dimension in the way we understand affects and emotions in relation to the making of science. Personally for me her work has a life-changing inspiration on my own exploration of how all forms of knowledge emerge from deeply personal and emotional place. I would like to quote here what Keller wrote to me in 2018 after reading the draft of my book *Affective History of the Gene: Who is the Scientist-Subject?* about to be published by Routledge at that time (Shah 2018). In her email, Keller applauded my ambition, calling me a brave woman, "Indeed, yours is an ambition that I shared quite fully in my early work in the history and philosophy of science. Certainly, to bring the affective springs of scientific reasoning to the fore was a central aim of my McClintock book, and even more explicitly so, of Reflections on Gender and Science." Keller described her book on Gender and Science as psychoanalysis of science. But then her tone turned somber when she wrote, "Ultimately, however, I gave up. The resistance in the history of science community to any kind of subjective analysis…was just too strong."

I don't think Keller ever gave up, and still she felt a sense of being defeated with regards to her pioneering work on emotions and science. I wish she were here. Because 4S awarding Ludwik Fleck prize to Donovan Schaefer's outstanding contribution on how thinking is feeling has vindicated the whole community of scholars working on similar themes. I feel deeply honored to have been asked to participate in this moment of celebrating Schaefer's book that has opened up this space for the discussion on the topic that at best is marginalized and at worst treated with suspicion and occasionally even subjected to antagonistic aggression.

Reading Donovan Schaefer's book has been an experience. At times I felt almost dizzy as his intellectual range is breathtaking. Schaefer effortlessly engages with a vast range of topics normally not discussed in one breath, in one book. He gives us an outline of cogency theory on how thinking feels and then debates it in relation to racialized reason, secularism, Darwin and Huxley

and their science, new findings in neuroscience, creationism and more. Each chapter is like a stand-alone episode of the most gripping drama, reading which I had a Hitchcockian feeling—never a dull moment. The book is a page-turner. The nuance, the richness, the poetics of Schaefer's prose—the book is indeed a testimonial of a passionate love affair with an idea.

Let me now elaborate on how Schaefer pulls the bow string as far and still, I intend to argue, not enough. I have three remarks/points of concern/comments specific to the outline of cogency theory. First, I want to unpack *how* thinking is feeling. If thinking is feeling is a process, how does it happen? Secondly, where do reframing science/knowledge as thinking is feeling take us? How can it reframe what we mean by science? And lastly, who is the thinking-feeling scientist-subject?

I want to first unpack what I believe is one of the key statements of the book, "knowledge-making is not just entangled with feeling, as some claim (*Feeling can shape how we think, under certain circumstances*), but encompassed by it (*Feeling is necessary for thinking; there is no thinking that is not feeling*)" (Schaefer 2022, 5, emphasis original). I see here two separate statements that I would like to discuss, although Schaefer does not make such distinction as I do. One, feeling is necessary for thinking, and second, there is no thinking that is not feeling.

Schaefer explores how thinking is feeling with interconnected tropes: click and agonism of emotions culminating into what he calls cogency theory. Schaefer writes how all forms of knowledge is formed by inseparably combined effect of thinking and feeling, how knowledge-making are versions of micro-level delight in the subtle click of things coming together. Click is how it feels when pieces of information coalesce, when knowledge-making is solving puzzles that draws us into pleasurable sphere of playing by thinking—Schaefer writes. Schaefer variously describes click—it happens when we organize a maze of noise into a tidy whole, we piece together secrets, when we derive the joy of solving good mystery, when a detective story neatly wraps up elements of a plot in a satisfying resolution, when we thrill to the immaculate resolution, we experience a symphony of details coming together in sudden alignment. Click is the conceptual trope that binds whole of Schaefer's book together. How do pieces of information coalesce? How do details come in alignment producing a click? Even when Schaefer's click relates to both product and process of the knowledge-making, it is more than often described as a product, an outcome of the knowledge-making—click is when details fall in alignment. I want to particularly focus on unpacking how click may relate to the processes of knowledge-making.

Here I would like to refer to Loraine Daston's concept of "jizz," which further refers Ludwik Fleck's work on genesis of the scientific fact, which is remarkably close to Schaefer's click, except that Daston and Fleck explore the processes by which thought and emotions collapse into a unity. Daston (2008) explains how the jizz in scientific observation is sure, swift, silent, and happens without

a pause for mental analysis. She further explicates how learning to see like a scientist is a matter of accumulated experience, a matter of habit formation, how the scientist has to go through a gradual process of training before experience turns into a habit whence for a mature scientist it becomes possible to see things all-at-once, like a jizz. Daston further clarifies, while the scientist is seeing things all-at-once, in a jizz, the conscious reason has no idea how this is done, how conscious knowledge is formed here by unconscious habits and intuitions. Daston here refers to Ludwik Fleck's interpretation of thought collective and genesis of the scientific fact explained in a musical metaphor. It all starts with the confused notes which is followed by hummed and inaudible tunes gradually turning into a melody. The melody turns into a thought collective once the "co-workers" listen and tune their sets until these became selective. The melody could then be heard even by the unbiased person—meaning, it crystallizes into a fact (Fleck [1935] 1981). There is a similarity between Daston's experienced scientist able to see in a jizz, all-at-once, because a chain of experiences turn into unconscious habit, and Fleck's genesis of the scientific fact emerging first as chaos, then confused notes, then collective tuning into a melody—both are processes, they are emergent, subjected to time, and both constitute a chain of embodied experiences that are silent and swift, intuitive and unconscious, while they form and structure rational reason (Shah 2017).

What is remarkable about Daston's exploration of jizz and Fleck's genesis of melody is that the subtle ways in which the irrational and rational, conscious and unconscious, feelings and thought, experience and fact are seen as coalescing in place. Here the emergence of the conscious scientific reason has to necessarily go through the path of the unconscious. The way perception turns into memory into experience into habit has psychology, and yet again, the conscious reason have no inkling how it is done when it is done. This all-at-once-ness in Daston (2008) is explained in the words of Descartes, "the arguments so speeded up that it bursts upon the mind as a single cognitive event" and "[n]o amount of explicit reasoning, even mathematical reasoning, can compete with it."

I interpret and relate Schaefer's first statement—*feeling is necessary for thinking*—with Daston and Fleck's exploration of the play of the conscious and unconscious, reason and intuition, in the making of knowledge. The jizz, all-at-once-ness, is not about some random eruption of unconscious upon the conscious. This is about the unconscious fundamentally structuring the conscious reasoning without the conscious even knowing it. In fact, this is about the indispensable necessity of embodied experiences turning into unconscious habit for any conscious knowledge to form.

Even when Schaefer acknowledges the role of unconscious in the making of knowledge, the thinking is feeling is still predominantly conscious process. The cogency theory is inspired by Micheal Polanyi's idea of tacit knowledge absorbed through embodied experiences and intuitions and persuasive passions rather than

abstract language. Schaefer in fact argues that selection of Thomas Kuhn and his work on structure of scientific paradigm instead of Michael Polanyi as gadfly of science studies is why feeling/thinking binary still has such power. Schaefer also further discusses the portrait of a multilayered self in affect theory—how this self is a continuum from micro to macro stretching both within and below our field of awareness (Schaefer 2022, 65) and how affect theory lights up the interior of reason below the threshold of consciousness awareness shapes how we think (2022, 70)—the cogency theory is still significantly tied down to a very conscious, carefully configured or constructed or calibrated, amalgamation, agonism, of emotions in the making of knowledge (2022, 34–36, 55). The role of psychoanalytical unconscious in creating multilayered affective self either emerging from or challenging Freudian traditions is briefly discussed in the chapter on affect theory and still not persuasively incorporated in the cogency theory. How the click happens as a complex, below the surface of consciousness, process in which how *feeling is necessary for thinking*, is not convincingly explored. This has consequences for the theory of emotions in the making of the scientist-subject or knowledge-seeker, I return to this point later in my remarks.

Discussing the second statement—*there is no thinking that is not feeling*—Schaefer makes repeated references to how thinking feels involve pleasure, joy, curiosity, playfulness, how it's a longing, a need, a passion for intellectual beauty. And then Schaefer also refers to what he calls countervailing forces such as shame, fear, anxiety, and frustration of not able to solve the puzzle, and embarrassment and nervousness the possibility of error or being wrong may cause. Schaefer argues how scientific rationality is a carefully constructed amalgam of feelings in tension; how science is born in an agonism of contesting emotions; how cogency means confluence and contest of forces in agonism, a struggle for different priorities; how good knowledge is the product of complicated operations of feeling—an agonism, an invisible clashes of forces holding our desire to know the world in tension; how agonism of intellectual feelings is the real driver of good knowledge; how knowledge production collapses when this emotional agonism is removed; how we feel our way to knowing via a rich agonism of pleasure and frustration (Schaefer 2022, 9, 12, 16, 19, 36, 41, 48, 54). Even when Schaefer anchors his cogency theory on emotional agonism that oscillates between excitement and frustration, my impression is that the exploration of feelings remains overwhelmingly on the side of the pleasure principle. Joy of discovery, exhilarating sense of science, research as compulsive excitement, intellectual passion comprising curiosity and playfulness, role of intuition, persuasive passion, love of elegance or beauty, the subtle joy of pieces of information snapping together, joy and exhilaration as guidelines of scientific method, discovery as an indescribable pleasure which pales the rest of life's joys, to pursue truth is pleasurable dancing with sentiment of rationality, transition from a state of puzzle and perplexity to rational comprehension is full of lively relief

and pleasure—these are some of the expressions repeatedly appear throughout the book. And of course there is agonism of countervailing emotions. But what about the rest, the whole complex spectrum of humanly possible emotions? What about the human subjectivity firmly grounded in complex interplay of emotions? And this is where I construe that Schaefer's cogency theory gets on the defensive backfoot. There are a series of statements in which this is visible.

If cognition is felt, it is *susceptible* to the affective ecology from which it emerges. Knowledge production is emotional, so it is always *at risk of messing* together with the *other spreading inkblots* of feeling making up our embodied lives (Schaefer 2022, 11). Wheeling matrixes of feeling *lure* us to *misbegotten* belief (2022, 11). If knowledge is felt, it is always in intimate proximity to *other things we feel*—things we want—including our *secretly savored prejudices* (2022, 29). Science as a felt process that registers the way things are in the world produces good knowledge, but it's also *susceptible to contamination*. It *messes together with the other emotions* that give form to our social, embodied lives (2022, 8). If reason is itself a passion, then it messes together with the *other feelings* that define our lives. It eats *(messes)* at the same table with them, and it *sloshes around (messes)* in the same cup, creating a strange brew of felt intuitions from which thought emerges, dripping (2022, 8). How powerful *prejudices "seep"* into our intellectual bedrock (2022, 11). Science as a felt process that registers the way things are in the world produces *good knowledge*, but it's also *susceptible to contamination*. It *messes* together with the other emotions that give form to our social, embodied lives (2022, 8). Good knowledge . . . is the product of complicated operations of feeling—an agonism, an invisible clash of forces holding our desire to know the world in tension and trying to outmaneuver *other felt priorities that might muddy the waters* (2022, 12). Knowledge is always *susceptible to being pulled off course* or landing awkwardly on skewed surface (2022, 18). By studying how rationality works by feeling its way along offers our best chance to name and cultivate the habits and dispositions that make up the sense of science and *veer away from the traps* set by our own sweet tooth (2022, 24). Scientific rationality as carefully calibrated configuration of intellectual passions in tension is always *susceptible to conservative tilt that can obscure truth* (2022, 55). Because rationality is affective, it's *susceptible to contamination* by *other emotional* ink-bolts surrounding it (2022, 70). Because secular reason is emotionally determined, it is susceptible to messing together with other desires and preferences (2022, 84). Science's *susceptibility to contamination* by racism and sexism (2022, 222). (Emphasis mine in all quotations.)

I want to unpack here the notions of other emotions and good knowledge. On the two extremes of the spectrum of agonism of emotions are pleasure and joy of the click on one side and shame, fear and frustration of failing to achieve click on the other side. Both these sides remain firmly anchored on the epistemic, intellectual components of doing science—click is when puzzle is solved and frustration, shame and embarrassment when it doesn't. In between

are other emotions that variously mentioned in the quotes above, and some of them produce—worry, concern, anxiety, unease. The cognition as feeling is susceptible, at risk of messing up with other feelings, including savored prejudices, it is susceptible to contamination, again, of other emotions. The choice of the lexicon contamination appearing throughout the book assumes there was something that was pure and then it got mixed up with bad stuff. If there was no origin purity then there can't be contamination; if it never was pure then it was always already contaminated; if there is no original then there can't be copy because then all are only copies. But even if the choice of the word contamination may not be that consequential as I might be projecting, still, there are multiple expressions in Schaefer's book involving susceptibility to other emotions that form the core of the discussions on conspiracy theory and racialized reasons. The message comes out of these discussions that make me feel uncanny—we need to be beware of other emotions, possible contaminations, risks, susceptibilities, prejudices, misbegotten beliefs. Oopps, there are emotions in the vicinity, be careful, when you open the pandora's box!

Anchoring cogency theory in the agonism of emotions as the conceptual lexicon is intriguing. The term "agonism" derives from the Greek "agon" meaning "painful struggle, conflict and competition or dispute" (Oxford English Dictionary). The meaning of agonism of emotions then presupposes separate categories of feelings or emotions in conflict—the good, bad and ugly—the pleasure, the countervailing shame or embarrassment, and the other. Note that these are very conscious emotions in carefully calibrated conflict. Unlike Daston and Fleck for whom the jizz and making of melody are rather unconscious processes—the mental faculty does not have a clue how it is done—Schaefer's process of knowledge making is a cogent contest of forces, a struggle of different priorities, a constant measuring and remeasuring of the felt weight of facts (Schaefer 2022, 9), the sense of science is a dynamic of multiple intellectual affects held in tension (2002, 34). This aspect of cogency theory again has consequences for the way we conceptualize the thinking-feeling scientist subject. Later in my remarks.

Schaefer then also relates different emotions in—what sounds like agonising agonism—with different forms of knowledge—good knowledge, successful or failed science, true and false science, science on and pulled off track. How scientific rationality is built—how it inculcates confidence or gets impaled by prejudice, how it succeeds or falls?—asks Schaefer. Feeling makes science work, but it also leads to the collapse of good knowledge (Schaefer 2022, 4), we can gain an even better understanding of the sense of science, though, by studying how it malfunctions (2022, 49)—writes Schaefer. My question then is: can we separate prejudice from non-prejudice in individual scientist's subjectivity? And is it possible to neatly relate them with successful or false science, good or bad science?

I have no doubt that Schaefer's intentions here are clear—he wants to show that inducting emotions in the making of knowledge does not make us awry. The most intriguing of all Schaefer's statements—in the flow of reading, I had to read it twice pondering why doesn't it make sense?—"From perspective of cogency theory, good knowledge, too, is made by emotions" (Schaefer 2022, 11). What did he say that good knowledge, too, is made of emotions? In my interpretation Schaefer refers here to how emotions conventionally get bad reputation; how reason and passion are generally arrayed in opposition and related to good and bad science; how we have been made to believe that science is rational because it is not emotional, it is objective because it is not subjective; how emotions are projected as nothing but personal bias that distort objective science; how personal passions and emotions are disruptive, dangerous, childish, feminine, irresponsible, they must be ignored, denied, controlled, relegated to literature and art and biographies, they are not generally the stuff that science is made of. Steven Shapin (2012) calls it a dustbin conception of subjectivity—the bin collects those stories that deflate, disrupt or disorder objectivity. Emotions only distort and make bad science—this is indelibly written of our collective unconscious. Schaefer challenges this mammoth tradition of logical positivism deeply entrenched in our believes, institutions, and intuitions about what is good and bad science. He insists how emotions do not distort science, how good knowledge, too, is made by emotions; how collapsing the binary of reason and emotion doesn't leave us adrift. In doing so, however, Schaefer's emotions get separated in pleasure, shame and other emotions, (good, bad, and ugly emotions), which then get connected to good and bad science.

Can we really separate different emotions in the always-emerging subjectivity of an individual scientist at work? Do scientists at work carefully calibrate various emotional responses? In my own book-length exploration of how in the creative struggles of five pioneering scientists, reductionism in the history of genetic science was sustained, adopted, questioned, and challenged—I did not find such conscious, carefully calibrated, agonism of emotions (Shah 2018). Doing science is indeed an emotional struggle—but it happens below the surface of conscious recognition in a way that it is not only impossible to separate emotions for the person experiencing it, but even for an outsider it may need several biographies to excavate those archaeology of emotions and relate them with the process and product of the making of science.

Schaefer himself discusses at length how in his manuscript *Origin* Darwin questions racial differences, while in his book *Descent* he affirms them; how "Darwin's staunch repudiation of anti-Black racism was undercut by his own embrace of cultural racism"; how Darwin never used the word white or Anglo-Saxon, but he wrote about finer gradations of intellectual and moral character of highest races and lowest savages (Schaefer 2022, 151). Not only that it has

taken several biographies to make these internal contradictions visible and still they can't be so neatly related to good or bad science.

Schaefer discusses philosopher Imani Perry's work on how racism is shaped by visceral responses that operate within the process of reason, how racist logic feels right, how it operates outside the conscious awareness of being inadequate and intentional. My own interpretation of Perry's work would be how racist thought is felt right in the body of the person thinking feeling it. Meaning, thinking as feeling when viscerally experienced, it feels right from the perspective of the one who is experiencing it. So from the perspective of the person producing knowledge, there are no prejudices then. There are no other emotions that may contaminate, mess with, slosh around.

By especially invoking the category of other emotions and its relation to bad or false or failed or pulled off track science, or with other forms of problematic knowledge like conspiracy theories or racialized reason, Schaefer is trying to maintain a moral hygiene by, paradoxically, rationally classifying emotions in different categories—there are intellectual emotions on two extremes and in between there are other emotions. This framing exactly prevents Schafer to pull the bow-string to the point of breaking. It's a love affair but not yet love. Because it is not possible to maintain such conscious and careful emotional hygiene in the making of any scientist's or knowledge-seeker's subjectivity. For the person experiencing thinking as feeling, it is a muddle. Embodied experiences of emotions manifest as muddling through, a chaos, all kind of emotions enmeshed beyond conscious recognition. That is why it is terrifying, by making thinking as feeling we open the can of whole worms. If we open the can, we cannot let the heads walk in and tails be separated and left behind. Can we? My own question then is to inquire—who is the thinking feeling scientist-subject? In my own book-length exploration of scientist-subject, I see knowledge-seeker as a self that is not only a fundamentally feeling, suffering, experiencing affective self but most importantly, this self is incoherent, contradictory, heterogeneous, split between the conscious and unconscious, both fictional and real at the same time, the self that comes into being only in relation to others. I have argued how this affective self is profoundly constitutive of the method and philosophy of science, but it goes both ways—making choices in science also means choosing the affective mode of existence. My claim is that intellectual paradigms are affect worlds, in other words, the conceptual theories are isomorphic with the world emotionally and existentially desired (Shah 2018).

I now want to reflect on the way Schaefer separates good and bad science/knowledge/logic. He interprets Perry's work on racism as visceral response that operates within the process of reason, how for the person experiencing it, racist logic feels right. Schafer also elsewhere in his book discusses at length how racialized reason are created by feelings. Rather than a set of propositional claims about who's up and who's down, racism is sunk deep into the bodies of

both its agents and its victims, Schaefer writes (2022, 24). However, despite this clarification, Schaefer interprets Perry's argument to claim how these visceral responses in intimate connection with emotional machinery of scientific knowledge production threatens to pull science off course (Schaefer 2022, 8). What would science being on and off course mean here? Schaefer writes that affects are the pulse of reason. This is both why science succeeds and why it fails (21). But then Schaefer writes that click can also totally derail the search for truth, licensing our prejudices as what feels true, sanctifying what we think we know and rendering it immune to challenge (24). We can gain an even better understanding of sense of science, though, by studying how it malfunctions (49). (All emphasis in quotations mine). There are thus various connotations of science—good and bad science, science gone off course, science that succeeds or fails, prejudices derailing search for truth, other intellectual emotions messing up our thinking, science that malfunctions.

"Can there be feminist science?", asks feminist epistemologist Helen Longino and answers it in the negative (Longino 2001). In principle it is possible to do science as a feminist, but there is no feminist science, says Longino. This is an old debate among feminist epistemologists in science studies, somewhat done and dusted (Shah 2013), but I want to rehearse some of the main points of this debate in relation to these different connotations of science appearing throughout Schaefer's book, to remark on what is understood as science and how may thinking as feeling alter it.

Following the publication of Keller's biography of McClintock in 1983 a debate started on whether there was distinctly feminine or even feminist science. According to the much-debated storyline, for Evelyn Fox Keller, McClintock was an exceptional scientist devoted to a holistic conception of life, whose science of maize genetics was sensitive, dynamic, interactive, and flexible, inspired by her "feeling for the organism" (Keller 1983). Keller's biography highlighted the role of intuition, feeling, connectedness, relatedness—the qualities that are stereotypically identified as feminine—in the making of McClintock's science. Although, the key part of Keller's story of McClintock›s life was an account of how McClintock's unorthodox scientific views challenged the gene as master-molecule dogma of the time and how they isolated her from the mainstream science (Keller 1983). But many understood Keller's story of McClintock as demonstrating how women saw scientific objects with feelings and hence their science was holistic and hence radically different from the reductionism of male-dominated science. Keller's account eventually made McClintock a feminist icon with which started a debate if stereotypically feminine gender traits have role to play in transformation of mainstream science—which otherwise displays stereotypical masculine gender traits, and if the feminine traits could make science more humane and more in tune with nature (Keller 1987; Richards and Schuster 1989). The same question was also

posed differently: Do women indeed do science differently than men? If so, could there be distinctly feminine method? In these propositions were confused and conflated feminine with feminist, science as content and practice, science as product and process, science as epistemological methodological entity and its political cultural context, constitutive values internal to science and contextual values external to science. What is science after all?

In the history and philosophy of especially natural science, since the French philosopher Gaston Bachelard, methodological or epistemological field of science is understood to consist of a set of interlinked, mathematicized concepts which intra-act with instrumentalities or experimental hardware to form what Bachelard called phenomeno-technique. Many women scientists doing hard science were angry with the suggestion of distinctly feminine science because such science was then branded as soft science—as non-mathematical science. They felt that in the name of better science, they were asked to do inferior science, not-even-science. Another way of saying women can't do science. Taking this argument back to where it all started—Keller's biography of McClintock—Nathaniel Comfort (2011) did second biography of McClintock two decades after Keller's biography, in which he called Keller's story of McClintock as being rejected, ignored, not listened to, ridiculed by the mainstream male-dominated scientific community, a myth. Comfort says that McClintock herself created this myth which Keller took it at face value. Comfort then spends entire biography to separate fact from fiction, to dismantle the McClintock myth. I have written an entire chapter and a paper on reading Keller and Comfort's biographies together to have my own analysis of McClintock's subjectivity and her relation to scientific objects, and I do not intend to repeat my arguments here (Shah 2018, chapter 5; 2016), but what I want to highlight here is that Comfort repeatedly shows in his biography that McClintock's theories were rejected, not listened to, ignored by the wider scientific peers not because she was a woman or that she had a holistic views of her scientific objects that challenged the male-dominated dogma, but because her arguments were speculative, because she did not have adequate data to substantiate them. McClintock' science was bunch of ideas, empirically it was not-yet-science (Comfort 2001). Long story short, methodologically and epistemologically, especially when we are speaking of hard or physical sciences, there cannot be good or bad science—there is science or none-science. Science as moral or political category is then only a product. Science that built atom-bomb is always science, what it did to Hiroshima and Nagasaki is morally and politically bad science.

There are three distinct connotations of science in the play here: methodological or epistemological, political or moral, and cultural or historical. science and technology studies have spent considerable amount of ink and paper to show how they overlap or interact. How to open the black box of science and to show how it is historically, culturally, politically, morally constructed is the

mainstay of much of science and technology studies. The point is when societal impact of science can be easily categorized morally or politically as good or bad or ugly—methodologically and epistemologically, there is only science, in other words, there is either science or no-science, science that works and science that does not—successful and failed science, perhaps yes, but good and bad science, no. The pendulum of the mind alternates between sense and nonsense and not between right and wrong—said Carl Jung. So then science as sense-making the good science is what makes sense and bad science as nonsense science or no-science.

I know that my commentary here is partial to natural or hard science. What about social sciences or humanities? And I must admit that in writing this commentary I had days of agonizing agonism—I felt a knot in my stomach imagining all forms of knowledge conflated together in Schaefer's book. Schaefer clarifies that knowledge and knowledge-making is a master term that applies to formalized knowledge like math, science, history, philosophy, humanities, and all other forms of non-formalized knowledge like informal everyday knowledge about the world like a group text, or a reddit thread, or the incident of people-watching at café (Schaefer 2022, 4), at another place Schaefer treats all forms of knowledge produced in a chemistry laboratory, a noisy pub, an operating room, or a wheat field, on par (2022, 55). The chapters on conspiracy theory and racialized reason refer to various formalized and non-formalized knowledge-sources such as reddit thread, emails, pamphlets. I felt concerned about treating formalized science and everyday forms of knowledge within one frame. Is there a way to distinguish cultural prejudice expressed in a public or private space from institutionalized science driven by prejudice? I was paralyzed for days writing this commentary pondering upon this point. Something quite does not feel right, what is it?

In my comments for the 4S panel at the Honolulu conference (authors meets critics debating Schaefer's book), commenting exactly on this point when I described myself as an old-fashioned Popperian, all my fellow panelists' heads turned, and there was much discussion on how science cannot be so neatly separated from other forms of knowledge and how there is a continuum in the way formalized and non-formalized knowledge unfold and bleed into each other. Yes, I agree—rationally. And then, I feel uncanny. As a feminist this unease rises deep inside my stomach as a feeling of insecurity about mixing science with other forms of knowledge. I believe that institutional forms of science acquires it authority and legitimacy from its method, its verifiability and replicability (yes, by applying Popperian falsification principle), its aspiration for truth and objectivity—while I use each of these terms very carefully qualifying that their meanings and how they manifest in a particular scientific case in a particular historical context are highly contested, they still constitute the bedrock of scientific institutions which then allow the possibility that science

can be systematically contested, challenged and reformed, it can be called on to be responsible and accountable. Because I believe in power of ideas to change the world—and this is why I work within the university—to debate, engage with, contest, teach, research scientific ideas. Cultural prejudices of a group of people watching a soccer game at a café or a thinking as feeling expressed in a reddit thread cannot be so systematically challenged, made accountable, responsible. Without this distinction of science and non-science, I fear that all knowledge may collapse into merely opinions—personal or collective, does not matter. This is precisely the reason why historians of science are wary of inducting feelings into making of science—what Keller was referring to. This is why I interpret that Schaefer gets onto a defensive backfoot about other emotions and their relation with bad science. I want to maintain a boundary in understanding how thinking feels applies radically differently in the making of different knowledge categories. And my commentary refers to only hard sciences deriving from my own work on history of genetic science.

For me, the pertinent questions to ask are: What is science and how it is made by emotions? How does the psychic life of the scientist-subject relate to the mode of science and how do practices of science constitute emotional subjectivity? How do agonism of feelings shape internal content of science, what Longino calls constitutive values, epistemological (truth-seeking) part of science? How do we open the black box of science and show how agonism of emotions are not external forces or products of science-making but how they are incorporated in the method, in the choice of hypothesis, in the choice of value-based assumptions forming theories, in the production and analysis of empirical results, in the making of theory? How do the affective is not only the motivational force or the end result of doing science but profoundly constitutive of the making of the science itself? How can we unpack science on case by case basis to show how embodied emotions are incorporated in its making, so that they can be made visible opening up a possibility to challenge them if necessary?

In conclusion, it has been a pleasure to read and engage with Schaefer's book. I cannot have enough. I hope this is a provocation enough for Schaefer to start thinking feeling next book because then he already has a dedicated reader anxiously and joyously waiting to read it.

References

Comfort, Nathaniel. 2001. *The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control*. Cambridge, MA: Harvard University Press.

———. 2011. "When Your Sources Talk Back: Toward a Multimodal Approach to Scientific Biography." *Journal of History of Biology* 44:651–69.

Daston, Lorraine. 2008. "On Scientific Observation." *Isis* 99 (1): 97–110.

Fleck, Ludwik. (1935) 1981. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Keller, Evelyn Fox. 1983. *A Feeling for the Organism: The Life and Work of Barbara McClintock*. New York: W. H. Freeman and Company.

———. 1987. "The Gender/Science System: Or, Is Sex To Gender As Nature to Science?" *Hypatia* 2 (3): 37–49.

Longino, Helen E. 2001. "Can There Be a Feminist Science?" In *Women, Science, and Technology: A Reader in Feminist Science Studies*, edited by Mary Wyer, Mary Barbercheck, Donna Giesman, Hatice Ozturk, and Marta Wayne, 207–13. New York: Routledge.

Richards, Evelleen, and John Schuster. 1989. "The Feminine Method as Myth and Accounting Resource: A Challenge to Gender Studies and Social Studies of Science." *Social Studies of Science* 19 (4): 697–720.

Schaefer, Donovan O. 2022. *Wild Experiment: Feeling Science and Secularism after Darwin*. Durham, NC: Duke University Press.

Shah, Esha. 2013. "Rosalind Franklin and Her Science-in-the-Making: A Situated, Sexual and Existential Portrait." *Yearbook of Women's History/Jaarboek voor vrouwengeschiedenis* 33:127–46.

———. 2016. "A Tale of Two Biographies: The Myth and Truth of Barbara McClintock." *History and Philosophy of the Life Sciences* 38 (18): 1–12.

———. 2017. "Who Is the Scientist-Subject? A Critique of the Neo-Kantian Scientist-Subject in Lorraine Daston and Peter Galison's *Objectivity*." *Minerva* 55 (1): 117–38.

———. 2018. *Who Is the Scientist-Subject? Affective History of the Gene*. London: Routledge.

Shapin, Steven. 2012. "The Sciences of Subjectivity." *Social Studies of Science* 42 (2): 170–84.

Zygon®: Journal of RELIGION & SCIENCE

van Emmerik, Mari. 2024. "Changing Minds: Cogency Theory and Secular Bodies." *Zygon: Journal of Religion and Science* 59 (3): 839–49. DOI: https://doi.org/10.16995/zygon.16859

Open Library of Humanities

Changing Minds: Cogency Theory and Secular Bodies

Mari van Emmerik, Research Associate, Faculty of Divinity, University of Cambridge, Cambridge, UK, mave2@cam.ac.uk

This reflection on Donovan O. Schaefer's *Wild Experiment* focuses on one angle of the multidisciplinary network of ideas explored in his book: the way cogency theory pertains to the theoretical discussion in psychology and secular studies. The beauty of cogency theory is identified to lie in its capacity to bring together conversations in philosophy, psychology, religious studies, and secularism studies, among other fields.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

OPEN ACCESS

Introduction

I am grateful for the opportunity to join this forum and offer a reflection on Donovan Schaefer's latest book, *Wild Experiment: Feeling Science and Secularism After Darwin* (2022). I first met Schaefer in 2017 as a graduate student at Oxford, where I enjoyed the privilege of being taught by him in the Science and Religion program. The discussions we had about different ways of thinking about religion, science, and secularism in the light-filled rooms of Trinity College in many respects shaped my academic sensibility and the trajectory of my research. Reading *Wild Experiment* brought back the joyful feeling of diving into those expansive and illuminating exchanges.

Wild Experiment is a captivating book. Much like Schaefer's other works, it offers the reader an impressive kaleidoscope of ideas, weaving the diverse fields of science and technology studies, psychology, affect theory, and secular studies, to name a few, into a single capacious conversation saturated with crisp theoretical and practical insight. Schaefer's (2022, 3) concept of "click"— the intellectual affect that consists in "the subtle joy of pieces of information snapping together" best describes my experience of following the central argument of this book unfold and blossom.

Wild Experiment reads as a natural companion to Schaefer's first book, *Religious Affects: Animality, Evolution, and Power* (2015), in which he offers a layered critique of the interiority and immateriality of the Western cultural notion of religion defined as "always and only a way of thinking," arguing that religion is affective at its very core:

> What if religion is not only about language, books, or belief? In what ways is religion—for humans and other animals about the way things feel, the things we want, the way our bodies are guided through thickly textured, magnetised worlds? Or the way our bodies flow into relationships—loving or hostile—with other bodies? (Schaefer 2015, 29)

Although some of these themes are echoed in *Wild Experiment* (I particularly enjoyed Schaefer's discussion of William James's (2022, 47) understanding of religion, which "emerges not from arithmetic but from our deep-running affective currents" where "we start with the feeling, then surround it with elaborate architectures of belief"), the new monograph expands this argument beyond the subject of religion. Here, Schaefer (2022, 20) builds his case to demonstrate that "everything we do as bodies is affective": religion, secularism, science, and politics are all saturated with feeling. Schaefer (2022, 16–17) writes that "from babies playing with blocks to giddy proofreaders picking out mistakes in a text," the desire to know the world is the mechanism that propels thinking (an argument I eagerly attest to from "empirical" observation as I alternate writing this reflection with looking after my infant daughter who is currently in her "explorer" stage):

> Babies and their games with shapes, words, dolls, blocks, faces, gestures, and toys are already learning to revel in click . . . All animals and all humans are rational in the sense of desiring, considering, and more or less effectively arranging information about our environments. This is because all animals have intellectual feelings, in different combinations and in varying degrees. What gets called rationality is really a combination of intellectual passions nudging us to know the world.

This approach shapes the main arch of the book's argument, which is centered around Schaefer's (2022, 10) *cogency theory*, defined as "a collection of perspectives on how thinking is made by feeling." Schaefer (2022, 5) takes this thesis further than the thinkers associated with feminist science and technology studies scholarship have done in the past, arguing that "knowledge-making is not just entangled with feeling . . . but encompassed by it."

This argument, supported by several case studies, is developed in two key moves: first, Schaefer deconstructs "the ambient belief" that saturates the Western way of approaching thinking and feeling as separate categories. Subsequently, he builds an intricate bridge between secularism and affect, presenting his case for the reimagining of the secular not as "contraptions of pure reason" but as "alloys of emotion." Max Weber's regularly misconstrued concept of disenchantment (or "demagnification"), according to Schaefer (2022, 95), is not based on annihilation of feeling, but on its reconfiguration or rearrangement:

> The cogency theory approach to secularism suggests that disenchantment is not so much the eradication of feeling in our understanding of the world. It remixes ways of feeling and thinking into new affective forms.

As a scholar of nonreligion with a background in cognitive science, I find this framework of approaching the secular through the lens of reconfiguring emotion/affect most intriguing. The angle that interests me most is the mechanism behind this reconfiguration. In my work, I explore the cognitive aspects of nonreligion, focusing on the way different theories of mind contribute to the production of nonreligious experience. The two central issues behind the majority of cognitive approaches to the study of nonreligion are the disembodiment behind their conceptions of the mind and the subtractionist approach to defining nonreligion. I argue that cogency theory is perfectly placed to address both issues. Therefore, this response will focus on the themes in *Wild Experiment* that are closest to my heart by engaging with the way this book sets up the conversations about cognition and secularism (particularly in chapters three and four) and exploring what cogency theory contributes to the psychological study of nonreligion.

My reflections are rooted in the 4E (more appropriately described as 5E) cognition paradigm, which is an understanding of the mind as embodied, embedded, enactive, extended, and emotional/affective. I use this framework as I reflect on the cluster of approaches to nonreligion Schaefer calls "material secularism," the term that gives the nod to "material religion," thereby presenting secularism as substantive, embodied, and felt.

Embodied Minds: Thinking as Feeling

"Before any statement about the mind, is an assumption about the nature of the reality of which it is part, and which is in some degree accessible to it as experience or as knowledge. Whoever controls the definition of mind controls the definition of humanity itself, and culture, and history," writes American novelist Marilynne Robinson (2010, 31–32) in her remarkable collection of essays *Absence of Mind*. Science (and cognitive science is no exception) is never neutral: Schaefer (2022, 231) writes that "no part of science happens in a vacuum. It's engineered by living bodies, saturated by culture, power, and history, and enfolded within feeling." Science always works within a specific cultural paradigm, which leads to a particular understanding of the mind; this, in turn, leads to a particular understanding of the human experience.

Schaefer (2022, 14) builds on this foundational theme throughout the book, first by putting Thomas Kuhn and Michael Polanyi in an elaborate conversation to show that science always relies on "tacit knowledge, the repertoire of unspoken background coordinates—absorbed through experience rather than language—that powerfully shape scientific knowledge production." All theories, without exception, are created by embodied, culturally embedded, historically and physically situated human minds. But what, according to cognitive science, is the mind?

According to the traditional cognitivist (i.e., computationalism) approach, dominant in cognitive approaches to the study of nonreligion, the mind is a complex information processing system, and cognition is usually understood as a form of computation. Philosopher Andy Clark (2012, 276) describes the traditional cognitive science model, as "brainbound." According to this model, the body is "just the sensor system of the brain, while the rest of the world is just an arena in which adaptive problems get posed, and the brain-body system must sense and act" (Clark 2012, 276). This broad commitment to the computational theory of mind in the cognitive study of nonreligion leads to an almost exclusive focus on the study of mental phenomena like the presence or absence of belief.

An alternative to this "brainbound" model is an embodied or 4E approach that views cognition as inherently linked to the body and its environment. Clark aptly describes this model as "porous." According to this porous model, the processes that facilitate cognition involve inextricable entanglements of

"feedback, feedforward, and feed-around loops that promiscuously criss-cross the boundaries of brain, body, and world" (Clark 2012, 278). From this perspective, the faculties linked to sensing, thinking, and feeling are inextricably formed and informed by our body-based interactions with the surrounding world. According to these 4E theories, the body (from its chemical composition down to the gut bacteria) and environment (expressed through both physical and cultural niches) are an inextricable part of cognition and emotion.

By his own admission, Schaefer (2022, 6) employs a "light-touch approach" to the term "cognition," using it "as a catchall to include many kinds of thought like reasoning, knowing, recalling, discovering, and learning." Despite the conceptual breadth of this treatment, Schaefer's understanding of the mind seems to broadly align with the embodied cognition approach. Even though the term "cogency" looks as if it is etymologically rooted in the "cogito" of computationalism, Schaefer (2022, 9) explains that this is not the case:

> It comes from the Latin roots *co-*, meaning "together," and *agō*, meaning "drive" or "act." Its descendants are English words like agent or action. The related Greek root *agō* gives the further sense of a "guide" or "duct" and enters English in words like the agonism of dramatic action, or the pedagogy of guiding the young, or axiom—a thing found to be weighty. Cogency as *cō* + *agō* corrals all these meanings. This confluence and contest of forces is agonism, a struggle of different priorities. To say something is cogent is not the self-evident testimony of truth to an abstract intellect. It spotlights forces streaming together, creating a tangible *feeling of truth* measured by the body. These forces interact on the microregister of feeling. *Believing* means one of these struggling currents of feeling has prevailed. It has, for now, been found *cogent*.

One minor concern I have about Schaefer's discussion of the psychological grounding of cogency theory is the use of metaphors that on occasion seem to obscure the conceptual precision of the argument. For instance, Schaefer (2022, 234) writes that thinking "buds" from feeling, thus echoing the idea that goes back to the founder of experimental psychology Wilhelm Wundt, who held that feelings come before thinking, arguing that "the clear apperception of ideas in acts of cognition and recognition is always preceded by special feelings" (Wundt 1907, 244). However, the argument behind cogency theory as well as Schaefer's (2022, 8) engagement with the key discussions within psychology and neuroscience support the perspective that affect and cognition are one and should be conceptualized as "a seamless garment."

In the "Feeling is Believing" chapter, Schaefer sets up a dialogue between psychology and cognitive theory that achieves two important goals: first, it demonstrates that contemporary cognitive neuroscience supports cogency theory, and second, it challenges the outdated theoretical frameworks within

the field of psychology that stem from the tacit knowledge that understands reason and emotion as autonomous categories. Schaefer's methodology is rooted in Samantha Frost's biocultural approach, which allows him to build a much-needed bridge between the humanistic and scientific perspectives on the mind. A biocultural approach, according to Frost, pays attention to an ongoing polyphonous conversation between an organism, its habitat, other organisms, and their history of experience: it melts and remolds the flat familiar binaries (including the binary of thinking/feeling) into relationships, networks, and flows. This theoretical framework is consonant with the logic of the 4E approach as it provides a way to conceive of the body as essentially and irreducibly relational, extended, and historied. "An organism," writes Frost (2016, 123), "can be seen as a literal corporealization of a conjunction between its transgenerational carried history and the environment within which it currently lives."

The layer I would like to explore further is the role that culture and environment play in the shaping of this "seamless garment" of thinking/feeling. In his discussion of the making and remaking of belief, Schaefer (2022, 232) describes cogency as "the raw material of conviction":

> It's the agonism of compulsions in our bodies, the traces of the rambunctious passion for reason messing together with the sticky affects that surround it to make weird conglomerates of belief. It affirms that thinking and feeling are not just intertwined . . . It's about how we make up our minds by feeling our way forward, and how new injections of feeling—including the feelings carried by information, fact or fiction—remake belief.

If feeling (re)shapes belief, then what (re)shapes feeling? How does one explore this process and the mechanism behind it without reiterating the binary that cogency theory deconstructs? One potential avenue for exploring it within the biocultural framework is the concept of imaginaries. It has a rich (yet not unproblematic) history of engagement in the work of Charles Taylor. However, the approach I find more productive is found in works of Leila Dawney (2011) and Kathleen Lennon (2015), who reimagine this concept through the lens of affect, using imaginaries as a tool of subject formation to argue that bodies make sense of their engagement with the world through the production of affective imaginary associations tied to their histories and shaped by participation in their environment. It seems to me that this approach may avoid Schaefer's critique of Taylor's understanding of imaginaries as an on/off switch of emotion.

This contemporary application invites a new conception of imaginaries as material entities: "They are not just a backdrop to the world," argues Dawney (2011, 535), "imaginaries are produced by bodies through practices and technologies and constitute the way in which we experience the world." From this perspective, imaginaries are more like intuition or gut feeling than

a rigid mental representational framework. Dawney (2011, 542) argues that conceptualizing imaginaries as "a way of collective production of sense" offers an insight into "how differently historied bodies come to experience and engage with the world in different ways" and leads to the possibility of thinking about the role of imaginaries in constituting different bodies with different capacities: differently historied bodies produce different imaginaries, which in their turn may increase or decrease their capacities to act in accordance with the histories that have affected their imaginaries in the first place.

Kathleen Lennon's *Imagination and the Imaginary* (2015, 73) defines imaginaries as "the shape or form in terms of which we experience the world and ourselves; a gestalt which carries significance, affect, and normative force." Lennon (2015, 138) argues that they are simultaneously *disclosed* by the world and yet *created* by us. We do not project or force imaginaries on the world; rather, they emerge from our embedding in the environment.

Lennon's (2015, 2) conception of the imaginary should be approached "not as a domain of illusion posited in opposition to a 'real', but rather as *that by which* the real is made available to us." The histories through which bodies are produced take place within histories of both ideas and material relations and should be considered in terms of how they position and affect bodies always already embedded in the environment. It is through the body's material relationship with the world that these categories emerge and "back-form reality," using Brian Massumi's (2002, 8) vocabulary. One of the exciting contributions of cogency theory is the way it reframes the imaginary of the mind by reconfiguring the thinking/feeling binary and thereby offering a toolkit for a less sterile and more embodied and affective psychological study of nonreligion not limited to the study of belief and its absence.

The Making of Secular Bodies

Before I proceed to discuss the implications of cogency theory for the understanding of the secular, I pause here for a quick terminological detour. In the current theoretical landscape of the fields engaging with secularism and nonreligion, there are three main clusters of terms: first, the ones that stem from the term "religion" (e.g., nonreligion or irreligion); second, the ones with "theism" as their root (e.g., atheism or anti-theism); and third, the ones stemming from the term "secular" (e.g., secularity and secularism). In my work, I follow Lois Lee (2012, 131), who argues that the term "nonreligion" is the broadest as it covers the largest number of nones under this multidisciplinary umbrella. There are overlapping issues within the fields of nonreligion and secularism, and I approach them as sibling disciplines here as I focus on those similarities.

One of the biggest issues common to both of these fields is the negative attitude reflected in what Chris Cotter calls a *subtractionist approach,* rooted in a

problematic historical attitude that treats nonreligion-related phenomena either as a problem or as an "uninteresting residuum." According to this perspective, different manifestations of nonreligion are viewed as an anomaly, a statistical outlier, a left-over category, a comparison group, or just an afterthought (Cotter 2020, 55). The way this attitude is reflected in the traditional secularization theory, as well as most psychological studies of nonreligion, is linked to the reification of religion: from this perspective, religion-related phenomena are treated as something "charged" and substantial, while any nonreligion-related phenomena are relegated to the "neutral" space empty of religion. In *Recognising the Non-Religious*, Lee (2015, 50) shows how this treatment of nonreligion is also the dominant approach in social science, which manifests in "religion-centric methodologies, in which the secular is viewed as a context in which religion exists and is enacted."

Schaefer's (2022, 4) cogency theory challenges the subtractionist approach found in the traditional narrative of secularization summed up "as the slow but steady fade-out of religion" and that "depicts rationality floating above the world and guiding history, immune to the local, the particular, the bodily, and—especially—the emotional":

> Cogency theory shines a light on the secular not as the gleaming fortress of reason, but as a humming network of tastes, dispositions, and moods laying down the rhythm that enables our memories, ideas, concepts, and beliefs.

In his engagement with secularism, disenchantment, and critique, Schaefer (2022, 84) first challenges the historical narrative of disenchantment found in the classic accounts of Max Weber and Peter Berger followed by introducing the work of scholars in the field of secularism studies to show that secularity is not a "flat landscape" left over after religion has been removed but "a something," or rather, "somethings"— habits, cultures, dispositions, affects.

The tradition Schaefer (2023, 704) calls "a material secularism" in one of his recent papers observes that secularism comes along "not just with a set of discourses, paradigms, beliefs and disbeliefs, but with its own ways of being embodied." To further disassemble this logic of subtraction, Schaefer engages with anthropological work exploring the mechanisms of formation of the secular subject and secular habitus in the works of Talal Asad and his students, the late Saba Mahmood and Charles Hirschkind. This intellectual tradition draws on Michel Foucault's conception of historically constituted subjects, which implies the process of embodied cultivation and subjectivation of perceptual and affective capacities.

Where the classic secularization theory separates religious and secular-relegating religion to emotion and secularity to "neutral" rationality, cogency theory suggests that nonreligion and secularism are also shaped by embodied

affects. Schaefer argues that a closer look at the relationships between secularism, enchantment, and critique reveals the affective currents of reasoning. Together with Ann Pellegrini, Jane Bennett, Janet Jakobsen, Monique Scheer, and Nadia Fadil, among others, Schaefer explores the textures of secular feeling. If Taylor's story of secularization offers a flat approach to feeling where it is either there or not, Schaefer's cogency theory offers an exploration of its fullest spectrum. Therefore, cogency theory acts as a bridge between secularism studies and affect theory: Schaefer (2022, 84) argues that from this perspective, what gets called disenchantment is not the destruction of feeling but its reconfiguration.

It seems to me that Schaefer's discussion of secularism through the lens of cogency theory essentially provides a layered response to Charles Hirschkind's provocative question originally raised in his essay titled *Is There a Secular Body?* (2011). Hirschkind similarly positions his material inquiry along the lines of secular embodiment, with the question reflected in the title of his essay. This question gets further unpacked in his elaborate call for attention to the embodied nature of the secular as he asks whether there is "a particular configuration of the human sensorium—of sensibilities, affects, embodied dispositions—specific to secular subjects, and thus, constitutive of what we mean by secular society?" (Hirschkind 2011, 633). Hirschkind's essay generated a lot of interest in the notion of the secular body, however, he left his question open. My reading of Schaefer's (2022, 105–6) engagement with the secular through the lens of cogency theory suggests (and I welcome his correction if I am wrong) that his answer to this question is "yes":

> In all its forms, secularism washes over us, rewriting the parameters of how our bodies feel their way through the world. Thought—the play of ideas, concepts, reasons, and evidence—is emotionally alive, neither immune to the public domain nor detached from it. If formations of the secular have something to do with the reconfiguration of frames of knowledge, they have everything to do with changing how the world feels. That's disenchantment. Certain affective forms are nourished, cultivated, detailed, disciplined refined to a level of sculptural precision. Others are left to wither and fall. Josephson-Storm writes that "we have never been disenchanted." But really, the conclusion to draw is that disenchantment, for Weber, never meant emotional emptiness. It was a new channel of feeling-by-thinking.

Conclusion

The shimmering beauty of cogency theory lies in its capacity to bring together conversations in philosophy, psychology, religious studies, and secularism studies, among other fields. In this reflection, I have focused only on one angle of the multidisciplinary network of ideas explored in *Wild Experiment*: the way cogency

theory pertains to the theoretical discussion in psychology and secular studies. Its reach is much wider than that, as other contributors to this special issue will attest.

Wild Experiment provides a fresh contribution to the understanding of secularism and nonreligion by offering a more subtle yet stimulating intellectual history of the secular that considers the affective and embodied formations of both belief and unbelief. Schaefer's cogency theory and the material secularism approach are perfect candidates for mapping out new strategies for the cognitive approaches to nonreligion as they engage much more closely with the role of feeling, materiality, and culture in shaping secular bodies and their experiences, habits, orientations, and sensibilities.

In this reflection, I have argued that cogency theory contributes to a new biocultural paradigm that could enrich and broaden the cognitive understanding of nonreligion as a substantive object of research and would work well in conjunction with the 4E cognition paradigm. Schaefer's approach shows that what constitutes nonreligion and the secular extends beyond the bland absence of beliefs or emotions by providing a new map for rethinking cognition and nonreligion: "Whatever secularization and disenchantment are, they're not so much the eradication of feeling as the replacement of one template of feeling by others" (Schaefer 2022, 24).

References

Clark, Andy. 2012. "Embodied, Embedded, and Extended Cognition." In *The Cambridge Handbook of Cognitive Science*, edited by K. Frankish and Q. M. Ramsey, 275–91. Cambridge: Cambridge University Press.

Cotter, Christopher R. 2020. *The Critical Study of Non-Religion: Discourse, Identification and Locality*, *Bloomsbury Advances in Religious Studies*. London: Bloomsbury Academic.

Dawney, Leila. 2011. "Social Imaginaries and Therapeutic Self-Work: The Ethics of the Embodied Imagination." *The Sociological Review* 59 (3): 535–52.

Frost, Samantha. 2016. *Biocultural Creatures: Toward a New Theory of the Human*. Durham, NC: Duke University Press.

Hirschkind, Charles. 2011. "Is There a Secular Body?" *Cultural Anthropology* 26 (4): 633–47.

Lee, Lois. 2012. "Research Note: Talking about a Revolution: Terminology for the New Field of Non-Religion Studies." *Journal of Contemporary Religion* 27 (1):129–39.

———. 2015. *Recognizing the Non-Religious: Reimagining the Secular*. Oxford: Oxford University Press.

Lennon, Kathleen. 2015. *Imagination and the Imaginary*. London: Routledge.

Massumi, Brian. 2002. *Parables for the Virtual: Movement, Affect, Sensation*. Durham, NC: Duke University Press.

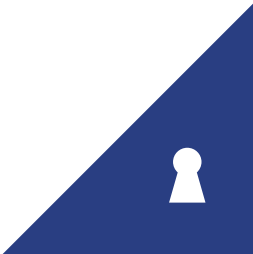
Robinson, Marilynne. 2010. *Absence of Mind: The Dispelling of Inwardness from the Modern Myth of the Self*, *Terry Lectures*. New Haven, CT: Yale University Press.

Schaefer, Donovan O. 2015. *Religious Affects: Animality, Evolution, and Power*. Durham, NC: Duke University Press.

———. 2022. *Wild Experiment: Feeling Science and Secularism after Darwin*. Durham, NC: Duke University Press.

———. 2023. "Voracious Secularism: Emotional Habitus and the Desire for Knowledge in Animal Experimentation." *Religion* 53 (4): 7007–23.

Wundt, Wilhelm. 1907. *Outlines of Psychology*. Leipzig: Wilhelm Englemann.



Zygon®: Journal of RELIGION & SCIENCE

Schaefer, Donovan O. 2024. ""They Made Up Quiet Songs": Classification, Feeling, and Technique in Science and Religion." *Zygon: Journal of Religion and Science* 59 (3): 850–70. DOI: https://doi.org/10.16995/zygon.17162

Open Library of Humanities

"They Made Up Quiet Songs": Classification, Feeling, and Technique in Science and Religion

Donovan O. Schaefer, Associate Professor of Religious Studies, University of Pennsylvania, Philadelphia, USA, doschaef@upenn.edu

In this rejoinder for the Book Symposium on *Wild Experiment*, I consider several themes that recurred across the five responses. These themes are: the way we classify and morally position science, religion, and secularity; the relationship between scientific rationality and its others; the roles of scientific emotions in scientific training; and the relationship between the cool or calm cognitive emotions and scientific practice.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

Receiving the Ludwik Fleck and International Society for Science and Religion prizes for *Wild Experiment* has been one of the greatest honors of my career. And I am humbled by the rich, detailed, and imaginative responses to the book offered by Mathew Arthur, Anne Pollock, Josh Reeves, Esha Shah, and Mari van Emmerik in this Book Symposium. This book was driven by my conviction that although plenty of work on the affective life of science existed, that research was happening in disparate corners of the academy. Those corners, I believed, could be brought together to help spark the conversation that science studies in all its forms—including science and religion—needs to be having about the nature of knowledge-production. So I am grateful for five conversation partners who have so expertly staged exactly the kinds of dialogue I hoped I would get to participate in after writing this book. It is a true pleasure to be allowed to explore new horizons in science and religion—as well as science studies more broadly—in such excellent company.

Several themes recurred across the five responses in this symposium, and I would like to reply to those themes in turn, rather than pivoting from one discussant to the next. (With apologies, this means that for lack of space I will not be able to engage every single agenda item raised in these fabulously interesting responses.) These themes are: the way we classify and morally position science, religion, and secularity; the relationship between scientific rationality and its others; the roles of scientific emotions in scientific training; and the relationship between the *cool* or *calm* cognitive emotions and scientific practice.

"The Great Advantage of Neutrality": Classifying and Moralizing Religion and Science

One of the recurring motifs of this dialogue was the question of how to classify—and evaluate—the categories of religion, science, and the secular. Coming from religious studies, it is widely acknowledged that the category *religion* needs to be studied at arm's length—not as an essential feature of reality but as a contingent classification. The ongoing historicization of the category of religion shows that the definition of religion is always a process rather than a consistent arrangement (McCutcheon 2003; Smith 2004; Masuzawa 2005, Nongbri 2013). Religion is deployed for different purposes, in different assemblages and with different elements, in different moments in history. Brent Nongbri (2013, 155) proposes that one of the defining questions of religious studies should be:

> What sorts of interests are involved in such decisions of defining religion? *Who* is doing the defining and *why*? In other words, a good focus for those who would study "religion" in the modern day is keeping a close eye on the *activity* of defining religion and the *act* of saying that some things are "religious" and others are not.

Science, too, has enjoyed its fair share of critical reframing in the past century. Paul Feyerabend's (1993, 1) assertion that "*the events, procedures and results that constitute the sciences have no common structure*; there are no elements that occur in every scientific investigation but are missing elsewhere" has widespread acceptance in science studies.[1]

I start to fret, then, when I read a passage like this one, by Mathew Arthur (Arthur 2024, 790–91):

> For both science and religion, change is at the mercy of histories already composed and in process. Both are guilty of methods that uproot situated stories and put them to work reproducing more of the same.

It is not that I disagree with the overall sentiment here: the things we call *religion* and *science* are definitely historical and definitely need to be met with a range of critical reactions. It is the syntax I wonder about. *Religion*—in the singular? *Science*—in the singular? As if either of these vast, polymorphous worlds could really be localized to a single domain? That worries me.

To be clear, though: I do not think Arthur believes that science is so monolithic as it sometimes comes across. His wonderful book *Everything Is a Lab* is introduced with the observation that "the bubbling-up of yeasty dough, the soft or hard coordination of lover's bodies, the same old house in freak weather or the glow of dusk light [is] each time a new technology" (Arthur 2023, ix). I cannot recommend this brief volume strongly enough—it is beautiful, bold, and hugely generative. We are on the same page, I think, that technology and science and their many offspring spill everywhere. But I want to use Arthur's piece to draw out some tendencies that I find operating in a number of domains in the humanities, and are refracted in some ways even in this symposium—in particular, a tendency to see science as a cohesive intellectual object that can be assessed as innocent or guilty. Science, I would say, is not localizable to one time and place, nor is it reducible to a single moral valence. It is woven into violence, exploitation, and oppression just as it is woven into peace, justice, and flourishing.

For decades, science and religion scholarship has pointed to a powerful consensus that the science/religion binary needs to be dissolved. As Peter Harrison (2015, 3f) writes, "the compartmentalization of modern Western culture that gave rise to these distinct notions 'science' and 'religion' resulted not from a rational or dispassionate consideration of how to divide cultural life along natural fracture lines, but to a significant degree has been to do with political power—broadly conceived—and the accidents of history." Studies of science and religion, at their best, offer a powerful rejoinder to the constant pressure of conventional wisdom—the basically modern ideology that sees science and religion as antonyms (Brooke 1991; Harrison 2015; McGrath 2019; Midgley 2002).

Some might say that as Euro-modern categories, *science* and *religion* can be aptly applied in those times and places—the contexts of emergence of the terms themselves. But that does not go far enough. Even in their initial moment of imagining, the categories of science and religion were not up to the task. What does it mean to split science and religion when the medieval European universities were mostly religious foundations? Or when Isaac Newton and Johannes Kepler were cross-cutting astronomy and scriptural commentary? Or when Darwinism was met with sincere embrace from some religious believers and devout rejection from some men of science? (Brooke 1991) The most important topos we can bring into any conversation about science and religion is, first and foremost, a rejection of the science/religion binary.

So, too, with the secular. Arthur writes that he "live[s] in one of the most secular cities in Canada," and then paints a textured portrait of Vancouver as "gridded with Teslas at EV charging stations, yogawear retailers, and lush greenspaces" while also struggling with homelessness, drug addiction, and stark economic inequality (Arthur 2024, 792). I was born in that same city in Canada, and I want to take friendly issue with Arthur's characterization of it as secular. Not because it is churchy—definitely not (see Block 2016; Todd 2022). But the lesson of secularism studies is that we cannot draw a firm line around *the secular* and invincibly wall it off from *the religious*. Mari van Emmerik (drawing on scholars like Christopher Cotter and Lois Lee) phrases this deftly in her response in this symposium. She challenges the "reification of religion" by which lights the secular is wanly defined as the blank tableau left over when religion withdraws: "from this perspective, religion-related phenomena are treated as something 'charged' and substantial, while any nonreligion-related phenomena are relegated to the 'neutral' space empty of religion" (van Emmerik 2024; see also Lee 2015).

Vancouver is an instructive case study for studying exactly why the category of *the secular* is so mercurial. So yes: a panorama of green-forward urban planning and gleaming condo towers and pop-up VC firms in skyscrapers and cryptocurrency vending machines in corner stores. But also: a city where colorful prayer flags hang in windows, Buddha statues smile from garden plots, a towering statue of Lord Krishna rises in the distance from Marine Drive, and psychedelic designs proliferate on posters and wall hangings; a city where a vivid, shapeshifting landscape of new religious movements, transnational cultural exchanges, and New Age efflorescence inspired me to study religion in the first place.

I think secularism studies goes even further in showing us that even things that don't even try to correspond to the historic labels of religious and secular end up blurring the lines between the two. So we could ask of planetary consciousness, psychedelica, or the yogafication of exercise: religious or secular? The churches may be in retreat, but all kinds of new fixations, new disciplines, new utopian projections, new articles of faith crop up in their wake. So too with that eminently Vancouverite construct: spiritual but not religious. The

clamor of conventional wisdom places spirituality as the binary opposite of religion—washed clean of power, absolved of institutionality, timeless rather than wedded to history. But as Linda Mercadante (2014, 235) writes, "rather than secularization or the 'death of God,' we may be seeing a shift in the location of the spiritual instinct." Secular people may begin by careening in the opposite direction of religion, but they often find themselves traveling full circle before slowly coming to rest right next to the place they left. This "secular paradox," as Joseph Blankholm (2022, 5) puts it, is "the weirdness of being secular."

I am taking this definitional detour because I think it has serious consequences for the way we imagine the relationship between science and power. I am exceptionally grateful to Arthur for sharing his familiarity with Indigenous STS—and Indigenous thought more broadly—in this space. One of the vibrant pleasures of reading his piece was learning with him from these approaches. His response, as I take it, is about challenging the innocence of science, about showing that science is not as neutral—epistemologically or morally—as it is sometimes held up to be: "The question remains," he asks, "of how to run with the claim that 'thinking feels' while accounting for science's bad side effects: lab rats, addictive pills, shitty bedside manner, racist algorithms, nuclear weapons, Elon Musk's plans to colonize Mars" (Arthur 2024, 797). So true, and Arthur is right that by centering the contrast with conspiracy theory, I may have painted science with an optimistic sheen in the first half of *Wild Experiment* (before the stark criticisms I make of race science, eugenics, and sociobiology in the second half). This was, I suppose, motivated by my concern over what I see as a pervasive reflex in some quarters of the humanities to swiftly lapse into moralism around science. It looks to me like the calcified residue of some kind of Weberian grammar, by which science is automatically bad, automatically guilty, automatically in hock to oppression, exploitation, colonialism, capital, and alienation.

There was more I wanted to say—about lab rats, about race science, about Silicon Valley sadists. It ended up on the cutting-room floor, as does so much with books that try to say too much. Some of it wound up in my later article "Voracious Secularism: Emotional Habitus and the Desire for Knowledge in Animal Experimentation," which expressly considered how the pleasure aspect of science is exactly what locks in violent dimensions of scientific inquiry (Schaefer 2023). That piece began with the work of two early-twentieth-century suffragettes, also anti-vivisection activists, who attended a lecture in London where a professor of medicine carved up a living dog on an operating table for the benefit of a room of students. But rather than the "serene dignity of science," the women described how "there is a spirit of jocularity prevailing, a loud conversation is going on, jokes and laughter everywhere" as the animal slowly died at the center of the theater of learning (Lind af Hageby and Sohartau 1903, 20f). The pleasure dimension I draw out of science through the epistemology of *click* was never intended to render science angelic.

So I am in full agreement that we cannot indulge in a portrayal of science as innocent. But I think in the process of making this point, Arthur goes too far toward isolating science and secularity as discrete intellectual artifacts with distinct moral valences. And I sense that part of Arthur's critique of *Wild Experiment* is exactly a concern for locating the innocence—or guilt—of science. He attributes the Vancouver landscape to "the 'great advantage of neutrality' . . . that evidence-based joins of science and democracy are said to provide" (Arthur 2024, 793). He docks me for working with "neuroscience and experimental psychology in order to prove that cognition includes emotion" (Arthur 2024, 794). He expresses concern over whether "science's stamp of approval [on animist cosmologies is] just another colonial gesture?" (Arthur 2024, 794). And he takes to task my assertion that it is important for humanities scholars to stay in touch with the natural sciences, noting:

> In a world already conscripted by technoscience with mortal stakes—life-saving drugs, agricultures of scale, labor automations, climate mitigations—this might hold true. But when science and the state are ontologically aligned, how are we to index or account for subversive forms of racialized reason embedded in "evidence-based" governance? How can we adjudicate feelings of truth in conspiracy versus, say, Indigenous lifeways that answer the question "who gets to be a person" (Wilkinson 2017) with rivers, rocks, and trees or Chinese medical syncretisms that cure depression with needles and bitter herbs that move invisible Qi? (Arthur 2024, 793)

To affirm the non-innocence of science, we also need to have a sense of strict bounding lines around *the* religion, *the* science, *the* secular. This authorizes the antithesis of technoscience-secular-state against traditional-religious-nonwestern. It dispenses with the lability of the categories—and renders them conspicuously well-suited to a tidy morality tale.

And this is where I want to offer a respectful rejoinder to Arthur's framing. We are not, it seems to me, "conscripted by technoscience." We are conscripted by capital. And capital (itself a highly polymorphous conglomerate of power-knowledge-affect, and maybe ultimately not a very useful term) will ruthlessly carve up whatever it wants, however it wants. Real conscription by technoscience would look very different. It would mean state actors taking climate change far, far more seriously than they do. Conscription by technoscience would mean the blanket rejection of policy-making that is oblivious to long histories of racialization—given the abundant scientific consensus that those histories matter in the ongoing unfolding of the present. Conscription by technoscience would mean sweeping transformations of food systems, energy, transit, health, and labor. Science's voice does not have the command that I think is being attributed to it here. And to forget the fluidity of science—and the pervasiveness

of science as something that belongs to everyone—seriously damages our capacity to name and respond to the real problems facing us.

Similarly, I am not convinced science is as invested in envisioning personhood in ways that are hostile to Indigenous or nonwestern perspectives. Arthur (2024, 795) writes that "the sciences have been hell-bent on pathologizing those who talk to trees and rocks." But in my view, that is just not what scientists tend to do in the daily work of science.[2] I know Galileo's (1957, 186) division of science and religion into *how the heavens go* and *how to go to heaven*—reaffirmed 400 years later by Stephen Jay Gould (1999, 6) in his non-overlapping magisteria paradigm—is not quite right. Facts and values, evidence and morality, cannot be so neatly compartmentalized. The interdigitations of what gets (and has been) called science and religion are too deep, too uncountable to cleave them into such tidy domains, even as a regulative ideal. But still: it is misguided to fetter science to a comprehensive view of personhood that I think few working scientists would claim as their own.

On the other side of the equation, I worry about the reflexive affiliation of, for instance, qi medicine and acupuncture with something inherently progressive or liberatory, some gentle fledgling in need of protection from the predations of science. This would seem to overlook the way that the Chinese state actively promotes the superiority of traditional Chinese medicine over so-called *western* medicine as part of an authoritarian, nationalist project (Yuan 2024). As a qi-based martial arts practitioner (another legacy of my not-quite-secular Vancouver youth), I have long been fascinated by the way the usually soft and squishy image of, say, Aikido quickly mutates when you start looking at the roots of the martial arts revival in pre-war Japan's escalating agenda of imperialism-cum-fascism. In my experience, qi/ki is deeply embedded in frames of power, violence, and hierarchy (Bodiford 2010; Mangan and Komegome 1999). This is where the rubber hits the road with trying to pin moral stakes on the science and religion binary. As an analytic lens, it obscures the far more sweeping, far more heterogeneous entanglements of science, religion, and secularity that define the transverse field of power-knowledge-affect. "Innocence," as Donna Harraway (1991, 157) writes, "has done enough damage."

Arthur (2024, 793–94) writes "[l]ike Western science, these knowledge practices have their own in-built agonisms or self-correcting checkpoints for appraising the 'felt weight of facts' (Schaefer 2022, 9)." The sudden slip from science to Western science in this passage is telling. It is exactly because I do not see science as an exclusively Western prerogative that I am not interested in a simplistic definition of science as univocally moral or immoral. I see science as our embodied orientation to understanding the worlds enfolding us. I see it as part and parcel of animal being-in-the-world, therefore entangled with many different forms of power and resistance, many different thoughts and actions, many different value systems. On some level, I see science as essentially neutral—precisely because it is so foundational to our existence that it is

indissociable from thought itself. My argument is just that we cannot understand knowledge-making processes without paying attention to feeling, and feeling has not been attended to nearly enough. We definitely cannot understand the relationship between science and power without looking into how feeling welds them together. The quickest way to misunderstand science is to isolate it as a morally simplex object. Guilt and innocence are the least interesting verdicts to offer.

I am looking at the frontispiece of my copy of Chanda Prescod-Weinstein's *The Disordered Cosmos* (2021), her memoir of life as a Black theoretical physicist. It is a sketch made by artist Shanequa Gay of her painting *We Were Always Scientists*, and it shows two Black women, working at a table, with liquids, glassware, and a mixing bowl in front of them, their hair wrapped in scarves, pieces of cloth hanging on a line behind them. One woman is holding two implements for liquid transfer, maybe a beaker and a dropper, and her face is the focal point of the image: her brow furrowed in deep concentration, that unmistakeable expression of the feeling of doing science.[3] And I take this to be one of the themes of the book, that science does not just belong to Euro-modernity, to men, to the wealthy, to the formally educated. Science—the careful study and refinement of our interactions with our environment—must be fully depedestaled, fully democratized, fully realized as part of how we all interact with our worlds.

The Purity of Science

Esha Shah's response comes at this same question from the other direction. Like Arthur, she wants to insist on a distinct profile of science as a discrete category. Unlike Arthur, she holds to this as a quality that defines the virtue of science: its capacity to hold itself above the fray of prejudice-laden jingoism, kneejerk conspiracism, and every variety of social media sludge. So I will reassert, as well, my basic argument for why science needs to be depedestaled. All modalities of knowledge-production are formalized to varying extents; science (as an ideal type) tunes itself as best it can to a more formalized process, matching, knowingly or not, the emotional blueprint I have laid out here—and doing so through the whole ensemble of training procedures scientists use to make more scientists. But I would add that formalized processes are happening everywhere. Every conversation, every exchange, every ad hoc process of trial and error is a study, a probing, a refinement. Even conspiracists debate better and worse ways to weave exhilarating delusions about the way the world works. And even the best scientists can succumb to groupthink, prejudice, lapses in judgment. Everything, as Arthur (2023) says elsewhere, is a lab.

I agree with Shah that the processes of institutional science make it more amenable to various forms of counter, dispute, and improvement. That is why we can resoundingly endorse the climate change consensus, or the validity of vaccines, or the literature on the damaging effects of racial discrimination, or a

previous generation of scientists' findings about acid rain, second-hand smoke, and ozone-depleting aerosols (Oreskes and Conway 2010). But I do not quite buy that formalized, institutionalized science and non-science are reducible to the binary between open and prejudiced. That does not take our everyday modalities of learning about the world seriously enough. And it risks draining the reservoir of humility that science needs to keep on hand at all times to maintain exactly the intellectual liquidity—"the possibility that science can be systematically contested, challenged and reformed, [so] it can be called on to be responsible and accountable" (Shah 2024, 837)—that Shah and I both see as indispensable.

This is also where Shah injects an important line of inquiry about contamination in the book. She notes—rightly—that this term activates an implication that science "was pure and then it got mixed up with bad stuff" (Shah 2024, 831). I remember tapping out the word *contamination* and knowing it was going to get me into hot water at some point, but I stuck with it. And I guess that is because somewhere in the back of my mind, I am holding to the prospect of a regulative ideal of science—a vanishing point on the horizon, a model for doing science that most effectively integrates and synthesizes the many intellectual affects that enfold the scientific knowledge-making process. (Click is one such emotion, but by no means the only one.) This is what I call *the sense of science*—a dynamic of intellectual affects.

So contamination does not mean *emotion gets into scientific rationality and messes it up*. It means *the emotional chemistry of the sense of science gets thrown out of whack and the wrong emotions are put in control*. All kinds of scientific malpractice—from bio-racism (Fields and Fields 2014, 4) to research fraud to academic bullying to groupthink to *p*-hacking—can be lined up with this. The wrong things start to matter—wrong from the stance of the regulative ideal of science—and I do not think it is wrong to call that contamination, as long as we are clear that *emotion gets into science* is not the mode of contamination we are talking about. Right or wrong, science is emotional.

Reeves, too, sounds off on this topic, noting that "*Wild Experiment* attributes a primary role to pleasure" via *click* and asking: "is the pursuit of pleasure inherently selfish, tied to personal prestige and status? . . . Is there a convergence between what is pleasurable and what is good, or is the pursuit of science always an extension of self-interest?" (Reeves 2024, 821). It is a compelling point. Fundamentally, I see pleasure as multiple, and science as oriented toward the pursuit of many different pleasures, in different degrees and kinds. Most importantly, scientific endeavor is not just about a pleasure we might take from beating out scientific rivals, getting a promotion, or earning a living. There is pleasure inside the very practice of working with ideas. And that pleasure messes together with all the other pleasures that go along with the scientific life. There are ways of training scientific intuitions—cultivating pleasures, what Monique Scheer (2012) calls *emotional habitus*—that really do matter for the doing of science.

That is a regulative ideal, though—a Weberian ideal type—not a real scientist. No real scientist is an exact compound of that formula. So I agree entirely with Shah's (2024, 833) rejoinder: the knowledge-seeker is "a self that is not only a fundamentally feeling, suffering, experiencing affective self but most importantly, this self is incoherent, contradictory, heterogeneous, split between the conscious and unconscious, both fictional and real at the same time." When Darwin was pre-empted by Alfred Russel Wallace in publishing the theory of evolution by natural selection, he immediately launched a conspiracy with his friends at the Linnean Society to get his own version on the record before Wallace's priority could be established. "It is miserable in me to care at all about priority," he wrote, shamefacedly, to one of his associates at the time (Darwin 1858; cf. Schaefer 2022, 144). But that did not stop him.

That kind of scientific ambition is indissociable from academic life, I believe. But it is also liable to contaminate the regulative ideal—the virtual horizon point on the highway that we will never reach, but that we can all benefit from holding in front of our field of vision—of the sense of science. I think Shah and I are on the same page here. Science is feeling all the way down; we hold the same bow-string, drawn all the way back.

Scientific Emotions, Social Emotions

While Shah insists that I am spreading science around too much, Reeves argues that I am not allowing it enough breathing room, suggesting that I see conspiracists and scientists as existing in two tidily discrete containers. I am going to push back on this characterization of my argument, though. It is precisely because I do not think scientists and conspiracists can be so easily siloed that I talk about them together. Reeves (2024, 816) writes that "scientists, like anyone else, are susceptible to confirmation bias." Definitely. I would never say that scientists have perfected the schema of emotional prioritization of *science* as an ideal type. I cite feminist critics who have dubbed the tendency of scientists to confirm their own ideological priorities (or to find answers that consolidate their careers) the Mulder effect to help make this point (Schaefer 2024, 8; see also Jordan-Young and Karzakis 2019). I wrote in Chapter 1: "Just as conspiracy theories can never really be disproven (and sometimes contain flashes of truth), there is no mechanism within science for manufacturing certainty, only a carefully constructed arena for staging the contest of forces" (Schaefer 2022, 54). But I also think that intellectual tendencies defining the two camps (such as they are), help to explain the real divergence between them, and I think those intellectual tendencies are best characterized in terms of feeling.

On the other side of the table, Reeves wants to attribute a science-like epistemic vigilance to non-scientists. Here too we are in complete agreement. But I want to stress that epistemic vigilance is not just an on-off switch. It is a complex set of intellectual tendencies—tendencies which can be cultivated through scientific training. And the best way to understand epistemic vigilance,

I think, is by thinking of it as fully enmeshed in emotion—not separate from it. This is where I want to try to reconnect with exactly what I think Reeves is sketching out. Reeves's picture of science is my understanding of science but with a different explanatory frame. For instance, Reeves (2024, 817) writes that ""Tribal rationality,' from this perspective, is not an illogical tendency to believe whatever your superiors tell you but is rather a useful shortcut that relieves one of the cognitive burdens of having to assess someone's trustworthiness." But what does that mean—relieving the cognitive burden of trustworthiness? What is a cognitive burden? My argument is that what we are really talking about here is an affective modality, an intransigent, deeply felt sense of frustration that comes along with having to uproot an existing way of thinking.

This, in other words, is the affective topography of rationality itself. Reasoning is not a computational system that automatically pathfinds a maximally efficient route. It is guided by feeling. We find easily navigable terrain here—ideas that click together effortlessly, right or wrong—and rough patches here—ideas that challenge, disrupt, frustrate, agonize—as we fumble our way forward. And that topography of resistances and openings is shaped by the emotional infrastructure of rationality. Reeves (2024, 817) actually reaffirms this in his next sentence, noting that "Many conspiracy theorists are driven not by the allure of 'click,' but by the desire to avoid work, which, like physical labor, can be unpleasant." I agree! What we are talking about is exactly a reframing of knowledge-making as a dynamic of desire and pleasure—and even pleasures. The point is that the arena in which this agonism of forces plays out is affective, rather than computational.

Reeves's (2024, 818) model is connected to the sociological account of science as optimally understood in terms of its sociality: "Given the limits individuals encounter in seeking truth," he writes, "we in modern societies solve the problem of distinguishing good from bad information by forming institutions and collectively practicing intellectual vigilance through the structured contestation and deliberation of different points of view." I do not disagree with this exactly. But I think it is an insufficiently detailed account. The bundle of interpersonal dynamics we collectively identify as the sociality of science is more precisely defined as an ensemble of affective processes.

Reeves (2024, 818), for instance, writes that "Scholars are motivated to expose subpar scholarship and to remain open to innovative and potentially groundbreaking theories, understanding that overlooking such theories could place them at a competitive disadvantage." Sure. But why are some people motivated by concern over "competitive disadvantage"? Why do some scientists want to join the herd? And why do others stick their necks out for a hunch? The nature of that motivation is a conglomerate of social emotions: shame and competitiveness, most prominently. To point to sociality as the explanatory terminus for the motivational structure of science is, in my view,

to step backwards, from a more unpacked explanation (social affects) to a less unpacked one (sociality, as such).

Nor is that the only thing that motivates scholars. This is a feature of my argument that I think Reeves has ignored: click is its own motivating force, its own locus of pleasure. Reeves (2024, 818) contends that "mutual oversight ensures that only competent scholarship is advanced, thereby enhancing the discipline's social authority." We know perfectly well that the history of scholarship is not a history of perfectly lineal forward motion. Click is both a powerful driver to better understand the world—shattering orthodoxies as new data grates on existing paradigms—and a profoundly conservative force, as the dominant paradigms demand fealty, locking in affectively charged patterns of interpretation that make new ideas impossible. Besides, the social emotions also exert their own retrograde vectors of force—as rivalries, jealousies, and ambition drive scholars to falsify or misinterpret their own work, consciously or otherwise.

And then there is science in solitude. Click is an exorbitantly powerful emotion. It can be coassembled with all kinds of social affects. But it can also be experienced on a desert island. Reeves (2024) suggests that there are far more scientific emotions in play than just click, a sentiment I think tracks with my definition of the *sense of science* as "a permanent struggle between the excitement of click and countervailing pressures—like fear or shame—about getting things wrong" (Schaefer 2022, 23f). There are many affective strands braided together in the sense of science—and the social emotions are vital elements of that braid. But not the whole story.

Anne Pollock makes a rhyming point with her request to return to the category of the erotic in conceptualizing science. For her, the erotic is another trope of sociality or relationality: a real connection rather than the predatory domination of some scientific practice. She introduces Evelyn Fox Keller's sense of the word; for her "the erotic is figured as a longing for a more relational and reciprocal way of knowing the universe than masculine science is structured to pursue—not 'having one's way with' the natural world, but communing in a profound way that might blur the boundaries of the knowing subject and leave us transformed by the encounter" (Pollock 2024, 805).

I take Pollock's point that the category of the erotic can be pitched in different ways, and I am all in favor of the normative injunction to produce a science that is less violent, less like domination, more like making love. Still, it is exactly the ambivalence of the erotic in the work of Black feminists like Sharon Patricia Holland (2012) that I find so constructive. Ambivalent in that the erotic contains pleasure and opens onto the forms of reciprocity that Pollock is right to remind us should always be squarely in our field of vision; but also points the way to the many forms of violence that are hinged to desire, and that pleasure itself makes possible (Schaefer 2015, 123). The desire to know has many domains of

violence (to return to Arthur's point above), and I want to keep those sequelae in the same frame as our desire to know in non-violent ways. That is my attempt at methodological symmetry.

I would say that sociality is not enough. And relationality is not enough. These explananda point to better approaches than the old positivist portrait of an individual genius staring in solitude at the world. But the approach I am suggesting wades even deeper into the complexity of sociality. Whereas the social sciences often see the dictum *it's social!* as a productive analytic terminus, I am arguing that it is only a move in the right direction. Sociality needs to be unpacked into all kinds of different processes that are better understood as affective: the pleasure of working in a well-tuned collaboration; the shame of knowingly publishing something that can be disproven; the sting of rivalry driving one to spend more hours at one's computer. Scientific emotions include, but are not limited to, social emotions.

van Emmerik phrases this in terms of an interesting question—one that I know I punted on in the book: the vectorial relationship between thinking and feeling. van Emmerik asks why I stress that thinking "buds" from feeling and notes the resonance of this concept with the early experimental psychology of Wilhelm Wundt. She goes on to suggest that a more consistent application of cogency theory would be to see thinking and feeling as fully interdigitated, with no firm priority awarded to either dimension. Put another way, she asks: "If feeling (re)shapes belief, then what (re)shapes feeling?" (van Emmerik 2024, 844).

van Emmerik is right that I do not offer a satisfying answer to this in the book, nor do I really have one now. My intuition is that there needs to be some kind of deep interrelatedness in the way we imagine the relationship between thinking and feeling. Ideas shape how we feel about things, just as feelings shape our mesh of cogency: what we find believable or preposterous. And yet, I cannot get away from the hunch that there is still some kind of priority to be located on the affective side of the equation. Maybe one way to say this would be that the affective dimension is better at creating conditions in which belief becomes possible or impossible than new information is at reconfiguring our emotional make-up. I think, in other words, that the substrate of our psyche is better understood as affective than as intellectual, and even to say it is simultaneously affective and intellectual is not quite right. Or maybe the category of *thinking* is too narrow, and the category of *feeling* just captures much more of the spectrum of thought. But that's just a hunch.

Shah offers another way of coming at this question, which is to bring in the dynamic of the conscious and the unconscious, and in particular the dynamic of *jizz*—a rapid intuition about an object of knowledge—theorized by various STS/HPS scholars. "It all starts with the confused notes," she writes, "which is followed by hummed and inaudible tunes gradually turning into a melody" (Shah 2024, 828). This is similar to the conclusion I'm fumbling towards—that we need to think of the affective as more foundational than the cognitive.

Nonetheless, I want to push back on part of how Shah (2024, 828) frames this, which is to emphasize that jizz indexes "the unconscious fundamentally structuring the conscious reasoning without the conscious even knowing it." I have come to believe that the binary of conscious and unconscious—whether that is coming from psychoanalysis or somewhere else—is fundamentally unhelpful. It suggests a thick curtain encircling some areas of experience, rendering them unconscious, and a floodlit chamber containing the rest—the conscious. It is tempting to correlate unconscious:emotion::conscious:cognition. But I think affect theory, at its best, swiftly dissolves this too-tidy grid. Many emotions can be mustered to conscious awareness, just as many thoughts can be forever foreclosed from our field of attention.

Science is not all or nothing. And although training can refine and improve it, it does not actually require us to be told that it must be done just so. This is because it is built into our basic emotional make-up. An emotional make-up that is fully epistemological just as it's fully affective. As van Emmerik (2024, 840–41) notes in her response, people who actually spend time raising children are confronted with this daily. I would say the same about paying attention to animals. Polanyi (Polanyi 1962, 140) is way ahead of us on this, as usual, in that he pays attention to both the ontogenetic (found among infants) and phylogenetic (found among nonhuman animals) precursors to scientific knowledge-production, "the inarticulate levels of intelligence of the animal and the infant, in which the personal coefficient of spoken knowledge is primordially preformed." In other words, we can come to better grasp what is at stake in the depedestaling of science by studying our kin. Science—feeling our way along—is part of what it means to be animal.

Varieties of Scientific Feeling

Reeves and Pollock are onto something, I think, when they note that this emotional pluralism of the sense of science can also be productively sized up from the vantage point of the question of hot and cold passions. I want to open a more detailed dialogue with them about what, exactly, we mean by these terms—or what Hume calls violent and calm passions—and how we might best operationalize this framework in theorizing science.

I am excited to return to this topic, because I think *calm passions* is both an eminently necessary concept and a mysterious one. It is a kind of contradiction-in-terms, right? Or an oxymoron—literally a *sharp-dull*—words at war, a thing that shouldn't be, passion that is dispassionate. And I think, on some register, it's also central to the puzzle of why emotions are so hard to understand, how it is that the rigidity of our language betrays us, interring some of the most powerful forces in our lives in shadow. It is the everyday topography of what matters and what does not matter, defined by calm passions, that does the vast majority of the work of conducting power—not the punctal moments of grief, fury, or elation. And I would say the same about science.

Reeves (2024, 813) writes that

> [h]ot emotions, like anger, fear, and love, move us to direct action because they are pleasurable. Cold emotions, which include the fear or shame associated with potential errors or deceit, also motivate action but on slower time scale. For instance, an individual grappling with addiction might opt for decisions that favor their long-term emotional health, despite potential short-term discomfort.

Hume's take on this (again, he's talking in terms of violent and calm passions) is somewhat different, though. He writes that the first criterion for distinguishing the varieties is not so much type of emotion, but intensity. But type seems to matter, too. So he proposes

> there are certain calm desires and tendencies, which, tho' they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation. These desires are of two kinds; either certain instincts originally implanted in our natures, such as benevolence and resentment, the love of life, and kindness to children; or the general appetite to good, and aversion to evil, consider'd merely as such. When any of these passions are calm, and cause no disorder in the soul, they are very readily taken for the determinations of reason, and are suppos'd to proceed from the same faculty, with that, which judges of truth and falsehood. (Hume 1960, 417)

This is the level of complexity Hume wants to bring us to: it is primarily about degree of intensity; secondarily about type of feeling. This is why Hume will affirm that reason is fundamentally an affective process, but because it primarily traffics in calm passions, we overlook the fact that it is, in fact, emotionally constituted: its "tranquility leads us into a mistake concerning them, and causes us to regard them as conclusions only of our intellectual faculties" (Hume 1960, 437). So the typology offered by Hume is not so much about cognitive and non-cognitive passions: it is a portrait of stronger, more palpable affects and subtler, often-unnoticed affects, all of which have cognitive consequences. And the latter—because they do not intrude on our awareness in the same way as their more garish counterparts—often do not get recognized as passions at all. That is partly why we have the thinking/feeling binary in the first place.

That is about as far as Hume gets. And when I read Hume on this, I hear him working something out—something he is not sure of, something our language and our philosophical inheritance has left in the murk. But I think Reeves adds something to the picture, which I would also want to draw out: time. I suspect that one of the ways of understanding the divergence between calm and violent

passions in Hume—and maybe the cold/hot metaphorics, as well—is to think about them in terms of timescales. Different cognitive affects may be latching on to different degrees of futurity and pastness.

Take the Stanford Marshmallow Experiment, in which children's ability to delay gratification was measured with a simple experimental set-up: they were offered a small treat (marshmallows, cookies, and pretzels were popular options) immediately or more treats if they would wait some period of time (Mischel, Ebbesen, and Zeiss 1972). The framing of this in terms of delayed gratification makes it sound like this is about putting emotions on pause. But that misses the deeply affective nature of the strategies used by the children to stay focused on the futural reward. As the experimenters wrote:

> They made up quiet songs ("Oh this is your land in Redwood City"), hid their heads in their arms, pounded the floor with their feet, fiddled playfully and teasingly with the signal bell, verbalized the contingency ("If I stop now I get _____, but if I wait I get _____)," prayed to the ceiling, and so on. (Mischel, Ebbesen, and Zeiss 1972, 215)

In other words, rather than avoiding emotion, the children seem to have spontaneously improvised a range of techniques to train their emotions on a futural object rather than a proximal one. Calm passions, futural passions.

Reeves, then, is onto something in his framing of calm passions in terms of timescales. The phenomenologist Edmund Husserl's vocabulary of *protentions* and *retentions* may be particularly valuable here. For Husserl, our experience in any given moment is actually a layering of three different processes: past-consciousness (retentions); original impressions of things experienced in the moment, and anticipations of future moments (protentions) (Beyer 2022). Husserl's conception of this is focused on a narrow band of recent and imminent events. But really, our field of awareness is fashioned from a sedimentation of all kinds of retentions and protentions in any given moment—at many different scales. Distant memories and far-reaching anticipations are melded together in every instant of our experience.

Most importantly for our purposes, all of these latches to different timescales are affective. As Michelle Maiese writes "anticipation of the future, or what phenomenologists call 'protention,' always involves motivation, an affective tone, and readiness for intentional movement…. This flow of intentional movement and lived sensations is driven by *bodily feelings of caring*" (Maiese 2016, 33). The way we experience the world around us, then, is actually a dense stack of superimpositions—many retentions, many protentions—all catching different moments, different memories and anticipations, all affectively charged. The cumulative sum of all those charges is, in essence, our mood, "*an affective lens*, affecting how we are affected" (Ahmed 2014, 14).

These time-bound affective figments—protentions and retentions—are trainable. The children staring down a cookie—and hoping it will become two—are seen praying, squirming, and making up songs to try to stay focused on a future moment. That reflects, I would say, both past emotional conditioning and an effort to self-train in the moment. It is about creating affective patterns of attention that will draw corresponding actions toward futural outcomes.

Reeves suggests that Kuhn's theory of scientific training can be grafted on to the idea of the sense of science—the unfurling agonism of scientific emotions—in *Wild Experiment*. Kuhn, he writes, asserts that "scientists have learned to use their cold emotions—a desire to avoid error and the shame that comes from mistakes—in their scientific practice" (Reeves 2024, 819). I agree on one level: I definitely think Kuhn's theory of scientific training can be adapted to this model. Kuhn himself, however, as I showed in *Wild Experiment*'s Introduction, tends to be deeply disdainful of any explanations of the scientific process that are routed through emotion. This, too, is an account of the demarcation problem—how to separate scientists from nonscientist. Like Reeves, I would say that "[a]lthough a natural curiosity is common to humanity, the mental discipline to pursue learning is not universal. Learning presents a challenge because it initially involves absorbing information that requires effort and does not immediately gratify" (Reeves 2024, 820). That training is an affective training—and it has everything to do with how we encounter the different layers of retention and protention that make up our quotidian experience.

We can see how this plays out in Reeves's new (and very helpful) typology of conspiracy theorists, which I will abbreviate to the cheerful, fearful, and social varieties. The project of detailing the plural attractions of conspiracism in all their emotional particularities is exactly the conversation I think we need to be having right now to truly grasp what conspiracism is and how it manages to snare us so effectively. Reeves contends, though, that *Wild Experiment* only really offers explanatory tools for the first variety, and so misses some of the deeper entanglements of conspiracism and religion.

But I think there are ways of connecting cogency theory to all of these types. Reeves (2024, 814) suggests that for fearful conspiracists, conspiracy theories "are often not enjoyable and cause spiritual trauma that comes from living with a pervasive sense of danger, mistrust, and paranoia." For my part, I am not sure it is so easy to divide pleasure and pain—even pleasure and trauma. One of the most profound insights to come out of affect theory is Lauren Berlant's startling idea of cruel optimism (Berlant 2011). For Berlant, we can be attached—affectively—to worlds, objects, and other people that are actively harming us. And the harm can be part of what renders the attachment so intransigent. Conspiracism is, for many, I suspect, a similar kind of exchange. You are caught in these very tight cycles of anticipation and dread, causing you to keep holding

a possibility in front of you to seize it, control it, keep it pinned by your attention so it stays in your sight. And that brings its own sense of power and thrill. They are kernels of white-hot passion, a fusion core of pleasure and pain.

Even Reeves's third type of conspiracist—the social conspiracy theorists who live in "skeptical information environments," such as certain religious communities, and have absorbed the conspiracist habits of thinking of their milieu by rote—can be encountered in terms of cogency. Reeves (2024, 815) asks: "do humans accept all beliefs because of emotional investment, or are some beliefs accepted 'secondhand' from trusted figures from one's communities?" and responds that many who come to conspiracism through religious communities "hold many beliefs that are emotionally inert." By way of examples, Reeves (2024, 815) offers his own convictions that

> water is composed of one water and two hydrogen atoms or that my mother was born in the state of Alabama. These beliefs are accepted without question until a sense of discomfort suggests that it might call for closer examination. Emotion, in this context, acts more as a detector for questionable beliefs rather than as a source of pleasure, but it takes too much effort to assess all our new beliefs. As two psychologists have recently argued, susceptibility to partisan fake news is better explained by lack of reasoning (i.e., laziness) than by motivated reasoning (i.e., a dopamine hit of pleasure).

My rejoinder to this is to ask, simply, why we have so much confidence that these processes are affectless. As with Reeves's points about the reducibility of scientific emotions to sociality I notice that Reeves's vocabulary inadvertently glides into the affective dimensions of the processes he is describing—the "sense of discomfort" when confronted with new beliefs, for instance. What's the emotional make-up of "lack of reasoning (i.e., laziness)"? What's the emotional make-up of "detect[ing] questionable beliefs"? This is where cogency theory calls us to fundamentally rethink the way we model emotion and cognition in the first instance. It is not just about feelings at the macro register, the emotions that wash over us, that we give names to. Our cognitive emotions define our encounter with every word, fact, and idea at the micro level, running over the granular texture of information and defining our responses.

Conclusion

Shah begins her response with a moving passage from a personal letter sent to her by the now passed Evelyn Fox Keller. Keller was responding to the publication of Shah's book *Who Is the Scientist-Subject? Affective History of the Gene* (2018), a brilliant study of the links between scientific knowledge-production and emotional dynamics. But Keller's letter is posed in a mood of tragic

resignation—a sense of defeat inculcated by decades of constant reminder that the whole apparatus of the humanities has been tuned to see feeling/affect/emotion as trivial, irrelevant—even embarrassing.

The letter's poignant reflection on the long tyranny of the thinking-feeling binary made me think of this passage in David Hume's *Treatise of Human Nature*:

> Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and to assert that men are only so far virtuous as they conform themselves to its dictates. Every rational creature, 'tis said, is oblig'd to regulate his actions by reason; and if any other motive or principle challenge the direction of his conduct, he ought to oppose it, 'till it be entirely subdu'd, or at least brought to a conformity with that superior principle. (Hume 1960, 413)

This paragraph—written almost 300 years ago—was part of a wider trend in eighteenth-century thinking toward taking emotion seriously—including in profoundly rethinking the relationship between thinking and feeling (Riskin 2002; Sullivan 2020). But three centuries later, we are still stuck on that same cusp. Western philosophy has, for three centuries, struggled with the riptide of the thinking-feeling binary. My hope is that the voices and counter-voices running through this Book Symposium point beyond this trap, underwriting new conversations that will help fashion a clearer picture of the landscape of how we think, believe, and make science in the world.

Notes

[1] Sergio Sismondo's *An Introduction to Science and Technology Studies*, for instance, takes it as a given that the anti-formalist approach of Thomas Kuhn is foundational to contemporary STS (Sismondo 2010, 12).

[2] With some conspicuous exceptions. See, e.g., Gazzaniga (2005) and Harris (2010), the latter particularly well known for the controversy it sparked (see Pigliucci 2013).

[3] See Darwin's discussion of the facial expressions corresponding to reflection in *Expression of the Emotions in Man and Animals* (Darwin 2009, 204–9).

References

Ahmed, Sara. 2014. "Not in the Mood." *New Formations* 82:13–28.

Arthur, Mathew, ed. 2023. *Everything Is a Lab: Doing Ordinary Science*. Lancaster, PA: Imbricate! Press.

———. 2024. "STS and Affect in the Wild." *Zygon: Journal of Religion and Science* 59 (3): 788–99.

Berlant, Lauren. 2011. *Cruel Optimism*. Durham, NC: Duke University Press.

Beyer, Christian. 2022. "Edmund Husserl." In *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), edited by Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/archives/win2022/entries/husserl/.

Blankholm, Joseph. 2022. *The Secular Paradox: On the Religiosity of the Not Religious*. New York: New York University Press.

Block, Tina. 2016. *The Secular Northwest: Religion and Irreligion in Everyday Postwar Life*. Vancouver, BC: University of British Columbia Press.

Bodiford, William M. 2010. "Belief Systems: Japanese Martial Arts and Religion Since 1868." In *Martial Arts of the world: An Encyclopedia, Vols. 1&2*, edited by Thomas A. Green, 382–94. Santa Barbara, CA: ABC-CLIO.

Brooke, John Hedley. 1991. *Science and Religion: Some Historical Perspectives*. Cambridge: Cambridge University Press.

Darwin, Charles. 1858. "Letter no. 2298." Darwin Correspondence Project. June 29, 1858. https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-2298.xml.

Feyerabend, Paul. 1993. *Against Method*. 3rd ed. London: Verso Books.

Fields, Karen E., and Barbara J. Fields. 2014. *Racecraft: The Soul of Inequality in American Life*. London: Verso Books.

Galileo, Galilei. 1957. "Letter to the Grand Duchess Christina." In *Discoveries and Opinions of Galileo*, translated by Stillman Drake, 173–216. New York: Anchor Books.

Gazzaniga, Michael S. 2005. *The Ethical Brain*. New York: Dana Press.

Gould, Stephen Jay. 1999. *Rocks of Ages: Science and Religion in the Fullness of Life*. New York: Ballantine.

Harraway, Donna J. 1991. *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Harris, Sam. 2010. *The Moral Landscape: How Science Can Determine Human Values*. New York: Free Press.

Harrison, Peter. 2015. *The Territories of Science and Religion*. Chicago: University of Chicago Press.

Holland, Sharon Patricia. 2012. *The Erotic Life of Racism*. Durham, NC: Duke University Press.

Jordan-Young, Rebecca M., and Katrina Karzakis. 2019. *Testosterone: An Unauthorized Biography*. Cambridge, MA: Harvard University Press.

Lee, Lois. 2015. *Recognizing the Non-Religious: Reimagining the Secular*. Oxford: Oxford University Press.

Lind af Hageby, L., and L. K. Sohartau. 1903. *The Shambles of Science: Extracts from the Diary of Two Students of Physiology*. 3rd ed. London: Ernest Bell.

Maiese, Michelle. 2016. *Embodied Selves and Divided Minds*. Oxford: Oxford University Press.

Mangan, J. A., and Takeshi Komagome. 1999. "Militarism, Sacrifice and Emperor Worship: The Expendable Male Body in Fascist Japanese Martial Culture." *The International Journal of the History of Sport* 16 (4): 181–204.

Masuzawa, Tomoko. 2005. *The Invention of World Religions: Or, How European Universalism Was Preserved in the Language of Pluralism*. Chicago: University of Chicago Press.

McCutcheon, Russell T. 2003. *The Discipline of Religion: Structure, Meaning, Rhetoric*. New York: Routledge.
McGrath, Alister E. 2019. *The Territories of Human Reason: Science and Theology in an Age of Multiple Rationalities*. Oxford: Oxford University Press.
Mercadante, Linda E. 2014. *Belief without Borders: Inside the Minds of the Spiritual but Not Religious*. New York: Oxford University Press.
Midgley, Mary. 2002. *Evolution as a Religion: Strange Hopes and Stranger Fears*. Abingdon, UK: Routledge.
Mischel, Walter, Ebbe B. Ebbesen, and Antonette Raskoff Zeiss. 1972. "Cognitive and Attentional Mechanisms in Delay of Gratification." *Journal of Personality and Social Psychology* 21 (2): 204–18.
Nongbri, Brent. 2013. *Before Religion: A History of a Modern Concept*. New Haven, CT: Yale University Press.
Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Press.
Pigliucci, Massimo. 2013. "New Atheism and the Scientistic Turn in the Atheism Movement." *Midwest Studies in Philosophy* XXXVII:142–53.
Polanyi, Michael. 1962. *Personal Knowledge: Towards a Post-Critical Philosophy*. London: Routledge.
Pollock, Anne. 2024. "Cultivating Calm Passions." *Zygon: Journal of Religion and Science* 59 (3): 800–10.
Prescod-Weinstein, Chandra. 2021. *The Disordered Cosmos: A Journey into Dark Matter, Spacetime, and Dreams Deferred*. New York: Bold Type Books.
Reeves, Josh A. 2024. "Kuhn, Conspiracy Theories, and Scientific Progress." *Zygon: Journal of Religion and Science* 59 (3): 811–24.
Riskin, Jessica. 2002. *Science in the Age of Sensibility: The Sentimental Empiricists of the French Enlightenment*. Chicago: University of Chicago Press.
Schaefer, Donovan O. 2023. "Voracious Secularism: Emotional Habitus and the Desire for Knowledge in Animal Experimentation." *Religion* 54 (4): 700–23.
———. 2015. *Religious Affects: Animality, Evolution, and Power*. Durham, NC: Duke University Press.
———. 2022. *Wild Experiment: Feeling Science and Secularism after Darwin*. Durham, NC: Duke University Press.
Scheer, Monique. "Are Emotions a Kind of Practice (And Is That What Makes Them Have a History)? A Bourdieuian Approach to Understanding Emotion." *History and Theory* 51 (May 2012): 193–220.
Shah, Esha. 2018. *Who Is the Scientist-Subject? Affective History of the Gene*. New York: Routledge.
———. 2024. "Love Affair with an Idea." *Zygon: Journal of Religion and Science* 59 (3): 825–38.
Sismondo, Sergio. 2010. *An Introduction to Science and Technology Studies*. 2nd ed. Oxford: Wiley-Blackwell.
Smith, Jonathan Z. 2004. *Relating Religion: Essays in the Study of Religion*. Chicago: University of Chicago Press.
Sullivan, Marek. 2020. *Secular Assemblages: Affect, Orientalism, and Power*. London: Bloomsbury.
Todd, Douglas. 2022. "Metro Vancouver Tops List of Non-Religious Big Cities in North America." *The Vancouver Sun*. October 27, 2022. https://vancouversun.com/opinion/columnists/douglas-todd-metro-vancouver-tops-list-of-non-religious-big-cities-in-north-america.
van Emmerik, Mari. 2024. "Changing Minds: Cogency Theory and Secular Bodies." *Zygon: Journal of Religion and Science* 59 (3): 839–49.
Yuan, Li. 2024. "Is Xenophobia on Chinese Social Media Teaching Real-World Hate?" *The New York Times*. July 4, 2024. https://www.nytimes.com/2024/07/04/business/china-hate-speech-xenophobia.html.

Zygon®: Journal of RELIGION & SCIENCE

Evers, Dirk. 2024. "Good Fences Make Engaging Neighbors." *Zygon: Journal of Religion and Science* 59 (3): 871–83. DOI: https://doi.org/10.16995/zygon.16716

Open Library of Humanities

Good Fences Make Engaging Neighbors

Dirk Evers, Professor of Systematic Theology, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany, dirk.evers@theologie.uni-halle.de

This article comments on the concept of science-engaged theology by highlighting several issues. It asks for the explanatory value of the sociological concept of secularization, points to the relevance of esotericism, elaborates on the importance of context with reference to developments around science and religion in Germany, and suggests an appropriate border regimen between science and religion by employing hermeneutic means to strive towards science-engaged theologies.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

I am grateful for the opportunity to offer some thoughts on John Perry and Joanna Leidenhag's exciting book on science-engaged theology. I am doing this as a German Protestant systematic theologian. From this perspective, I find it particularly encouraging that Perry and Leidenhag's proposal for new and bold modes of science–theology interaction moves away from traditional dichotomies such as religion vs. science or natural vs. revealed theology. I have organized my thoughts and comments along six points.

Secularization as Explanation?

Science-engaged theology stresses historical contextualization of both "science" and "theology," and of their ways of engagement. In their book, Perry and Leidenhag (2023, 9) refer to Peter Harrison when they write: "science and religion are not transhistorical categories that we can track throughout different epochs, but imagined concepts that, as a result of certain theories of secularism, have come to be defined in opposition." Here as well as at other places Perry and Leidenhag (2023, 20) refer to a meta-narrative of secularization. According to the presentation by the authors, Harrison also subsumes "the story of science and religion . . . within larger accounts of secularization and the loss of Aristotelian-Thomistic teleology." According to this meta-narrative the 'imagined' concepts of science and religion in their modern meaning "were *invented* as tools of an ideology called secularism" (2023, 21). All this insinuates that the ideology of secularism has invented the imagined concepts of religion and science, and that if we get rid of that ideology we might leave the conflict behind. However, the sociological concept of secularization or secularism is not so much a descriptive historical category as it is an analytical term which became common with Sigmund Freud, Max Weber, Émile Durkheim, Peter L. Berger, and others as a characterization of the development of modernity away from religious dominance in society.

Usually, three aspects of secularism are distinguished: (1) Functional differentiation: "the process of societal modernization as a process of functional differentiation and emancipation of the secular spheres—primarily the state, the economy, and science—from the religious sphere and the concomitant differentiation and specialization of religion within its own newly found religious sphere" (Casanova 2008, 19). (2) The decline of religious convictions and religious behavior. (3) An increasing privatization of religion so that the religious sphere is reduced to the narrow realm of private life. In his famous analysis from 1917, Max Weber (1985, 594), for example, linked these developments to the rise of science leading to the increased intellectualization and rationalization of modernity and thus to the "disenchantment of the world," in which technical means and calculation provide tools to master nature. For Weber, this development had positive and negative aspects. It gained technical

progress, but it also enclosed the modern individual into a "shell as hard as steel" (Weber 1922, 203), or, as in the famous translation by Talcott Parsons, into an "iron cage" (Weber 1930, 181). Others have questioned the thesis of a general secularization of the Western World on principle or on empirical grounds. Today, the debate about secularization and whether or not we already live in a post-secular age is an open one (Costa 2022).

In any case, one should bear in mind the warning of the great church historian Owen Chadwick from his Gifford Lectures:

> Umbrella terms, however doubtful, are useful. I do not think it an abuse of such a term to call this radical process, still in part so obscure to the enquirer, still in part undefined and possibly in part undefinable, by the name of secularization; on the one condition (and it is an absolute condition) that the word is used, neither as the lament of nostalgia for past years, nor as propaganda to induce history to move in one direction rather than another, but simply as a description of something that happened to European society in the last two hundred years. And what happened, and why, must still be matter for much enquiry by students of history and religion and society. (Chadwick [1975] 2000, 265–66)

Therefore, it seems to me a normative and in itself anachronistic construct to identify a secularist ideology or agenda as the cause behind the debates and conflicts around science and religion in the nineteenth century, and a problematic move to insinuate that if only we overcome this ideology those tensions between the imagined constructs of religion and science at the expense of religion would go away or be transformed into fruitful engagement. Even if one agrees that the modern concepts of science and religion have been shaped in the nineteenth century and that they were used within developments and discourses which today many describe as secularization, such a view is no explanation and requires careful analysis of the complexity of the conflictual constellations between reason, faith, empirical sciences, scripture, scientific world-views, claims for religious tolerance, democratic change, etc. It is, indeed, complicated (Perry and Leidenhag 2023, 16–21) and not the result of ideology.

What about Esotericism?

Holding that caveat in mind, one can subscribe to the claim that science and religion, at least in the way we understand them today, are concepts of relatively recent coinage. Any thesis about a conflict between these two concepts, let us say, in the seventeenth century, when Galilei stood against the Vatican astronomers and argued against their Aristotelian concepts, or even at the time of early Christianity confronted with Hellenistic philosophy, is, as Peter Harrison (Harrison 2015, 3) writes, an "anachronistic application of modern concepts to past eras." However, in addition to Harrison and others, and

following Heidelberg professor of religious studies, Michael Bergunder (see Bergunder 2020 and his earlier study, Bergunder 2016), I would add esotericism or occultism to the concepts of science and religion as a third concept that was shaped and constructed in the nineteenth century in public discourses around religious and scientific worldviews and their meaning for the future of a changing society. Concepts of religion, science, and esotericism have a common history emerging in the course of the nineteenth century and particularly in its latter half. Modern esotericism as the idea of a re-enchanted worldview is an antagonist movement of amalgamation of religion and science against the religion–science divide established in those days. Recent research in the history of esotericism emphasizes that in its modern form it is no uninterrupted continuation of early modern concepts of secret knowledge and the like but has its roots in the second half of the nineteenth century, when Spiritism, Occultism and Theosophy as decidedly esoteric movements were formed and developed an intensive exchange of ideas with each other (Bergunder 2020, 54).[1] Kocku von Stuckrad (2014) argues that European secularism in fact has been religiously productive itself. The nineteenth century saw the emergence of professional knowledge about religion in a global perspective, while science was used to develop comprehensive worldviews, was propagated as a resource of meaning, and religious notions were merged with scientific concepts and practices. Von Stuckrad speaks of the *Scientification of Religion* between 1800 and 2000, which is still effective in present perspectives of a consonance between science and religion.

According to Bergunder, the concept of religion in the nineteenth century was constructed as a contrast to that of science, while in response different forms of esotericism propagated the intrinsic affinity and unity of science and religious attitudes and beliefs. This took place in the framework of a global discourse that not only constructed science and religion as antagonists but at the same time gave rise to a countercurrent that wanted to merge science and religion. Analogous discourses can be identified in the major global religions like Islam, Hinduism, and Buddhism, which in the nineteenth century in the wake of colonialism saw movements of renaissance and reform claiming a new amalgamation of what in the West seemed to become more and more antagonist forces, namely religion and science in their modern sense. In all religions, including Christianity and in the new freethought and free-religious movements, efforts emerged to justify religion on scientific grounds and to substantiate extrasensory phenomena and religious states of mind like meditation empirically and scientifically.

This constellation means that what today we call science and religion is not only shaped by claims of an intrinsic conflict between science and religion constructed in the nineteenth century, but also by alternative claims of an intrinsic consonance between science and religion, which at the same time argues for a

transformation of both, science and religion, in order to merge them at a higher level. For later representatives one might think of physicist David Bohm, who had been a follower of Jiddu Krishnamurti, an Indian philosopher promoted by Theosophists, or Stuart Kauffman's well-known book *Reinventing the Sacred: A New View of Science, Reason, and Religion* (Kauffman 2008). Bergunder as a historian of religion even files physicist and icon of the science and religion field John Polkinghorne under the label esotericism as well as Teilhard de Chardin and other forms of Anglo-Catholicism—a provocative, but maybe stimulating perspective on the discourse around science, religion, and theology and its roots in the nineteenth century. Recently John Milbank followed the tracks of esotericism when he referred to magic as what had been squeezed out in order to reach a truce between "religion" and "science" (Milbank 2022, 141). However, drawing on Antoine Faivre's understanding of Western esotericism he traces it back to early modernity and identifies in it a constant subtext to the development of modernity. He concedes that it "is not possible to respond to the arrival of 'science' in the 19th century merely by going back to Aristotle, the Fathers or Aquinas" (Milbank 2022, 119), but claims orthodoxy for esotericism and sees in it the means to overcome disenchanted immanence: "Magic alone unites science with religion" (Milbank 2022, 143). Perry and Leidenhag criticize Milbank on theological and hermeneutical grounds when he tries to establish theology as "master discourse" (Milbank 2008, 5–6) and degrades (secular) science as "a bad imitation of theology" (Milbank 2008, 22). However, in my view he also stays within the limits of nineteenth century discourse when he tries to overcome the ideology of secularization by means of "esotericism" not seeing that also esoteric notions are shaped and transformed in relation to secularization. Science-engaged theology would then not only have to contextualize concepts of science and religion, but also those of esotericism which are trying to re-enchant reality by inventing scientific concepts of religion and insinuating a religious understanding of science.

Context, Not Provinciality: The German Development

Another implication of historical contextualization is that constellations are different in different contexts. Contexts in the nineteenth century were different, for example, in Great Britain and in Germany. In Germany, public discourses on materialism, on Darwinism and on Monism shaped the scene for science and religion since the 1840s. There developed what was among the contemporaries identified as *Weltanschauungskampf*, an ideological struggle about worldviews (Evers 2024). A number of *ideologies* were fighting for political and public influence, among them socialist and democratic political movements. Science had built up considerable social prestige and many scientists (*Naturforscher*)[2] became public figures and defended the claim that science was the modern means to explain the totality of reality. Among them were public figures like Karl Vogt, Jakob

Moleschott, Ludwig Büchner, and Ernst Haeckel (see Chadwick [1975] 2000, 163–82) who often proclaimed socialist or liberal ideas and were decidedly anti-clerical. Numerous volumes on scientific worldviews appeared and different public pressure groups, from socialist to religiously conservative, were founded. Among them was the German Monist League (*Deutscher Monistenbund*) founded in 1906 as an association for the dissemination of a scientific *Weltanschauung* (worldview). In the founding manifesto written by Heinrich Schmidt it reads:

> The tremendous advances of the natural sciences have displaced and eliminated all outdated dogmatic and mystical ideas about the world and human beings, about body and spirit, creation and development, becoming and passing away of material things. Outdated dualistic ideas are increasingly being replaced by monistic ones, and those unsatisfied in the worldview (*Weltanschauung*) sanctified by tradition are looking for a unified worldview (*Weltanschauung*) based on natural science. This [worldview] rejects the belief in outdated, traditional dogmas and revelations and puts pure reason in their place. (Drehsen and Zander 1996, 218, my translation)

After the breakdown of German idealism, discourses were also shaped by the diversity of German philosophical schools: first Neo-Kantianism, then Neo-Positivism with Gottlob Frege, Rudolf Carnap, and the Vienna circle; later new existentialism in the Heideggerian version with sympathies for the upcoming national-socialist move towards a Germanic *Weltanschauung*; and a variety of attempts to establish philosophy as a moderator of *Weltanschauung*, like Wilhelm Dilthey's *Weltanschauungslehre*: the historico-philosophical discipline of worldview studies. However, after hefty controversies in the wake of World War I about Germany's cultural, religious and political fate in the first decades of the twentieth century, the Nazis finally claimed the monopoly of *Weltanschauung* and established Alfred Rosenberg in 1934 as "the Führer's representative for the supervision of the entire spiritual and ideological (*weltanschauliche*) training and education of the NSDAP," often referred to as *Amt Rosenberg* (Rosenberg Office). The whole development of German Protestant theology in the first half of the twentieth century, the insistence of dialectical theology not to turn religion into a comprehensive worldview, its rejection of natural theology, etc. must be seen against the futility and final outcome of such public and political discourses.

Debates on science and religion in the UK with its strong history of natural theology prominent even among natural scientists, including the young Charles Darwin, were less ideological. Even the debates about Charles Darwin's theory of evolution were less heated in the UK than in Germany, where Haeckel made use of Darwin for promoting his naturalist and monist worldview, which had already been formed years before Darwin published his theory of evolution. Darwin himself remarked in a letter that

> a week hardly passes without my hearing of some naturalist in Germany who supports my views, & often puts an exaggerated value on my works; whilst in France I have not heard of a single zoologist except M. Gaudry . . . who supports my views. (Darwin [1870] 2023)

In the UK, with its broad empiricist tradition and its elastic Anglican main church, public debates gained momentum in the late nineteenth century when John Tyndall and other scientific naturalists made the argument that the uniformity of nature, natural law, and causality, which natural theology had understood in theistic terms, rather point to a non religious understanding of the natural world (Stanley 2019). The so-called prayer gauge debate about petitionary prayer in 1870 made this a public controversy and linked it to the question if the efficacy of petitionary prayer can be measured scientifically (Tyndall and Galton 1876). Tyndall himself, who had studied in Germany, was aware of the German movement of Neo-Kantianism that argued for an agnostic stance towards religion from a scientific perspective and thus was far away from developing what in Germany would have been called a *Weltanschauung*.

For science-engaged theology, Perry and Leidenhag stress the importance of different historical or geographical contexts for science as well as theology. There will always be contingent social factors and one has to be "aware of the local, political and sociological issues surrounding particular studies and interpretations of data" (Perry and Leidenhag 2023, 65). But such awareness calls for expansion and dialogue beyond the particular context in order to understand it in its specific particularity. Science-engaged theology must not indulge in provinciality but should also develop a critical sense of its own context. It must seek interaction with other contexts and also with the broader questions of epistemology and ontology and thus with philosophical discourses which provide tools of analysis beyond contextual particularity. Theology should not fall prey to communal and cultural self-evidence nor to cultural relativism. In a global and inter-religious perspective and with reference to philosophical reflection, even science-engaged theology can, in the long run, not refrain from introducing hermeneutical and philosophical perspectives which point beyond historical context. Historical contextualization allows for deconstructing dominant and persistent narratives, but it is not a relief from the burden of constructive engagement with substantial philosophical questions including those of epistemology, philosophy of science, and ontology, as well as with doctrinal claims of theology.

Borders and Fences: On the Dialectics of Engagement

Perry and Leidenhag's understanding of science-engaged theology questions the truce between science and religion as many understand it. Especially in my context of German academic theology, such a truce as built on a categorical

separation between science and religion is seen as self-evident. This view is understood as a major progress started by Schleiermacher and ratified at the beginning of the twentieth century that science and religion no longer interfere and refrain from participating in an endless attrition warfare. In his open *Letters to Dr. Lücke*, Schleiermacher (Schleiermacher 1981, 64) proposed an "eternal covenant between the living Christian faith, and completely free, independent, scientific inquiry, so that faith does not hinder science and science does not exclude faith." This covenant is usually understood as a categorical separation between religion and science, so that theology as a reflection on and articulation of religious consciousness as feeling of ultimate dependence can in principle not be in conflict with science, because both have nothing to do with each other.[3] Or as systematic theologian Ulrich Barth claimed a few years ago in an essay with the speaking title "Farewell to Cosmology: Liberation of Religion to Itself": "Since Schleiermacher—this can be stated without exaggeration—the dispute between theology and the natural sciences over questions of cosmology has basically ceased to be an issue" (Barth 1995, 35, my translation). In his view the religious concept of creation is concerned with reflections on finitude (*Endlichkeitsreflexion*) and has nothing to do with theories of matter in space and time or the origin of the universe. A similar case but on the other side of the theological spectrum are Karl Barth's remarks at the beginning of his doctrine of creation "that there can be no scientific problems, objections or aids in relation to what Holy Scripture and the Christian Church understand by the divine work of creation. . . . There is free scope for natural science beyond what theology describes as the work of the Creator. And theology can and must move freely where science which really is science, and not secretly a pagan Gnosis or religion, has its appointed limit." (Barth 1958, 1–2). However, his preface to *Church Dogmatics* vol. III/1 was written in October 1945, only months after Germany's unconditional surrender when the "pagan Gnosis," the *Weltanschauung* of the Nazis, had left Europe in ruins. It must be read against the background of a century of *Weltanschaungskampf* culminating in nationalist, ethno-racist and militarist ideologies built on pseudo-science.

Against any notions of decoupling religion from science or, as they call it, *Non-Overlapping Magisteria* (*NOMA*) *on steroids*, Perry and Leidenhag argue for theology's qualified engagement with science. For this they propose to move beyond the territories of science and religion by deconstructing the antagonist categories as "deeply misleading," (Perry and Leidenhag 2023, 21). On the other hand, they concede that "a lack of clear borders does not make the label of science meaningless" and that "this blurred border" does not "license an anything goes relativism" (2023, 47). For me, the term *engaged* in science-engaged theology indeed presupposes science as something distinct from theology so that engagement becomes possible at all. Theology's potential commitment to science refers to the hermeneutical dialectics that, to use a famous line from poet Robert Frost, "good fences make good neighbours"

(Evers 2019) and thus separation and relative independence are necessary for mutual engagement. Christian theology begins with faith seeking understanding (Evers 2019, 63) and is not and cannot be an explanatory reconstruction of certain phenomena of reality in a third-person perspective. It is also taking first-person perspectives (such as faith as basic trust and personal involvement) and second-person conditions (such as expressivist, non-designative aspects of language and semantics, cultural formations and hermeneutical approaches to verbal and non-verbal communication) into account. Theology engages in open discourses which aim at integrating third- and first-person perspectives with critically reflected and elucidated second-person conditions. Science, on the other hand, must develop and fulfill empirico-mathematical methodological requirements. These are not so clear-cut as many think, and they often depend on the subject-matter one is dealing with, but in any case they are meant to relate formal models to reality in an objectified third-person perspective. Such methodological distinctions between theological discourse and scientific investigation improve and strengthen possible forms of engagement and even make them possible, while blurring distinctions can render fruitful exchange impossible, because it becomes unclear what exactly we are talking about.

My move would be not to blur borders, but to de-politicize and de-militarize them. A fence is not a wall, *a fortiori* no Berlin wall with border police and spring guns. Fences are permeable, albeit selective in their allowance of thorough passing. There is always peaceful large and small border traffic—between science and religion for example in terms of models and images borrowed from the other side. And fences require constant mending, re-negotiating and at times even moving. The territories of science and religion must not be understood in ways of political entities governed by legislation, police, and politicians, but as free and intrinsically diversified spaces which together with others like music, politics, arts, language, food, etc. form the landscape of human culture. In this landscape science and theology are indeed neighbors. They share a common history and common interests of rationality and truthfulness. And it is important that they are not seen as archenemies, but as engaged in constructive dispute—at times even within a person—about the limits and potentials of human knowledge. Then both are able to contribute their respective and particular gifts, especially when they engage in third places: in the mayor hall, in the marketplace, in the theatre, in parliament, in the media, etc.

Hermeneutics as Foundation: Science and "Wissenschaft"

For any science-engaged theology reflections on hermeneutics, that is reflections on the theory and methodology of interpretation and understanding, are key. Its approach makes it necessary to distinguish between different levels of semantics and conversation. That starts with reflections on the understanding of science as a source for theological reflection: "What, then, is it that science-engaged theologians are meant to engage *with*?" (Perry and Leidenhag 2023, 36) What we are talking

about when we talk about science: Consolidated scientific knowledge? Scientific hypotheses? Scientific data? Scientific research programs? Science as authority? Science as a cultural force? Science based or science-engaged worldviews? Science-flavored narratives? And science in which arena: at the university, in research labs, in public political discourse, as a culture (Snow [1959] 2000)? . . .

Let me again share a notion of my German context. While in the English-speaking world "science" has adopted a somewhat narrow sense, *Wissenschaft* in German means something different: a system of doctrines and knowledge regarding certain subject matters and methods related to them. *Wissenschaft* can be any ordered, systematic body of knowledge. Right from its beginning in the eighteenth century this notion had two aspects, the "material" aspect of knowledge, and the formal aspect of certain methodological disciplines to gain reliable knowledge with respect to certain objects of knowledge. The German constitution in article 5(3) grants freedom of *Wissenschaft* (academic freedom), and in 1973 the *German Federal Constitutional Court* defined *Wissenschaft* as "any activity which, according to its content and form, is to be regarded as a serious planned attempt to discover/determine the truth";[4] the court even links this broad and open definition to the "fundamental incompleteness of any scientific knowledge"[5] (Bundesverfassungsgericht 2023, my translations). As I see it, this resonates with Perry and Leidenhag's vision of unity and pluralism in science and beyond. In the German context, *Wissenschaft* divides into different sub-forms such as *Naturwissenschaften* (natural sciences), *Geisteswissenschaften* (humanities), *Kulturwissenschaften* (cultural sciences), *Sozialwissenschaften* (social sciences), *Sprachwissenschaften* (linguistics), etc. Often, the medieval higher faculties such as medicine, law, and theology are seen as practical disciplines dedicated to certain practical and institutional tasks and for this purpose employing methods and knowledge from different other disciplines across the spectrum. Mathematics, philosophy, and academic theology are often regarded as special disciplines of *Wissenschaft* which fall outside such sub-categories related to specific subject matters, while other *Wissenschaften*, like psychology, come in different varieties, either as a natural or a social science. Other disciplines are hybrids from the start like computer-linguistics combining empirical scientific methods with linguistics.

Theology in my understanding must be developed as "a serious planned attempt to discover the truth," and thus must be *Wissenschaft*. However, there are different levels and categories of truth. While *empirical sciences* deal with those forms of knowledge which are in such a controlled and subtle relation to empirical reality that they can take on the form of mathematical regularity and allow for empirical falsification and corroboration, *theology* refers to God as ultimate truth and therefore to truth as a transformative power from beyond human knowledge, experience, and calculation. If we want to establish "a serious planned attempt" to discover this kind of truth, we are dependent on revelation, on God making Godself accessible. While all human

scientific knowledge is intrinsically incomplete, theological truth claims fall even more under a fundamental reservation for a number of reasons. They cannot refer to God as an object which can be pinned down by empirical or methodological means. They can only try to articulate how our views of reality and our individual and communal ways of living change if what is understood as a revelation of God unfolds its orienting power. Thus, the theological quest for truth takes on different forms of opening and orienting human existence on personal, communal, and public levels, and it draws from the sources of religious traditions and practices seeking disclosures of truth (Ramsey 1973). If I understand science-engaged theology correctly, theological truth claims must not take the form of a super-science to explain reality, but they must point towards the mystery of God by developing empirically meaningful and historically particular presentations and analyses of reality.

Science-Engaged Theology as Cool Theology

Dewi Z. Phillips has distinguished between warm and cold philosophy. In a nutshell, warm philosophy is passionate in pursuing goals for shaping human existence, while cold philosophy simply describes different philosophical options in a historical or analytical perspective. Phillips identifies his own philosophy in contrast to these two extremes as cool philosophy. With this he refers to Ludwig Wittgenstein, who wrote in 1929: "My ideal is a certain coolness. A temple providing a setting for the passions without meddling with them" (Phillips 1999, vi). The sciences help to cool the temple of theological discourse too often heated by religious and other ideologies. They help keeping religious phantasies at bay and prevent them from running wild. They help to refute political usurpations of religion and theology, and they bind theology's moral and ethical discourses to concrete reality so that they can gain traction. They help to calibrate personal, spiritual, subjective perspectives on reality with references to empirical reality and intersubjective conditions of communication and argument. They help to prevent religious notions from developing into wishful thinking and inspire non-foundationalist and pluralist forms of "relaxed metaphysics" (Dalferth 2017). They prevent theological reflection from turning into a kind of pseudoscientific argument for trans-scientific objectives. In these respects, science-engaged theology is cool.

Notes

[1] This stands against the theses of Antoine Faivre, the founder of the study of Western esotericism, who saw the historical origins of today's esotericism in the fifteenth to seventeenth century (Faivre 2010). He presumed a subsequent continuous history, starting with high-degree Freemasonry, Swedenborgianism, and Mesmerism in the eighteenth century, through Romantic natural philosophy at the beginning of the nineteenth century and Spiritism/Occultism as well as modern Theosophy in the second half of the nineteenth century, to today's contemporary esotericism. As we will see, John Milbank refers to Faivre in his account of early modern esotericism. However, careful studies like that of Friedemann Stengel on Swedenborg show that Faivre's criteria of esotericism are won from early modern phenomena and do not apply to later forms (Stengel 2023).

[2] The term *Naturforscher* was already used in the eighteenth century. Kant, for example, uses this term frequently. The term *Naturwissenschaftler* (scientist) only emerged in the second half of the nineteenth century.

[3] But see the careful analysis of Schleiermacher's view by Pedersen (2017).

[4] In German: "alles, was nach Inhalt und Form als ernsthafter planmäßiger Versuch zur Ermittlung der Wahrheit anzusehen ist."

[5] In German: "prinzipielle . . . Unabgeschlossenheit jeglicher wissenschaftlichen Erkenntnis."

References

Barth, Karl. 1958. *Church Dogmatics: Volume III.1 (§§ 40–42)*. Edited by Geoffrey W. Bromiley and Thomas F. Torrance. Edinburgh: T & T Clark.

Barth, Ulrich. 1995. "Abschied von der Kosmologie: Befreiung der Religion zu sich selbst." In *Urknall oder Schöpfung? Zum Dialog von Naturwissenschaft und Theologie*, edited by Wilhelm Gräb, 14–42. Gütersloh: Kaiser/Gütersloher Verlagshaus.

Bergunder, Michael. 2016. "'Religion' and 'Science' within a Global Religious History." *Aries: Journal for the Study of Western Esotericism* 16:86–141.

———. 2020. "Umkämpfte Historisierung: Die Zwillingsgeburt von 'Religion' und 'Esoterik' in der zweiten Hälfte des 19. Jahrhunderts und das Programm einer globalen Religionsgeschichte." In *Wissen um Religion: Erkenntnis–Interesse: Epistemologie und Episteme in Religionswissenschaft und Interkultureller Theologie*, edited by Klaus Hock, 47–131. Leipzig: Evangelische Verlagsanstalt.

Bundesverfassungsgericht. 2023. "Entscheidungen des Bundesverfassungsgerichts: Hochschul-Urteil (1973)." https://www.servat.unibe.ch/dfr/bv035079.html.

Casanova, José. 2008. *Public Religions in the Modern World*. 6th ed. Chicago: University of Chicago Press.

Chadwick, Owen. (1975) 2000. *The Secularization of the European Mind in the Nineteenth Century*. Cambridge: Cambridge University Press.

Costa, Paolo. 2022. *The Post-Secular City: The New Secularization Debate*. Paderborn, Germany: Brill Schöningh.

Dalferth, Ingolf U. 2017. "A Relaxed View of Metaphysics: Neo-Aristotelian Thomism and the Theological Legacy of Thomas Aquinas." *Zeitschrift für Theologie und Kirche* 114 (1): 49–81.

Darwin, Charles. (1870) 2023. "Darwin Correspondence Project: Letter No. 7204: To Armand De Quatrefages 28 May [1870]." https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-7204.xml.

Drehsen, Volker, and Volker Zander. 1996. "Rationale Weltveränderung durch 'naturwissenschaftliche' Weltinterpretation?" In *Vom Weltbildwandel zur Weltanschauungsanalyse: Krisenwahrnehmung und Krisenbewältigung um 1900*, edited by Volker Drehsen and Walter Sparn, 217–38. Berlin: Akademie Verlag.

Evers, Dirk. 2019. "'Good Fences Make Good Neighbours': Why the Differences of Science, Religion and Theology Must Not Be Blurred." In *Our Common Cosmos: Exploring the Future of*

Theology, Human Culture and Space Sciences, edited by Zoë Lehmann Imfeld and Andreas Losch, 21–35. London: T & T Clark.

———. 2024. "Apologetics and '*Weltanschauung*' in Germany Since the Nineteenth Century." In *Science, Religion, the Humanities and Hope*: *Essays in Honour of Willem B. Drees*, edited by Michael Fuller and Anne Runehov, 133–51. Cham, Switzerland: Springer.

Faivre, Antoine. 2010. *Western Esotericism: A Concise History*. Albany, NY: State University of New York Press.

Harrison, Peter. 2015. *The Territories of Science and Religion*. Chicago: University of Chicago Press.

Kauffman, Stuart A. 2008. *Reinventing the Sacred: A New View of Science, Reason and Religion*. New York: Basic Books.

———. 2008. *Theology and Social Theory: Beyond Secular Reason*. 2nd ed. Malden, MA: Blackwell.

Milbank, John. 2022. "Religion, Science and Magic: Rewriting the Agenda." In *After Science and Religion: Fresh Perspectives from Philosophy and Theology*, edited by Peter Harrison, John Milbank, and Paul Tyson, 75–143. Cambridge: Cambridge University Press.

Pedersen, Daniel James. 2017. *The Eternal Covenant: Schleiermacher on God and Natural Science*. Berlin: de Gruyter.

Perry, John, and Joanna Leidenhag. 2023. *Science-Engaged Theology*. Cambridge: Cambridge University Press.

Phillips, Dewi Zephaniah. 1999. *Philosophy's Cool Place*. Ithaca, NY: Cornell University Press.

Ramsey, Ian T. 1973. *Models for Divine Activity*. London: S. C. M. Press.

Schleiermacher, Friedrich Daniel Ernst. 1981. *On the Glaubenslehre: Two Letters to Dr. Lücke*. Atlanta: Scholars Press.

Snow, Charles Percy. (1959) 2000. *The Two Cultures*. Edited by Stefan Collini. Cambridge: Cambridge University Press.

Stanley, Matthew. 2019. "God and the Uniformity of Nature: The Case of Nineteenth-Century Physics." In *Science Without God? Rethinking the History of Scientific Naturalism*, edited by Peter Harrison and Jon H. Roberts, 97–110. Oxford: Oxford University Press.

Stengel, Friedemann. 2023. *Enlightenment All the Way to Heaven: Emanuel Swedenborg in the Context of Eighteenth-Century Theology and Philosophy*. Royersford, PA: Swedenborg Foundation.

Tyndall, John, and Francis Galton. 1876. *The Prayer-Gauge Debate*. Boston: Congregational Publishing Society.

von Stuckrad, Kocku. 2014. *The Scientification of Religion: An Historical Study of Discursive Change, 1800–2000*. Berlin: de Gruyter.

Weber, Max. 1922. "Die Protestantische Ethik und der Geist des Kapitalismus." In *Gesammelte Aufsätze zur Religionssoziologie Bd. I.*, 2nd edition, 17–206. Tübingen, Germany: J. C. B. Mohr (Paul Siebeck).

———. 1930. *The Protestant Ethic and the Spirit of Capitalism*. Edited by Talcott Parsons. London: G. Allen & Unwin.

———. 1985. "Wissenschaft als Beruf." In *Gesammelte Aufsätze zur Wissenschaftslehre*, 6th edition, edited by Johannes Winckelmann, 582–613. Tübingen, Germany: Mohr.

Zygon®: Journal of RELIGION & SCIENCE

Leung, King-Ho. 2024. "Theology and Policing." *Zygon: Journal of Religion and Science* 59 (3): 884–92. DOI: https://doi.org/10.16995/zygon.17702

Open Library of Humanities

Theology and Policing

King-Ho Leung, Lecturer in Theology, Philosophy and the Arts, King's College London, London, UK, king-ho.leung@kcl.ac.uk

This article assesses the notion of "policing" in John Perry and Joanna Leidenhag's *Science-Engaged Theology* to consider the contemporary practice of theological inquiry as a theoretical endeavor. Drawing on their construal of science-engaged theology, which involves a twofold resistance towards "authoritarian" tendencies of asserting theology as the queen of the sciences as well as modernity's "border police" who seek to exclude religion and theology from public discourse, this essay suggests that Perry and Leidenhag's vision for science-engaged theology not only brings to light one of the ways theology is sometimes deployed as a political mechanism for policing but also how theology itself can perhaps even be considered an act of "theological *de*-policing."

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

At the heart of John Perry and Joanna Leidenhag's vision for the relationship between science and religion in their recent *Science-Engaged Theology* is the geopolitical imagery of the national border and border policing. As Perry and Leidenhag (2023, 47) write in a pinnacle passage in their short but insightful book:

> [W]e do not think that the only way for theologians to speak with boldness and positively contribute to the contemporary academy is to be queen of the sciences, empress of natural philosophy, or to remember with nostalgia the bygone days of Christendom. Christendom no longer exists, and what exists on roughly the same landmass that it once did is the fragile and imperfect European Union. The EU provides something of a metaphor for our vision of the university. It is a trading zone where national borders remain, and cultural and linguistic differences can be celebrated, but where there is also a shared set of standards allowing for increased collaboration and the relatively free movement of goods and peoples.

This passage in brings together a few of the recurring themes and motifs in Perry and Leidenhag's fascinating book.

First, Perry and Leidenhag maintain that theologians must speak with boldness and positive integrity, and not simply give in to scientism or adopt scientific standpoints in an uncritical manner, but to do so not in a dominating fashion. Second, science-engaged theology, as they see it, is a theological approach that is distinctly post-Christendom or indeed modern, with a contemporary as opposed to historical or genealogical understanding of "science." Third, to practice science-engaged theology properly is to be attentive to the place of theology as an academic discipline within the modern—or even "secular"—university. As with their metaphor of the national borders within the EU, while science-engaged theology is committed to distinct borders and seeks to avoid "the border policing" or "gatekeeping" of academic disciplines. Understood as such, science-engaged theology is committed to (1) theological boldness and integrity, with the caveat of not recasting theology as the queen of the sciences, (2) an acute awareness of the peculiar conditions of modernity, and (3) a commitment to maintaining clear disciplinary "borders"—but doing so without falling into the practice of "policing." These three points will be treated one after another in what follows.

In many ways, Perry and Leidenhag's vision for science-engaged theology is presented as a critique of John Milbank ([1990, 6, 380] 2006, 6, 382), particularly his account of theology as "a master discourse," or even "the queen of the sciences,"[1] most famously articulated in his now classic *Theology and Social Theory*—whose 1990 first edition was notably and provocatively dedicated to "the remnants of Christendom."[2] Part of Perry and Leidenhag's worry is that Milbank's theological vision explicitly does not allow for any "autonomous

reason," whereby one can approach reality without any reference to God, which, in Milbank's view, opens up the possibility of secularity. For Milbank, there is no area of life or reality that can stand independently apart from God, for all created reality only exist by virtue of participation in God who is Being itself. Accordingly, just as there is no thing that can be ontologically independent from God, the proper knowledge of all things must also always be connected to God—and, by extension, to theology as the study and knowledge of God. In view of this all-encompassing theological vision, Perry and Leidenhag (2023, 34) argue that "Milbank's methods—his attempt to regain theology's place as queen—tend towards authoritarianism," drawing in part on an earlier critique put forward by Linn Tonstad (2020).

While Perry and Leidenhag are alarmed by the authoritarian tendencies of those who posit theology as the queen of the sciences, they are also critical of those who uncritically subordinate theological discourse to scientific findings: "On the one hand, some have used science as a proxy for rationality or emancipation and religion as a proxy for superstition and oppression. On the other hand, others have sought to declare theology queen of the sciences" (Perry and Leidenhag 2023, 7). As Perry and Leidenhag point out, this narrative of *science* as a means for emancipation is a predominant theme of the modern political spirit held by those such as Thomas Jefferson. Such thinkers, they say, "joined Descartes and Locke on their quest for certainty and joined Voltaire and Jefferson on their quest for emancipation. They *created* something they called 'science' that was good, and something called 'religion' that was bad" (Perry and Leidenhag 2023, 23). This brings us to the second point noted above: the peculiar conditions of modernity.

Following this modern Jeffersonian narrative, Perry and Leidenhag (2023, 22) point out that many theologians "retreated into a realm of private spirituality," holding that theology is no longer a public but "private matter." This is, of course, a characteristically modern condition which we sometimes call *secular*. And, indeed, it is one that is played out not only in the division between religion and science, but also—perhaps even more notably—between religion and politics. As Perry and Leidenhag (2023, 21) note, this purported division or conflict between religion and science "is quintessentially a part of the political, cultural and philosophical shifts that came to be known as modernity."

The parallel between the religion/science opposition and the religion/politics division in modernity is particularly notable in a section titled "'Science' and Modernity's Border Cop" in Perry and Leidenhag's (2023, 24) book:

> [Thinkers such as Descartes, Locke, Voltaire, and Jefferson all shared the modern goal of] advocating for a universal and neutral criterion that could sort the rational wheat from the superstitious chaff. If deductive certainty or subjective taste preferences are the only options, into which box do theological

> and ethical claims fit? It was never clear. William Wood labels those who would exclude theology and ethics from the rational discourse of public institutions, such as universities, the field's "border cop."

Not unlike the "secular" exclusion of religious views from the public political sphere, Perry and Leidenhag (2023, 24) describe the exclusion of theology and ethics from the rational—"scientific"—discourse of public institutions as "modernity's standard policing strategy."

As we can see from the passage quoted above, Perry and Leidenhag explicitly draw on the notion of "border patrolling" or "policing" from William Wood's (2021) recent *Analytic Theology and the Academic Study of Religion* to speak of the prohibition or even exclusion of religious and theological voices from modern secular discourse. A similar notion of the secular policing of religion and theology can already be found in John Milbank's aforementioned *Theology and Social Theory*. According to Milbank ([1990, 106] 2006, 106), there is "a secular policing or 'encompassing' of religion within the flattened dimension of modern public space," whereby religion and theology are "'policed', or kept rigorously behind the bounds of the possibility of empirical understanding." In Milbank's ([1990, 2] 2006, 1) own view, "to contest this secular positioning of theology," is not an authoritarian gesture but, to the contrary, an anti-authoritarian act of resisting such "secular policing" (cf. Wood 2021, 287–88).[3] However, the "policing" which Milbank's theology opposes is not only that of "secular policing", but also what he at one point calls the "arch-reactionary" position of neo-scholastic "two-tier" Thomism (Milbank 2005, 26)—which is also of relevance for some of the themes of Perry and Leidenhag's theological vision, as will be made clearer below.

Milbank's rejection of secular or autonomous reason is sometimes coupled with or even expressed in his theological refutation of the neo-scholastic Thomist notion of *natura pura*, or pure nature, in his theology of grace and nature. Just as there is for Milbank reason can never be neutral, autonomous or independent from God and God's revelation (cf. Wood 2021, 104–6), nature itself is so to speak always already graced, it can never be independent or separated from God's grace: there is therefore no such thing as a "*pure* nature" (see Milbank [2005] 2014).

Milbank's rejection of "pure nature" would presumably sit uncomfortably with Perry and Leidenhag's vision for science-engaged theology. For Perry and Leidenhag not only seek to maintain the "national borders" between theology and other academic disciplines (especially the natural sciences), but also explicitly call for theologians to "remember nature": that one "could think of science-engaged theology as a *memento naturam*" which reminds theologians that "science is one of the ways that Christians have learned to listen to God" (Perry and Leidenhag 2023, 1, 6). But while Milbank's rejection of pure nature may be deemed too anti-modern or even too authoritarian by Perry and Leidenhag, at

the same time, Milbank's position on grace and nature has also been severely critiqued by some neo-Thomists for being too modern and too liberal (see Feingold 2010; Long 2010; Hütter 2012; cf. Healy 2008; Levering 2014). And, indeed, one might argue that the kind of neo-scholastic Thomist critiques of Milbank may be regarded as a kind of theological policing itself: policing not just in the sense of guarding the borders between grace and nature or between divine revelation and natural reason, but, to deploy Perry and Leidenhag's term, even in some authoritarian sense of using theological doctrines to police non-theological viewpoints and disciplines.

Indeed, such an account of the "authoritarian" use of theology could be found in Mark Jordan's 2006 book on Thomist theology, which opens with a chapter entitled "St Thomas and the Police," whose opening sentences are worth quoting in full:

> If only we could read Thomas Aquinas without encountering some other of his readers—especially the police.
>
> "The police" refers literally or figuratively. Figuratively we use the term to describe self-appointed guardians of social norms, as in "the decency police" or "the style police." Literally we use it to refer to the forces that keep internal order—municipal or state officers, the army on civic duty, and every other monitor or enforcer with the power of approved violence. . . . It is a remarkable fact about Thomas Aquinas's texts that they have been quoted so regularly by the police of various regimes—by papal or local inquisitors, of course, but also in service of Franco's victory in Spain or of the Argentine security forces during the 1970s and 1980s. (Jordan 2006, 1)

What we find in Jordan's critical depiction of the authoritarian scholastic or theological police is very much an inverse image of the kind of "secular policing" described by Perry and Leidenhag—as well as Milbank and Wood before them.[4] In this case, it is not religion or theology that is being "policed" by secular forces, but theology that "polices" the secular or indeed the world *qua saeculum*.

As opposed to Perry and Leidenhag's (2023, 2) "science-engaged theological" contention that the findings in natural sciences can serve as "correctives" to empirical claims that are sometimes made in theology, the theological police depicted and identified by Jordan argue that theology—and especially Thomist theology—can provide many scientific insights and indeed correctives for contemporary natural sciences (see Jordan 2006, 33–59). Whereas Perry and Leidenhag seek to remind theologians of "nature" and argue that nature can supplement or even occasionally correct theology, the theological police would hold that nature always needs superadded correction and regulation of grace and revelation, of the *sacra doctrina* as espoused by theological authorities such

as Aquinas. To paraphrase St Thomas himself, one might say that, for the "theological police," grace does not destroy nature but polices it (cf. *Summa Theologiae* I.1.8, *ad* 2).

The point of mentioning all this is not to attack Thomism or its neo-scholastic defenders and exponents. Instead, it is to highlight that theology is not always necessarily on the receiving end of policing by secular forces, but that theology itself can sometimes be (intentionally) used as a mechanism for policing—as Jordan points out, both figuratively and literally.[5] From this perspective, Perry and Leidenhag's "anti-authoritarian" theological vision can very much provide theologians with a model which can facilitate a kind of resistance against the type of theological policing mentioned above.

In the closing parts of their book, Perry and Leidenhag (2023, 54) argue that a science-engaged theology should be a "risky theology" and that "theologians already routinely make empirical claims, perhaps more often than they realize." Theologians are to take "risks" by allowing their claims to held accountable to empirical validation or verification. For according to them, "accountable theology is better theology" (Perry and Leidenhag 2023, 56). Instead of remaining content in appealing to a certain set of theological authorities and to deploy pre-set theological frameworks for gatekeeping or indeed policing, to do good theology in this view is not only to take risks, but to be "open to critique and accountability from others" (Perry and Leidenhag 2023, 54).

In Perry and Leidenhag's vision, science-engaged theology is not about subscribing to a particular set of methodological or metaphysical commitments or appealing to a particular school of thought or figure of authority. It is, as they say, "a disposition or mindset" of openness (Perry and Leidenhag 2023, 64). To engage theology with science is not a way of equipping theologians with scientific tools and theories to police others and close down conversations, but instead to open theological inquiry and conversations beyond their traditional silos. As opposed to using theological doctrine to police others who view the world differently or to police the disciplinary borders of theology, to practice science-engaged theology is to cultivate the "epistemic virtue" of "opening oneself up to multiple sources of correction" (Perry and Leidenhag 2023, 4).

But while Perry and Leidenhag's theological approach clearly differs from the kind of "theological policing" which Mark Jordan identifies among neo-scholastic Thomists, a question can perhaps be raised on whether Perry and Leidenhag's (2023, 32) pointed critique of "Milbank and countless others," who posit theology as the queen of the science may itself constitute a kind of "theological policing": one which ensures that theology stays within its "borders," putting theology back in its place within the modern secular world in the name of anti-authoritarianism and theological integrity. Indeed, echoing Deleuze and Guattari's (1987, 54) saying that "deterritorialization [always] has reterritorialization as its flipside or complement," if Perry and Leidenhag's

theological vision may be said to be an act of "theological *de*-policing" both against any assertion of theological dominance over other disciplines and against any secular policing over theology, one may more generally question whether all acts of *de*-policing may always have a kind of policing as its flipside or complement.[6]

We began this brief reflection by considering Perry and Leidenhag's metaphor of the European Union as a model for envisaging the borders between theology and other disciplines within the secular university and the secular world. Perry and Leidenhag speak of the EU as "*a trading zone* where national borders remain… but where there is also a shared set of standards allowing for increased collaboration and the relatively free movement of goods and peoples" (Perry and Leidenhag 2023, 47, emphasis added). But the EU is of course not just a trading zone: It is not just an economic body but also a "political" enterprise. To envision the relation between theology and other disciplines as national borders through the geo-political imagery of the EU is thus also, at least implicitly, a comment on the political aspects and implications of theological practice.

This short paper has sought to unpack some of the political dimensions of Perry and Leidenhag's *Science-Engaged Theology*. The division between science and religion in modernity, as Perry and Leidenhag point out, is as much a political phenomenon as it is a cultural and philosophical one. To the extent that their construal of science-engaged theology involves a twofold resistance towards "authoritarian" tendencies of asserting theology as the queen of the sciences as well as modernity's "border police" who seeks to exclude religion and theology from public discourse, Perry and Leidenhag's visionary and stimulating book brings to light one of the ways theology is sometimes deployed as a political mechanism for policing. In this way, though *Science-Engaged Theology* may not be written as a treatise in political theology, Perry and Leidenhag's meta-theological insights offer an important and timely intervention in our understanding of the place of theology in contemporary secular society as well as a political vision for the future of theological reflection and practice as theologians—"science-engaged" or not—seek to comprehend and engage with the ever-changing world.

Notes

[1] In addition to their worry about casting theology as the queen of the sciences, drawing on Arne Rasmusson's (2021) critical analysis of Milbank's political theological outlook (cf. Perry 2017), Perry and Leidenhag (2023, 54–56) also criticise Milbank for making empirical claims that lack empirical validity.

[2] See Milbank (1990, vi). In the second edition of *Theology and Social Theory*, Milbank (2006, viii) clarifies that his book is dedicated to the Christendom Trust, and "*not*, as some have mistakenly supposed, to the memory of the Middle Ages!"

[3] To the extent that God is Being itself—and not a being among other beings—and therefore exists in non-competitive or non-contrastive relationship with other beings, and that Christian Trinitarian theology posits that God is a triune being of harmonious difference, Milbank holds that (Christian) theology is not a discipline whose object of study is in competition with other disciplines and is in principle the only "master discourse" which can ensure that all other discourses can co-exist in harmonious difference. "This is why it is", Milbank ([1990, 6] 2006, 6) writes, "so important to reassert theology as a master discourse; theology, alone, remains the discourse of non-mastery." It is important to note that, for Milbank, theology is "the discourse of *non*-mastery"—following a *non*-competitive conception of the relation between God and creation, which forms the basis for a *non*-competitive account of the relation between theology and other discourses—and not of "*un*-mastery" (see Perry and Leidenhag 2023, 34–35), which would imply that theology would "*un*-master" any other rival discipline which seeks to assert "mastery" over theology. While Sarah Coakley's theological work is (understandably) not mentioned in Perry and Leidenhag's short book, it is perhaps worth adding here that the term "*un*-mastery" is notably deployed by Coakley (2013, 43) to describe her conception of theology explicitly in distinction from Milbank's.

[4] More recently, the leading Thomist scholar Brian Davies suggests that one reason why many use Thomas Aquinas' thought for "policing" other views may not be (solely) due to political or authoritarian incentives, but to what he describes as "academic laziness" (Davies 2020, 639).

[5] While Mark Jordan focuses on right-wing cases from the 1970s and 80s, one might also think of similar appeals to Aquinas in recent right-wing politics. See, for instance, the Conservative politician Jacob Rees-Mogg's surprising reference to Aquinas in support for his views on Brexit in his keynote address at the UK National Conservatism conference in May 2023: "St Thomas Aquinas, whose political work focussed on the state, in his fundamentally important explanation of a just war, he made it clear that legitimate authority belongs to the state rather than some ethereal international organization… I think I am the first person to claim Aquinas as a Brexiteer! It may surprise some of the Catholic theologians but I'm going to claim him as a Brexiteer anyway" (Rees-Mogg 2023, at 0m41s–1m00s and 31m58s–32m15s).

[6] Of course, one could by extension question whether interpreting Perry and Leidenhag's work as an act of "policing" would itself also constitute another act of policing.

References

Coakley, Sarah. 2013. *God, Sexuality, and the Self: An Essay "On the Trinity."* Cambridge: Cambridge University Press.

Davies, Brian. 2020. "Aquinas in His Place." *New Blackfriars* 101 (1096): 637–39.

Deleuze, Gilles, and Félix Guattari. 1987. *A Thousand Plateaus*. Translated by Brian Massumi. London: Continuum.

Feingold, Lawrence. 2010. *Natural Desire to See God According to St. Thomas and His Interpreters*. Ave Maria, FL: Sapientia Press of Ave Maria University.

Healy, Nicholas J. 2008. "Henri de Lubac on Nature and Grace: A Note on Some Recent Contributions to the Debate." *Communio: International Catholic Review* 35 (4): 535–64.

Hütter, Reinhard. 2012. *Dust Bound for Heaven: Explorations in the Theology of Thomas Aquinas*. Grand Rapids, MI: William B. Eerdmans.

Jordan, Mark D. 2006. *Rewritten Theology: Aquinas after His Readers*. Oxford: Blackwell.

Levering, Matthew. 2014. "A Note on John Milbank and Thomas Aquinas." *New Blackfriars* 95 (1059): 525–34.

Long, Steven A. 2010. *Natura Pura: On the Recovery of Nature in the Doctrine of Grace*. New York: Fordham University Press.

Milbank, John. (1990) 2006. *Theology and Social Theory: Beyond Secular Reason*. 2nd ed. Oxford: Blackwell.

———. (2005) 2014. *The Suspended Middle: Henri de Lubac and the Renewed Split in Modern Catholic Theology*. 2nd ed. Grand Rapids, MI: William B. Eerdmans.

Perry, John. 2017. "Grandpa Milbank." *Marginalia Review of Books*, October 13, 2017. https://themarginaliareview.com/grandpa-milbank/.

Perry, John, and Joanna Leidenhag. 2023. *Science-Engaged Theology*. Cambridge: Cambridge University Press.

Rasmusson, Arne. 2021. "Radical Orthodoxy on Catholicism, Protestantism and Liberalism/Liberality: On the Use of Historical Narratives and Quantitative Methods in Political Theology." *Modern Theology* 37 (1): 44–61.

Rees-Mogg, Jacob. 2023. "National Versus International Governance." *National Conservatism*, May 17, 2023. YouTube video, 33:34. https://www.youtube.com/watch?v=Jw3ZKh_B7Xo.

Tonstad, Linn Marie. 2020. "(Un)wise Theologians: Systematic Theology in the University." *International Journal of Systematic Theology* 22 (4): 494–511.

Wood, William. 2021. *Analytic Theology and the Academic Study of Religion*. Oxford: Oxford University Press.

Zygon®: Journal of RELIGION & SCIENCE

Boeré, Robyn. 2024. "Empirical Realities? A Question of Basic Principle." *Zygon: Journal of Religion and Science* 59 (3): 893–900. DOI: https://doi.org/10.16995/zygon.16820

Open Library of Humanities

Empirical Realities? A Question of Basic Principle

Robyn Boeré, Associate Professor, Faculty of Theology, University of Oslo, Oslo, Norway, r.e.boere@teologi.uio.no

This article is a response to John Perry and Joanna Leidenhag's book, *Science-Engaged Theology*. In this response, I focus primarily on the basic principle of science-engaged theology laid out by the authors on the first page of their book. Specifically, I invite both the authors and readers to think more deeply about (1) what it means to make scientific claims, (2) what it means for something to be an empirical reality, and (3) from what fields of study the insights of empirical investigation stem.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

John Perry and Joanna Leidenhag's book *Science-Engaged Theology* is an accessible and entertaining introduction to an emerging field of study, showing how the discipline of science-engaged theology is both traditioned and new. I can say with honesty that I enjoyed reading this little book, and if I have a complaint that is really more of a compliment, it is that the book is too short. It constantly teases a much fuller and more interesting story lurking behind each little titbit of information given to readers. Helpfully, in the book, which is available in both print and digital editions, they reference and link to many of these bigger stories.

In my response, I focus primarily on the basic principle of science-engaged theology laid out by the authors on page 1. Specifically, as it is obvious throughout the book that the authors have put considerable thought into this, I invite both the authors and readers to think more deeply about (1) what it means to make scientific claims, (2) what it means for something to be an empirical reality, and (3) from what fields of study stem the insights of empirical investigation. Even though it is obvious that Perry and Leidenhag have thought much about this, it is my contention that by the end of the book, it is still unclear what is meant by an "empirical reality." A less charitable way of characterizing my response is that I am here to quibble about the basic principle laid out on page 1. Here, the authors say the following: "The basic principle of science-engaged theology is that whenever theologians make claims about created, empirical realities, they should incorporate the insights of empirical investigation into their analysis" (Perry and Leidenhag 2023). This is, in my reading, related to the opening line of the second paragraph on the same page, where they say, "theologians make scientific, or science adjacent, claims all the time."

To quibble with the basic principle may seem perhaps churlish, or invite the speculation that a certain respondent did not manage to make it through the whole book. And, of course, a basic principle by definition cannot express nuance. Yet, these questions about the basic principle are important because the basic principle is what readers are going to take away long after they forget (or never read in the first place) the nuance of the following chapters. To be honest, I find myself asking these questions because I think they are fascinating and I do not myself clearly see an answer, even though I will offer a rough sketch of an alternate position.

In connection with the two aforementioned first-page quotes, I want to pose several, somewhat related, questions:

1. What is an empirical reality (as opposed to a nonempirical reality)?
2. What is empirical investigation, and how does this relate to empirical claims (as opposed to nonempirical investigation and nonempirical claims)?

3. What is a scientific claim, and how does it relate to empirical reality or empirical claims or the insights of empirical investigation? What is a science adjacent claim?

In addition to these questions, I must admit the desire to quibble about another term in the basic principle: whenever ("*whenever* theologians make claims about created, empirical realities"). But why? The scientific method(s) is one way of knowing about sin or an ocean or pregnancy or the human mind, but hardly the only way of knowing. There are also poetic, artistic, logical, intersubjective, and other modes of knowing. In fact, if one were to write a womanist account of theological anthropology, would one really want to turn to psychology, a field that not only has an acknowledged replicability crisis, but even more importantly, an acknowledged race and class issue? Similarly, there are good reasons to avoid scientific and medical data when writing on pregnancy. Is one really "pruning," the metaphor of choice in the section on accountability to the sciences in Chapter 5 (Perry and Leidenhag 2023, 57), or is one instead capitulating to a different lens that has, both historically and in the present, distorted women's realities and ignored women's experiences? While Perry and Leidenhag are right to reject Wittgensteinian fideism, it is nonetheless true that there are competing accounts of the human and that much if not most scientific and medical writings on pregnancy begin from the assumption that pregnancy is the gestation of a fetus in a container, not the intersubjective relationship between mother and child. What would empirical accountability then look like for that latter claim?

It is important to note that the authors are adamant that scientific methods of knowing are not the only ways, nor the only valid ways, of knowing, and they are adamant that theologians must remain theologians. To be clear for those who have not read the book, nowhere do Perry and Leidenhag suggest that knowledge gained from the sciences should get a trump card or that the sciences should be given a special verificatory role. In fact, the authors write: "Science-engaged theology does not imply that all theological claims need to be verified, or even corroborated, by the sciences in order to be meaningful or rational" (Perry and Leidenhag 2023, 4). But I still question the use of "whenever." That said, what is meant by this "whenever" depends significantly on what is meant by "created, empirical realities," and what it means to make claims about them. To say that theology as a discipline must engage with the sciences is different than saying every theologian must engage with the sciences on every topic of inquiry, but it is the latter that seems implied by the basic principle.

This is not to argue against the idea of the natural sciences as ready resources for theological study, including in womanist projects or theologies of pregnancy. The projects named and briefly described in the friendly blue boxes throughout this book beautifully demonstrate that recourse to scientific knowledge can be

very helpful and even necessary for many theological projects. A softer version of the basic principle is also given several times in this book: the authors encourage theologians to ask "what [scientific] methods or tools could help me improve this claim I am making about the world?" (Perry and Leidenhag 2023, 1, 8, 48). Nor is this to argue with Perry and Leidenhag's insight that many of our concepts are entangled, as they put it, "that cannot be understood as either scientific or theological in meaning and origin, but only as both" (2023, 13). As a side note, I appreciate that the authors strengthen this connection by naming the logical corollary to this, which is that scientists should then incorporate theological and philosophical reflections in their work (2023, 16).

Further, in posing these questions, it is necessary to acknowledge the reasons, which become obvious throughout the book, for why there is no definition for terms like "scientific" and "science" given in the introduction. As Perry and Leidenhag demonstrate throughout the volume, science (and theology, for that matter) is not a transhistorical and transgeographical *thing* whose universal essence can be found and objectively defined. Thus, there is no simple definition of "science," nor even "the scientific method," to be offered up here without vastly misrepresenting the regional, historical, and disciplinary complexities of the natural sciences.

But while the authors of this volume spend considerable time nuancing the definition(s) of science, and exploring some of these historical, regional, and disciplinary complexities, they do not explain what it means to make a scientific claim, and they leave essentially uninvestigated the meaning of "empirical," "empirical reality," and "empirical investigation."

To lay my cards on the table in a simplified form, I present the following argument, which is perhaps an alternative or counter-position to the basic principle laid out in this volume: that what makes something a scientific claim is not the subject matter—whether sin, anthropology, or ecclesiology—but rather that the claim is made using (or based on) scientific methods. Thus, against Perry and Leidenhag's (2023, 1) claim on the opening page of their book that "theologians make scientific, or science adjacent, claims all the time," I argue that theologians rarely make scientific claims, even though they make claims about objects and subjects about which the sciences also make claims. (As an aside, is that what they mean by science adjacent?) Similarly, perhaps what makes an empirical claim is that it is made using empirical methods, though it is not yet clear what it is meant by empirical methods. Perhaps a better way of asking it is this: Is "empirical" a type of reality or is it a method(s) of coming to know? Is it a category of objects or a category of epistemology? The Peter Godfrey-Smith passage they quote suggests the latter (Perry and Leidenhag 2023, 50).

Here is another way of saying this: scientific (or empirical) knowledge is really helpful, depending on the question. It is the question that is the operator, not the "reality." This is because, of course, reality is not given in

experience but in answers to questions. Unless we are naive realists, reality is not already out there waiting for us to discover, as it were. There is abstraction that does not make recourse to empirical datasets; however, it is not the reality (object, topic) about which one is asking that is or is not empirical but rather the mode of questioning.

So let us return to those questions I raised near the beginning of this response. First, what is a "created, empirical reality?" Is there a difference between created and empirical? That is, is there anything created that is nonempirical? (I.e., is the empirical/nonempirical simply the same as creator/created? If so, why say both in the basic principle?) What would be the alternatives? Is there also uncreated, empirical reality? One could argue, perhaps, that God is an uncreated, empirical reality, in that God is uncreated, and people can experience God. In fact, people even use scientific methods to investigate these experiences. Why, then, is this not included in the basic principle? Is there, on the other hand, created, nonempirical reality? Here, one wonders what this would be—a created reality that cannot be experienced? But is there any reality that cannot be experienced?

This second example especially highlights the third question raised, which is about the relationship between scientific claims and empirical claims. Surely any person who has experienced God can make claims that rely on those experiences (the private, interior realities discussed in Chapter 5, for example). And people historically have made such claims, whether expressing their visions or artistically expressing their feelings of relationship with God, and so on. It is safe to say, though, that we cannot investigate God by the use of scientific methods but only the human experience thereof. Is this true then of all empirical methods, or just scientific ones?

What, then, is an "empirical reality"? Is it the same thing as just "reality"? In the final chapter, Perry and Leidenhag (2023, 64) offer the following examples of (theological) claims about empirical realities: "Spiritual practices, character formation, claims about sin, claims about other creatures, claims about church polity, liturgy, or sacraments, claims about the incarnation, about birth and death." Thus, we can reason that, according to this, the following are "empirical realities": spiritual practices, character, sin, other creatures, church polity, liturgy, sacraments, the incarnation, birth, and death.

However, two pages earlier, in a footnote, they characterize the claim made by some Thomists that God infuses the soul with feminine and masculine attributes that remain independent of, or perhaps an objective reality even among, the complexities of sex and gender, or rather as an excuse to ignore those complexities, as nonempirical (Perry and Leidenhag 2023, 62n189). Here, interestingly, Perry and Leidenhag offer this as an example of a nonempirical sign. Is this (the sign) the same as a reality or as a claim? But what makes this nonempirical? The context suggests that it is because it is not verifiable by empirical methods. But this presents a problem, I would think, for the

previously given list because many of the claims we make about birth and death and sacraments and so on are not empirically verifiable (at least according to scientific methods). So, what exactly makes this nonempirical? Is anything we say about the soul non empirical? Is the soul a nonempirical reality? Or would this example then be characterized as a non empirical claim about empirical realities? But again, what makes the object of knowledge here empirical?

Let us say, perhaps, that all created reality is empirical. There is good reason to make this claim, for is there any part of reality that we cannot experience? The list Perry and Leidenhag give in this chapter is pretty all-encompassing. Not only that, but as the authors make clear, our concepts are frequently (always?) entangled. It is hard to think of a concept that is not entangled, in that it would not be related to both scientific and theological claims insofar as they are claims about reality. This would mean, for the basic principle elucidated on page 1, that every claim a theologian makes ought to incorporate the "insights of empirical investigation." Is it really the case that when I want to say anything about the world theologically, I should or must include the sciences? But it does not seem like the authors want to make this claim, or maybe I misunderstand, when they say that "[s]cience-engaged theology does not imply that all theological claims need to be verified, or *even corroborated*, by the sciences in order to be meaningful or rational" (Perry and Leidenhag 2023, 4, emphasis added).

Left unanswered still is the relationship between this verification or corroboration by the sciences with the insights of empirical investigation. In other words, this leads back to the question about the relationship between empirical and scientific: not all empirical investigation, one would think, is scientific investigation. Surely a painter painting a scene is also conducting an empirical investigation. So are philosophers and theologians when they reason on an *a posteriori* basis. So, does this basic principle also apply to the arts, for example? The artistic modes of coming to know represent an important depiction of empirical reality that can maintain a multiplicity of meanings, which are inherently more flexible and all-encompassing than scientific claims. I imagine that Perry and Leidenhag would not mind this as a corollary basic principle, though outside the scope of this work, since they speak many times of how theology relies on many different sources. Theology is inherently promiscuous.

But to return to this question of the relationship between the empirical and the scientific, when looking over the many instances of "empirical" throughout this book, the implication, I believe, whether intended or not, is that "empirical" and "scientific" mean the same thing. One can look, for example, to passages where the authors use such terms as "empirical sciences" and the "methods of empirical investigation" contrasted with the methods of scriptural investigation, and again "empirical studies" (Perry and Leidenhag 2023, 52–54, 63). In these contexts and others, the authors switch back and forth between "science/

scientific" and "empirical" in a way that suggests that these terms mean the same thing. So, while Perry and Leidenhag are commendably careful not to talk about "scientific realities," at least so far as I could find in the book, and have a careful and attentive nuanced view of science, they essentially smuggle in the concept of scientific reality and its attendant scientism through the use of the word empirical. One of the questions I put to them, then, is whether they would consider this paraphrase of the basic principle acceptable: "[w]henever theologians make claims about created, *scientific realities*, they should incorporate the *insights of scientific investigation* into their analysis."

By way of conclusion, then, I simply repeat the proposal made earlier: there is no such thing as a "scientific reality" nor "empirical reality" as if these were a question of object. That is, there is no object whose investigation makes it inherently scientific or empirical. Rather, what makes something a scientific claim is not the subject matter (whether sin, anthropology, or ecclesiology) but rather that the claim is made using scientific methods. About what exactly an empirical reality is I remain (curiously and delightedly) in the dark.

Reference

Perry, John, and Joanna Leidenhag. 2023. *Science-Engaged Theology*. Cambridge: Cambridge University Press.

Zygon®: Journal of RELIGION & SCIENCE

Leidenhag, Mikael. 2024. "Considering the Place of Apophaticism within Science-Engaged Theology." *Zygon: Journal of Religion and Science* 59 (3): 901–11. DOI: https://doi.org/10.16995/zygon.16735

Open Library of Humanities

Considering the Place of Apophaticism within Science-Engaged Theology

Mikael Leidenhag, Science and Theology Editor, School of Divinity, University of St Andrews, Fife, UK, mikael.leidenhag@st-andrews.ac.uk

Despite its firm embeddedness in the Christian tradition, apophatic theology has been strangely absent in the science-and-religion dialogue. Given that the apophatic theologian eschews the idea that we can fully comprehend God, or positively articulate God's being, this may not be too surprising. Indeed, the model-based logic of science and religion and its ambition to articulate, for example, divine activity through the latest scientific advancements seems to run contrary to the spirit of apophaticism. This article discusses whether Perry and Leidenhag's science-engaged theology may provide a more suitable home for those who emphasize the ineffability and mysteriousness of God's being.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

Introduction

Since the days of Ian Barbour's groundbreaking *Issues in Science and Religion* (1966), the wider field of what is often labelled "science-and-religion" has focused on the possibilities and challenges of harmonizing theological doctrines and beliefs with the deliverances of the natural sciences. As Perry and Leidenhag discuss in the first section of their book, this has yielded a number of impressive projects and volumes over the last 40 years or so. A main area of focus within this rich research discourse has been *divine action*: given what we know about the causal structure of the universe, the evolutionary development of life, and the complexities of the human mind, how is it possible to claim from a theological perspective that God acts in the world and interacts with humanity? If the world is governed by non-teleological laws and regularities, what place is there for God to act *purposefully* in the natural world?

A number of theological models have been devised to address this perplexing question. Some of these models are based on emergence theory and neuroscience (Clayton 2004; Peacocke 1993), others on process metaphysics (Griffin 2004), and still others on quantum mechanics (Russell 2006; Polkinghorne 1988; 2008), etc. There is no shortage of proposals as to how divine activity may be construed through scientific categories. However, a notable absent voice in this discussion has been apophatic theology, despite its firm embeddedness in the Christian tradition. In one sense, this is no surprise given that the apophatic theologian eschews the idea that we can fully comprehend God, or positively articulate God's being. Hence, "human beings *cannot* speak of God […] and cannot truthfully, or even intelligibly, say *what* God is" (Simon Hewitt 2020, 3). This seemingly runs contrary to the dominant spirit of science-and-religion, which seeks to construct models, depicting how divine activity can be positively understood through current science.

My first argument is that the models-based science-and-religion, despite its ambitions, ends up inviting scientism; that is, by understanding divine causality as equivalent to scientific causality, one ends up surrendering theology to the culture of scientism. In this way, the ontological deflationism of apophaticism should be of significant interest as it may avoid such scientistic implications.

Second, I evaluate the ability of science-engaged theology (SET), as primarily expressed in John Perry and Joanna Leidenhag's *Cambridge Element*, to include apophatic theology. My argument is that the epistemological flexibility of SET can more easily (a) give apophatic theology a seat at the table and (b) avoid the problem of scientism.[1] Indeed, Perry and Leidenhag (2023, 2) emphasize that they "do not want science-engaged theology to become a backdoor for naïve scientism within theology." At the same time, I also pose the question as to whether SET might nevertheless commit itself to a similar form of "the metaphysical matching game" through its notion of entanglement?

The Problem of Playing the Metaphysical Matching Game

Many philosophers play what Huw Price (2011, 3) calls a "metaphysical matching game": "Imagine a child's puzzle book . . . The left-hand page contains a large sheet of peel-off stickers, and the right-hand page shows a line drawing of a complex scene. . . . For each sticker . . . the reader needs to find the unique outline in the drawing with the corresponding shape." The aim of this game, thus, is to place all the stickers in their correct place. This, Price continues, bears a striking resemblance to the current game of naturalistic philosophers. They take a statement that we hold to be true about the world (the sticker), which we then try to place in its correct location (the scientific story about the world). Yet, some statements are extremely difficult to fit in such a naturalistic picture of the world, and thus we get what Price (2011, 188–89) calls a *placement problem*.

Although this argument, or observation, is made within the context of secular, naturalistic philosophy, I want to make the case that previous and current attempts to square divine action with the deliverances of modern science lead to similar "placement problems" and contradictions. Such divine action theories are in abundance and there is no space to survey all of them, so here I restrict the focus to Philip Clayton and Arthur Peacocke's emergentist conception of divine activity, and Sarah Lane Ritchie's theistic naturalism.[2]

Albeit in slightly different ways, both Peacocke and Clayton rely heavily on emergence theory in developing a model of divine action whereby God acts through natural processes in a non-competitive manner. An emergentist understanding of our world, according to Peacocke and Clayton, commits itself to a broad "monism," in that the world is made out of stuff (Clayton 2004, 60) and no "extra *entities* or *forces*" are needed to account for the natural world (Peacocke 2007, 12). However, despite such a general monism, this view also suggests that reality is ultimately "layered," such that higher levels are irreducible to lower levels; for example, the behavioral sciences are irreducible to chemistry or fundamental physics. This view resists any attempt at regarding a higher-level entity, such as the human mind, "as less real in comparison with some favored lower level of 'reality'" (Peacocke 2007, 13). Similarly, Clayton (2004, 60–62) recognizes eight characteristics of emergence, one of which is the belief in "hierarchical complexity," meaning that "more complex units are formed out of more simple parts." Moreover, and this is perhaps the key claim of strong emergence theory, higher-level properties can exert downward or top-down causation on their lower-level constituents. For example, mind can exert downward influence on the body (see Clayton 2004, 62; Peacocke 2006, 266–73). It is here that Clayton and Peacocke, in constructing a panentheistic vision of God, make the argument that God can exert downward influence on the world through natural processes, in a similar way to the mind-body relationship, without breaking the metaphysical commitment to monism—without this entailing

divine interventionism. Although Clayton and Peacocke's emergentist proposal provides a valuable pushback against reductionism, it also raises significant metaphysical problems. As Scot Yoder (2015) and I (Leidenhag 2014; 2021) have argued, strong emergence and the notion of downward causation ultimately break the monistic commitment of causal closure, thus introducing the sort of dualism that both Clayton and Peacocke consider scientifically untenable.

It is clear that both Clayton and Peacocke, by virtue of their monistic commitment, are forced to play their own metaphysical matching game; affirmations about God's active involvement in the natural world (the sticker) have to be placed in their correct locations (the emergentist construal of the world). Yet, this effort gives rise to a "placement problem" as it is difficult to square divine action with a purely monistic understanding of emergentism.

Although Ritchie, in *Divine Action and the Human Mind* (2019, 85), vows to go beyond such emergentist proposals—which she deems not "naturalistic enough"—her account faces somewhat comparable placement issues. Overall, Ritchie (2019, 37–38) adopts a similar monistic outlook by virtue of affirming *causal closure*, "the methodological assumption for working scientists" that stipulates that "all physical events have physical causes." Hence, the notion that God can directly intervene in the natural order is considered "blatantly unscientific" (2019, 37), as it would "undermine current and future scientific practice" (2019, 37, 38). Ritchie (2019, 350), while reserving most of this book for evaluating current divine action proposals, gestures towards how a theistic naturalism may provide a more fruitful framework for imagining divine activity in a world of science, according to which "the physical is involved with God's active presence." In order to avoid a possible clash between divine action and the causal closure principle, Ritchie (2019, 347) is keen to push back against a strict dichotomy between the natural and supernatural: "To be natural is to be involved with God." Although Ritchie has wisely avoided making the plausibility of divine action dependent on any one scientific model (including emergence theory), her theory raises similar issues. On her model, it sounds as if divine activity and natural processes are concurrent. That is, divine activity, G, and natural regularities, N, both sufficiently produce an event, E. However, in this case, neither G nor N are explanatorily necessary for E, as E would have obtained in the absence of G and in the presence of N, or in the absence of N and in the presence of G. Depending on how one looks at the situation, either G or N would be redundant; we would have no way of telling. However, if this account allows for E to obtain through G in the absence of N, then this account would contradict Ritchie's commitment to causal closure. So, either divine activity might be superfluous, or it contradicts causal closure. Ritchie avoided one sort of metaphysical matching game but has encountered another serious placement problem whereby divine activity either becomes superfluous or it breaks causal closure.

This is not to say that Ritchie's proposal ultimately fails as the short argument above clearly does not accomplish this. Instead, given these problems of Clayton/Peacocke and Richie's proposals I want to raise the broader question, "why engage in such metaphysical matching games at all?" By playing this game, we risk subjugating theology to naturalism, secularism, and scientism.

Taede Smedes (2004, 207) has developed such a critique in relation to various attempts at reconciling divine action with science, suggesting that several notable theories of divine action fail to respect and understand the "internal logic" of religious language and theology. In fact, by subjugating theological claims regarding divine activity to "criteria of meaning" of naturalistically conceived scientific theories, we are letting scientism in through the backdoor. Such scientism distorts the grammar of theology by suggesting that divine action must be understood in terms of scientific causality, such that God's action is wrongfully taken to be "similar to creaturely, causal and/or intentional action, including the limitations inherent to the creaturely condition" (Smedes 2004, 207). The implications of Smedes's diagnosis are clear; by playing the metaphysical matching game, seeking to articulate divine activity through the latest scientific advancement, we are making a category mistake, thus forcing theology into the logic of scientism.

Going beyond the Metaphysical Matching Game: Apophaticism

Through a brief overview of some attempts at reconciling divine action with modern science, we have seen how playing the metaphysical matching game leads to placement problems due to the ways in which these models tacitly assume a scientistic criterion of meaning. One way to avoid the metaphysical matching game altogether is to adopt an apophatic approach to divine action.

Here, I briefly unpack some core features of apophatic theology, before relating it to SET. Simon Hewitt (2020, 13) aptly points out that negative theology is part and parcel of the tradition, and not an alternative way of framing Christian theology; indeed, it should be understood as "working out of Christian theology in an attempt to be faithful to the tradition." Although apophatic theology (AT) lacks a singular definition, it typically comprises a few key ideas regarding our (in)ability to describe God through human language. As indicated in the introduction, the main thrust of AT is that God is beyond human language. Here, we can quickly see how AT relates to divine ineffability, which suggests that humans are unable to describe God in words (Hewitt 2020, 13). Yet, as Jonathan Jacobs (2015, 165) points out, "it is not our limitations that ground God's ineffability." It is not that we may be able to bridge this epistemological gap in the future, perhaps with the help of a supercomputer or if human beings evolved towards greater cognitive capacities. Rather, it is God's "*transcendence* that grounds his ineffability" (Jacobs 2015, 165, emphasis added). Hewitt (2020, 14) makes a similar point when he suggests that the doctrine

of divine ineffability is closely related to (although distinct from) the doctrine of divine transcendence, which adds the further claim that "God is radically dissimilar from creatures."

This is not to say that God and human beings merely differ in some number of properties. More strongly, this doctrine suggests that "there is no intelligible basis for comparison between God" and creatures (Hewitt 2020, 14). Given that God does not share a logical space with human creatures, it is a "category mistake" to apply a predicate associated with human creatures to God given God's radical dissimilarity (Hewitt 2020, 101). Of course, much more can be said about the apophatic tradition, but this short description suffices to identify a seeming tension between AT and the above-mentioned divine action proposals.

From an AT perspective, which has already been touched upon through our discussion of Smedes, these proposals commit a category mistake by applying scientific accounts of causality, and what we otherwise know about human beings and consciousness, on divine activity.[3] Indeed, the purpose of these accounts is not simply to shed light on how God interacts with the world and creatures, but they assume that the veridicality of theological beliefs in divine action and providence depends on them being successfully articulated through scientific theories or accommodated within, broadly speaking, naturalistic discourses that take a general monism to be true (as could be seen in Clayton's, Peacocke's, and Ritchie's frameworks). AT, given its emphasis on the limitations of human language due in the face of divine transcendence, pushes back against the metaphysical matching games being played in science-and-theology, which assume, to use Hewitt's phrase, that God and the world share a "logical space." In refusing to play this game, AT can, furthermore, bypass the scientistic implications of subjugating theological claims to secular and naturalistic presuppositions.

Exploring the Relationship between Apophatic Theology and Science-Engaged Theology

As has been seen so far, playing the metaphysical matching game inadvertently invites a scientistic understanding of the relationship between science and theology, rendering the latter epistemically subordinate to the former, whereby theological beliefs need to be accommodated within a secular and naturalistic discourse. AT, an underrepresented voice in this debate, emphasizes the particularity of the logical space of Christian theology, and so rejects the methodological principles underlying the metaphysical matching game. Here, I probe three issues: (1) Does SET avoid the metaphysical matching game? (2) Can SET give apophatic theology a seat at the table? (3) Can there be a science-engaged apophatic theology?

In analyzing the first issue, let us start off with a key feature of Perry and Leidenhag's (2023, 16) articulation of SET; namely, its recognition of "the

gospel of complexity." Drawing on the historical work of John Hedley Brooke, Peter Harrison, and David Livingstone, Perry and Leidenhag (2023, 19) seek to deconstruct the monolithic terms of "science" and "religion," showing their historical situatedness by concluding that "the search for a definitive relation between science and religion is a dead end." Indeed, modernity merely used those concepts "as badges or ciphers to separate rationality from irrationality" (Perry and Leidenhag 2023, 21). In emphasizing the irreducibility of "science" and "religion," SET, furthermore, opposes unificationism and the attempts of unifying "scientific disciplines and methods into 'one thing'," as most clearly manifested in A. J. Ayer's logical positivism and, to some extent, Karl Popper's falsificationism (Perry and Leidenhag 2023, 36, 38). As a result of such scientific unificationism, theology is relegated to the realm of irrationality unless such beliefs and doctrines could, somehow, be accommodated within the logic of scientific methodology, be that either Ayer's or Popper's criterion for scientific acceptability. Here, we can see how the spirit of pluralism may function as "an equalizing force in interaction between scientific and theological research" (Perry and Leidenhag 2023, 40). As a result of such a pluralist understanding, Perry and Leidenhag break with the problematic idea "of a single map for the territories of science and religion" (2023, 42). Without firmly committing to it, Perry and Leidenhag's (2023, 42) argument leads them towards Ronald N. Giere's scientific perspectivism, suggesting that different scientific instruments and theories provide different, partial perspectives of the world, and that whether something is the "best" perspective cannot be universalized but must be evaluated in relation to the goal at hand or "what type of question one is seeking to answer."

Here, we can see how SET, in stressing the pluralistic nature of both scientific practice and theological discourse, avoids partaking in the metaphysical matching game. In the same way that Perry and Leidenhag (2023, 38) stress that science and religion do not necessarily share the same one map, apophatic theology would object to the logic of the metaphysical matching game by stressing that scientific theories and theology do not engage in the same game or play by the same rules, and that scientific and theological claims cannot be housed within a "single logical language system," This breaks clearly with the divine action proposals of Clayton, Peacocke, and Lane Ritchie, which render the meaningfulness of theological claims dependent on scientific accommodation. Thus, if theologians fail in playing the metaphysical matching game—such that a theological claim cannot be matched with a scientific fact—this would undermine the rationality of theology. To put this in another way, unlike SET, these divine action proposals assume that theology (divine action) and science (physical and mental causation) share the same map or play by the same rules. It is such a unificationist endeavor that opens the backdoor for scientism within theology by insisting on a "neat and wholistic translation from the language of one discipline to that of another" (Perry and Leidenhag 2023, 44).[4]

However, by refusing to play the metaphysical matching game, does SET not simply succumb to relativism and isolationism? Is SET trading scientism for quietism? Here, Perry and Leidenhag (2023, 64, 43, 44) are careful to emphasize the possibility of constructively engaging with science, but in a "highly localized way" within particular "trading-zones" where local customs and norms can be negotiated. There are still better and worse ways of doing SET as "not all maps are equally good" (Perry and Leidenhag 2023, 42).

AT and SET share in their refusal to play the metaphysical matching game, conceding that theology and science do not necessarily share the same logical space and that any attempt at globalizing their relationship in terms of a metanarrative should be rejected; theology should not be uncritically subjected to or evaluated in terms of the language, norms, and rules of other disciplines. At the same time, AT and SET arrive at these conclusions partly due to different starting points. AT's rejection of the metaphysical matching game is grounded in its affirmation of the ineffability of God and the radical dissimilarity between God and created phenomena, whereas SET follows the "gospel of complexity" in problematizing transhistorical notions of science as a stable and monolithic entity.

Can there then, given this significant area of agreement, be a science-engaged apophatic theology? In answering this, it is worth revisiting a challenge against SET as such, identified by Perry and Leidenhag (2023, 46, 66): that is, by rejecting the narrative of unity and preserving the distinctive voice, or "native tongue," of theology, does SET not simply close off the possibility of meaningful dialogue? No. Again, it is possible to engage with science by staying local and specific, such that theological claims are not forced into some "mythological categories or metanarratives" (2023, 65), including causal closure or methodological naturalism. It is, similarly, possible to engage science from an apophatic perspective without committing "the category mistake" of ascribing properties of the natural order to God or determining the meaningfulness or rationality of theological claims on the basis of non-theological norms and criteria. One can still point towards consonance between a theological doctrine and a scientific theory, in the sense that there is no logical connection between the two and the former does not need to be accommodated within the latter, but they may "fit in with each other very well" (van den Brink 2020, 226). As such, AT is not some "*NOMA on steroids*" (Perry and Leidenhag 2023, 5).

Although there is significant agreement between SET and AT in terms of how the science-theology dialogue should *not* be conducted, some claims of SET may potentially cause concerns among those theologians who emphasize metaphysical dissimilarity between God and the created order. In avoiding the conclusion that SET merely amounts to "empirical fact checking" (Perry and Leidenhag 2023, 57), or that it constitutes a sub-category of the dialogue or integration model (2023, 14), the authors stress the *entanglement* of theological and scientific concepts, whereby a concept "cannot be understood as either

a scientific or theological in meaning and origin, but only as both," meaning that "the tools of more than one discipline" is needed to understand the phenomenon (2023, 13, 48). This may go contrary to the apophatic spirit of safeguarding the unique logical space of Christian theology, if the idea is that scientific and theological concepts ought to be used conjunctively to account for some phenomenon, thus, seemingly implying that scientific and theological operate on the same level of reality or that they share the same logical space. My question to Perry and Leidenhag is: How should we understand this notion of entanglement, and may it provide a novel idea beyond that of sheer "compatibility" between science and religion, or might it unintentionally lead us back to the perilous metaphysical matching game?

Acknowledgements

Parts of this work appear in Leidenhag, M. 2024. "Can There Be an Apophatic Science-engaged Theology?" *Theology and Science* 22:1–15. https://doi.org/10.1080/14746700.2024.2436777.

Notes

[1] I adopt an approach of constructive criticism for two reasons. First, I have been part of framing the prospects of science-engaged theology through managing an online resource in this area; see https://www.theo-puzzles.ac.uk/. In fact, Leidenhag and Perry's book utilises several theological puzzles to exemplify the current trajectory of SET. Second, and most importantly, I want to maintain the peace at home.

[2] It should be noted that Lane Ritchie (2018) has been crucial to the development of science-engaged theology, through co managing grants in the area and through a co written article (with John Perry), in which she sought to situate SET in relation to critiques of methodological naturalism.

[3] See also Smedes's discussion of "category mistakes" in a response to Ian Barbour (2008).

[4] On page 49, Perry and Leidenhag (2003) make a brief connection between unificationism and scientism; "Given the arguments made in the preceding sections, we cannot rely on scientistic assumptions regarding the superiority or unity of science."

References

Barbour, Ian. 1996. *Issues in Science and Religion*. New York: Vantage.

Clayton, Philip. 2004. *Mind and Emergence: From Quantum to Consciousness*. New York: Oxford University Press.

Griffin, David R. 2004. *Two Great Truths: A New Synthesis of Scientific Naturalism and Christian Faith*. Louisville, KY: Westminster John Knox Press.

Hewitt, Simon. 2020. *Negative Theology and Philosophical Analysis: Only the Splendour of Light*. London: Palgrave Macmillan.

Jacobs, Jonathan. 2015. "The Ineffable, Inconceivable, and Incomprehensible God: Fundamentality and Apophatic Theology." In *Oxford Studies in Philosophy of Religion*, vol. 6, edited by Jonathan Kvanvig, 158–76, Oxford: Oxford University Press.

Lane Ritchie, Sarah. 2019. *Divine Action and the Human Mind*. Cambridge: Cambridge University Press.

Leidenhag, Mikael. 2014. "Is Panentheism Naturalistic? How Panentheistic Conceptions of Divine Action Imply Dualism." *Forum Philosophicum* 19 (2): 209–25.

———. 2021. *Naturalizing God? A Critical Evaluation of Religious Naturalism*. New York: SUNY Press.

Peacocke, Arthur. 1993. *Theology for a Scientific Age: Being and Becoming—Natural, Divine, Human*. Minneapolis: Fortress Press.

———. 2006. "Emergence, Mind, and Divine Action: The Hierarchy of the Sciences in Relation to the Human Mind-Brain-Body." In *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, edited by Philip Clayton and Paul Davies, 266–73. Oxford: Oxford University Press.

———. 2007. "Emergent Monism." In *All That Is: A Naturalistic Faith for the Twenty-First Century*, edited by Philip Clayton, 12–16. Minneapolis: Fortress Press.

Perry, John, and Joanna Leidenhag. 2023. *Science-Engaged Theology*. Cambridge: Cambridge University Press.

Perry, John, and Sarah Lane Ritchie. 2018. "Magnets, Magic, and Other Anomalies: In Defense of Methodological Naturalism." *Zygon: Journal of Religion and Science* 53 (4): 1064–93.

Polkinghorne, John. 1988. "The Quantum World." In *Physics, Philosophy and Theology*, edited by Robert J. Russell, 333–42. Vatican City: Vatican Observatory.

———. 2008. *Quantum Physics and Theology: An Unexpected Kinship*. New Haven, CT: Yale University Press.

Price, Huw. 2011. *Naturalism Without Mirrors*. Oxford: Oxford University Press.

Russell, Robert J. 2006. "Quantum Physics and the Theology of Non-Interventionist Objective Divine Action." In *The Oxford Handbook of Religion and Science*, edited by Philip Clayton and Zachary Simpson, 579–93. Oxford: Oxford University Press.

Smedes, Taede A. 2004. *Chaos, Complexity, and God: Divine Action and Scientism*. Leuven: Peeters.

———. 2008. "Taking Theology and Science Seriously Without Category Mistakes: A Response to Ian Barbour." *Zygon: Journal of Religion and Science* (40): 271–76.

van den Brink, Gijsbert. 2020. *Reformed Theology and Evolutionary Theory*. Grand Rapids: MI: William B. Eerdmans Publishing Company.

Yoder, Scot. 2015. "Emergence and Religious Naturalism: The Promise and Peril." *American Journal of Theology and Philosophy* 35 (2): 153–71.

Zygon®: Journal of RELIGION & SCIENCE

Perry, John, and Joanna Leidenhag. 2024. "Replies to Some of Our Friendlier Critics." *Zygon: Journal of Religion and Science* 59 (3): 912–22. DOI: https://doi.org/10.16995/zygon.16729

Open Library of Humanities

Replies to Some of Our Friendlier Critics

John Perry, Senior Lecturer in Theological Ethics, University of St Andrews, St Andrews, UK, jmp24@st-andrews.ac.uk

Joanna Leidenhag, Associate Professor in Theology and Philosophy, School of Philosophy, Religion and History of Science, University of Leeds, Leeds, UK, j.leidenhag@leeds.ac.uk

In this Book Symposium, Dirk Evers, King-Ho Leung, Robyn Boeré, and Mikael Leidenhag have discussed John Perry and Joanna Leidenhag's book *Science-Engaged Theology*. Here, Perry and Leidenhag respond to their comments.

Zygon: Journal of Religion and Science is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

OPEN ACCESS

Introduction

What a privilege to have four excellent scholars give such careful attention to our book; we are deeply grateful. Each response is appreciative, but also identifies the intersecting points where we needed to say more on what we mean by key terms like secularism, policing, empirical, and entanglement. This Book Symposium started as an "Authors Meet Critics" panel at the European Academy of Religion in St Andrews 2023, and so we thank the EUAoR and *Zygon: Journal of Religion and Science* for facilitating and publishing this conversation, as well as Dr Danielle Jansen for editorial assistance.

To help us zero in on the most fruitful passages, we have selected one or two direct quotes from each response. We do this both to keep our rejoinders narrowly focused and also as a way to honor our critics' words. They have developed our thinking on these matters, and as a way to show gratitude, we will now dive more deeply into some of their salient points.

Response to Dirk Evers

Let us start with Dirk Evers, who pushes us to think more about historical contextualization in several different ways.[1] We focus here on the first such way, when he writes:

> Here as well as at other places Perry and Leidenhag refer to a meta-narrative of secularization. . . . However, the term secularization or secularism is not so much a descriptive historical category as it is an analytical term which became common with Sigmund Freud, Max Weber, Émile Durkheim, Peter L. Berger, and others as a characterization of the development of modernity away from religious dominance in society. (Evers 2024, 872)

What Evers objects to is those places where we speak of secularism as an ideology as opposed to the description of a historical trend, as we do in the opening lines of Chapter 3: "the deconstruction of modern terms like 'science' is necessary because in this conversation the terms were invented as tools of an ideology called secularism" (Perry and Leidenhag 2023, 21).

In our book, we tried to reserve secularism (note the *-ism* ending) to refer to the view that orthodox Christianity will (and should) inevitably die out to be replaced by something, allegedly neutral and presuppositionless, called science. Thomas Jefferson is the poster-boy for this view, but he is not the only one; two others that we mention in the book are Voltaire and, much more recently, Russell McCutcheon. We needed some word to denote this "Jeffersonian" notion, and we chose secularism.

Secularism in this sense is conceptually distinct from the theories of secularization proposed by the sociologists listed by Evers above, to which we may add other thinkers such as Charles Taylor, Alasdair MacIntyre, and José Casanova.

But we still needed some word to refer to whatever Jefferson and Voltaire were up to. Evers does not like our choice but does not propose an alternative. For what it is worth, the usage of both Berger and Taylor sounds closer to our terminology. Berger (2008) counts three different forms of secularism, even referring to them as ideology, as we do. Taylor (2018, 1–24) distinguishes Secularity 1 (akin to our secularism) from Secularity 2 (declining religious practice) from Secularity 3 (Taylor's own focus in *A Secular Age*). However, hidden behind Evers's quibble about terminology is a much more interesting argument. Reading his response gave a new handle on something that was in the back of our minds while writing the book but which we could never put into words until now.

There has been a swinging pendulum in science and religion historiography. The pendulum swings between (1) emphasizing meta-narratives as explanations for the relationship between science and religion (usually simplistic stories, like inevitable conflict or harmony) and (2) emphasizing complexity. Everyone involved in the discussion in these pages is broadly on the side of complexity, including Peter Harrison, Evers, and us. But simply asserting "It is complex," and nothing more, is not really an explanation of anything. As Harrison (2019, 223) writes, "while highlighting historical complexity is entirely appropriate as a first step, historians of science and religion need to do more than simply identify historical episodes that falsify unpalatable narratives." Ronald Numbers (2019, 235) puts the point sharper: "complexifying history seems to have little to recommend it besides its truth." Again, we think we are on the same page here as Evers. But what comes next?

For Harrison, what comes next is a swing of the pendulum back from complexity toward, ever so slightly, patterns or meta-narratives as partial explanations. He continues, "the primary task of the historian is not the searching out of complexity per se, but the attempt to render complexity intelligible as far as that is possible" (Harrison 2019, 226). So, historians must do more than simply tell episodic stories from long ago. We agree. However, what gave us pause in the original writing was the insight that, among all explanatory patterns, meta-narratives of secularization are somehow different (see Perry and Leidenhag 2023, 9). During the writing of our book, we well remember us meeting over Zoom (this was during the pandemic) and wrestling with exactly this question: Why is a meta-narrative of secularization different than the other stories, like the conflict model, that historians rightly reject? Maybe it is because the term secularism is flexible, as in our quibble above. Maybe it is because it is a pattern that allows for the terms under consideration (science and religion) to change in intelligible ways that transhistorical conflict or complexity do not. Maybe it is because it seems to explain how we got to the conflict and independence models in the first place. Maybe it is just that we are fans of the work of Berger, Taylor, and MacIntyre. We do not feel we have quite got to the bottom of this oddity yet, and we welcome the fact Evers intuited this.

We still think Harrison is right. Positing a given theory of secularization as an explanation could be a fruitful way to avoid extremes in the pendulum: extremes like "conflict!" or "complexity!" But hearing Evers's point made so lucidly drives home to us that our thinking about science-engaged theology is still in process. The closest we have got so far is as follows. Harrison (2019, 230) once wrote: "This is not the place to set out a complete account of what the history of science–religion relations would look like when viewed through [a given theory of secularization], but it does point to one strategy for moving the discussion on from reiterations of 'the complexity thesis'." Yes, exactly, and the first half of our book was an attempt to get that process underway. That is, given the visibility and influence of post-Barthian, post-liberal, Radical Orthodox theologies—all of which turn to some degree on theories of modernity—what would an account of the history of science–religion relations look like when view through those lenses?

Response to King-Ho Leung

We very much enjoyed King-Ho Leung's response as well, though for different reasons. He so thoroughly dove into two of our metaphors—policing and the EU—that it almost feels like *he* was the one who devised them, not us. He writes:

> if Perry and Leidenhag's theological vision may be said to be an act of "theological *de*-policing" both against any assertion of theological dominance over other disciplines and against any "secular policing" over theology, one may more generally question whether all acts of *de*-policing may always have a kind of policing as its flipside or complement. (Leung 2024, 889–90)

An excellent suggestion, one which we did not consider anywhere in the book! To put Leung's suggestion more pointedly, he is saying, *even though Perry and Leidenhag say they want to allow theology to stir up the reigning order, they have stacked the deck so that theology will be kept in its place*. The only theology that Perry and Leidenhag (2023, 24) will end up with is theology that has "learnt its lesson and now plays by science's rules." We hope that's not true, but we accept his challenge that our vision might lead to some things that we did not intend.

We have two reasons to believe otherwise. First, Leung (2024, 887) assumes that we object to "Milbank's rejection of 'pure nature'." If we believed nature was accessible to scientists via some "autonomous reason" (Leung 2024, 887), then Leung would be correct about us. Instead, we assume that knowledge of nature (or anything else) comes to us thanks to God's orderly creation of all things. So, we agree with Milbank's rejection of pure nature and autonomous reason, as far as we understand him. The quote from Milbank that we do challenge repeatedly, and which may have misled Leung to believe that we think that pure nature is accessible to human reason, is this notorious soundbite:

"unless other disciplines are (at least implicitly) ordered to theology . . . they are objectively and demonstrably null and void, altogether lacking in truth" (Milbank 2000, 45). *Implicitly* is doing an awful lot of work here! As we ask in the book about this very quote, what does "demonstrably null and void" imply about, say, a particular theory in biology or sociology? Further, what does it mean for another discipline to be ordered to theology? Is it working from or towards theological presuppositions? Does it make a difference whose theology is in view? If Leung or Milbank would answer these questions, perhaps giving us a demonstration (or example), we might be able to sort out our points of agreement and disagreement more fully.

Second, and relatedly, we speculate that part of Leung's worry originates in our willingness to accept the standard disciplinary boundaries of the university and—his implication seems to follow—controlling those that sneak across the disciplinary borders. We will get to this just below when we respond to Robyn Boeré, but we can say here that not all disciplinary divisions are consequences of modernity, that is, the policing strategy of Jefferson et al. We accept the division that originated in the ancient liberal arts not because of anything Jefferson said, but because of what the ancients and medievals believed. An academic discipline (*sciencia*) "treats only of one class of subjects" (Thomas Aquinas, *ST* 1:1.3, quoting Aristotle, *Posterior Analytics*, part 12). And why is that important? Not because the academic police might get you if you step out of line, but because what counts as evidence in, say, physics is not the same as what counts for evidence in biology, to say nothing about psychology or theology. Each has its own set of practices, traditions, virtues, and criteria for acceptance. This means that the boundaries, such as they are, are primarily internally drawn. Does that mean that our vision for science-engaged theology doesn't need a police department to enforce boundaries—which might alleviate Leung's worries? We think that our vision does not need a police force, but Leung's suggestions have us intrigued. It could be worth exploring in a later project.

Response to Robyn Boeré

Two passages stood out to us from Boeré's response. The first reads:

> [B]y the end of the book, it is still unclear what is meant by an "empirical reality." A less charitable way of characterizing my response is that I am here to quibble about the basic principle laid out on page 1 . . . "The basic principle of science-engaged theology is that whenever theologians make claims about created, empirical realities, they should incorporate the insights of empirical investigation into their analysis." (Boeré 2024, 894; Perry and Leidenhag 2023, 1)

Let us start with her puzzlement about what we mean by empirical realities. Our meaning was nothing spooky; indeed, it is quite mundane. Empirical realities are what can be known through our senses, together with the tools extending and

systematizing them, like microscopes and computers, and reasoning about them, like the process of cause and effect. And so, because new tools for extending our senses are always being developed, the scope of empirical realities may be subject to change. This is why empirical overlaps with, but is not identical to, material, created, or as discussed below, experiential. Empirical is a term defined by practice, not ontology.

Thomas calls this empirical knowledge (*scientiam experimentalem*). In fact, he even gives examples of this process, referring to how Jesus might have learned about ocean tides:

> Although all sensible things were not subjected to Christ's bodily senses, yet other sensible things were subjected to His senses; and from this He could come to know other things by the most excellent force of His reason, in the manner described in the previous reply; just as in seeing heavenly bodies He could comprehend their powers and the effects they have upon things here below, which were not subjected to His senses. (*ST* 3:12.1)

We can now answer Boeré's follow-up questions by some examples. "What is an empirical reality?" The moon and the ocean. "As opposed to a non-empirical reality?" Angels and the highest heaven. Why aren't the latter empirical? Because angels don't have bodies and, about heaven, "whatever we cognize about the heavens [= the sky and outer space], this is either via observation or via change; but the highest heaven [= where Jesus ascended to] is subject neither to change nor to observation" (Aquinas, *Scriptum super Sententiis*, lib. I, d. II, q. 1).

Perhaps an easier way to get at Boeré's puzzlement is via another passage from her reply: "Let us say, perhaps, that all created reality is empirical . . . for is there any part of reality that we cannot experience?" (Boeré 2024, 898). Not all experience is empirical. Following what we understand as Thomas' usage, we reserve *empirical* for what makes an impression on the *proper sensibles*, his term for the senses. Angels are not empirical realities, but humans can still experience them via a "prophetic vision—that is, according to imagination" (*ST* 1:51.2).[2] Nonetheless, the brains of humans who have experienced angels or God are empirical realities and could be studied, if Boeré or Thomas were interested enough, because they can be known through our senses, together with the tools extending and systematizing them.

Here is the other passage that stood out to us from Boeré (2024, 895):

> If one were to write a womanist account of theological anthropology, would one really want to turn to psychology, a field which not only has an acknowledged replicability crisis, but even more importantly, an acknowledged race and class issue? Similarly, there are good reasons to avoid scientific and medical data when writing on pregnancy.

Start with the first of these sentences. Why does she worry about scientific disciplines that have already acknowledged their problems? All that means is that there are disagreements among psychologists. Boeré's sentence works equally well the opposite way round:

> If one were to write a psychological account of the human person, would one really want to turn to womanist theology, a field which not only is rooted in texts that are acknowledged to be misogynist, but also has acknowledged race and class issues?

Of course, the point of much womanist theology is to critique these core texts, which, yes, proves that there is disagreement among theologians and biblical scholars, but so what? What we really should worry about is fields of inquiry that are not accountable to their own evidence and have no way of seeing and rooting out their own biases. The field of feminist science studies—which, we guess, is what Boeré is gesturing towards—depends profoundly on Thomas Kuhn and Helen Longino, who are two of the most important philosophers of science for our vision of science-engaged theology. Both take seriously and appreciate the socially situated nature of scientific knowledge.

We think that the point Boeré is aiming for may be better conveyed by the second sentence above. "Similarly, there are good reasons to avoid scientific and medical data when writing on pregnancy" (Boeré 2024, 895). She says this specifically regarding our passage in Chapter 5 that having clear success criteria allows theologians to prune wrong or idolatrous ideas from theological claims. In that passage, we (2023) wrote, "The tools of the empirical sciences allow theologians to be risky, and thus serve the church, in exactly this way." Boeré (2024, 895) responds: "much if not most scientific and medical writings on pregnancy begin from the assumption that pregnancy is gestation of a fetus in a container, not the intersubjective relationship between mother and child. What would empirical accountability then look like for that latter claim?" Excellent, precise question.

First, we affirm what Boeré is doing when she brings in perspectives beyond obstetrics in her understanding of pregnancy. We do not think, and never implied, that theology should only engage the empirical sciences. Multiple perspectives can be helpful in our quest for accountability. If Boeré is correct in her claim that pregnancy is an intersubjective relationship, then changing current obstetrics in line with this would not get rid of obstetrics, as she seems to hope; it would just lead to better, more accountable, medicine. Second, and even better, Boeré *is* making a risky claim! The claim that pregnancy is an intersubjective relationship is the sort of claim that can be tested, fitting better or worse with biological information about gestation as well as with the testimony of expectant mothers. That said, it seems to us that the two models

of pregnancy (uterus-as-container vs. intersubjective relationship) are the kind of hidden background assumptions or models that can both make sense of some of the information currently available concerning pregnancy. As such, it is not something Boeré needs to give up on, just because it is not the prevailing model in obstetric practice and discourse. Switching the model or metaphor in use will be a matter of showing obstetricians and others that the intersubjective relationship has better explanatory power and better health outcomes for patients than the current model.

Response to Mikael Leidenhag

Mikael Leidenhag's response first critiques the "metaphysical matching game" of other science-and-religion scholars for distorting the grammar of theology and consequently inviting scientism. We agree, and for exactly the reason Leidenhag identifies; namely, scientific pluralism undermines any notion that "science" can set the metaphysical or epistemological agenda. Second, Leidenhag outlines apophatic theology's (AT's) conviction that, because of divine transcendence, God, and the world do not share "logical space," such that science and theology will not use the same grammar or language. As Leidenhag (2024, 908) points out then, "there is significant agreement between [science-engaged theology] and AT in terms of how the science-engaged theology should *not* be conducted." But then he asks could there be a deeper partnership between science-engaged and apophatic theology? "Can there be a science-engaged apophatic theology?" (Leidenhag 2024, 906).

This brings us to our chosen quote to focus on from Leidenhag, which identifies a potential barrier to the idea of a science-engaged apophatic theology. He writes:

> [Perry and Leidenhag] stress the *entanglement* of theological and scientific concepts, whereby a concept "cannot be understood as either a scientific or theological in meaning and origin, but only as both," meaning that "the tools of more than one discipline" is needed to understand the phenomenon (2023, 13, 48). This may go contrary to the apophatic spirit of safeguarding the unique logical space of Christian theology, if the idea is that scientific and theological concepts ought to be used conjunctively to account for some phenomenon, thus, seemingly, implying that scientific and theological operate on the same level of reality or that they share the same logical space. (Leidenhag 2024, 909)

This is an excellent question, and the extended example from divine action debates drives the point home. What is the difference, he could have asked, between our notion of entanglement and Sarah Lane Ritchie's theistic naturalism, whereby "the physical is involved with God's active presence" (Ritchie 2019, 350; quoted by Leidenhag 2024, 904). Why couldn't Peacocke or Clayton simply

claim that emergence is an entangled concept? If we say, *there is no real difference, apart from our rule of thumb to be as specific and local as possible*, then we too might justly be accused of piecemeal scientism.

However, we do not think that entanglement is a kind of localized matching game. Unlike the metaphysical matching-game, our notion of entanglement is not about fitting a phenomenon from one metaphysical system into a different system. Entanglement is about using different disciplinary tools, or a plurality of perspectives, to account for a single phenomenon. What is the difference? In the case of entanglement, the existence of the phenomena is not in doubt. Entanglement is quite different, therefore, from placement problems or attempts to save the phenomena. This is because placement problems, matching games, and saving the phenomena are all attempts to hold onto things that seem incompatible with a metaphysic system (one *map*) that is taken as the universal criterion of rational enquiry, such as naturalism. But science-engaged theology does not accept that any one metaphysical map has all the rules that would then place certain phenomena (e.g., ethics, miracles, angels) under threat and in need of saving.

The difference may become clearer through some examples. Let us see how far we get if we try and transform Clayton's proposal into an acceptable form of science-engaged theology. Science-engaged theology does not ask, *where can I fit divine action into the metaphysical picture of causal closure provided by physics?* Instead, a science-engaged theologian could take Clayton's proposal and check if his proposal captures the scientific theory accurately (see Leidenhag 2021, 15–34). Alternatively, the science-engaged theologian might try and examine specific instances of emergence from theological and scientific perspectives. For example, Tom McLeish (2023, 149–66) explored emergence in the fractional quantum Hall state/effect and entangled polymer ring fluids and linked these to wisdom-based theology of science from Job 28. There is no reason to presuppose that it takes the theologian to the area of special divine action. McLeish (2023, 163) even states that his approach "can circumvent the tortured issue of divine action." This is a very unlikely place for the enquiry to arrive at since, at least within the orthodox Christian tradition, divine activity is not a product of evolution or material complexity.

Is it a problem that Christian models of divine action do not seem to be entangled with the theory of emergence? Not at all. Because entanglement is about maximizing accountability, and not the need for verifiability or rationality, there is no problem with saying that something, namely the divine nature, is not entangled with quantum mechanics or emergent phenomena. There is no obvious reason to suppose that quantum mechanics or emergence theory are the right tools to hold theologians accountable to on the question of how God acts in the world. It matters very much for moral reasons that we are accountable to the right people and rules.

Let us try a more promising example, taken from our book. A science-engaged theologian might say, *how does what I am saying about race (perhaps, as a category invented by Christian supersessionism) relate to what biologists say about race?* Very plausibly, both discourses have something important to contribute to how we understand racial categories. Of course, the same question could be asked with false theologies of race, such as those based on the descendants of Ham. Science-engaged theology is not a guarantor of truth, in part because the theology that one starts with or the scientific theory one is engaging might be complete bullshit. There is no way around this problem. Science-engaged theology is not valuable because it offers certainty—it does not—but because it tries to minimize error by making theology open to critique from a wide variety of perspectives. Modern genetics can help theologians see the theory about Ham's descendants as wrong. Theology can help explain how and why we started to organize people into racial groups, thus showing that scientists searching for signs of natural superiority are wrong. Together, we are not only slightly less prone to error, but we get a fuller picture of how the variation of human morphology relates to the racist organization of society. Of course, theology and the sciences are not the only two voices in this trading zone of ideas, and others are also needed.

Notes

[1] In contrast to our other respondents, Evers made six separate points in his piece, so us replying to only the one direct quote seemed a touch stingy. Let us briefly comment on Evers other five points. First, we do not know what to say about esotericism. Second, his examples of what happened in Germany in the nineteenth and early twentieth centuries fit very well with what we say about "the unifiers" (37–39), where the German context does make brief appearances—with Evers's help we clearly could have said more here. Third, Evers's argument that "good fences make for good neighbours," sounds correct to us, but makes us think Leung and Evers might have an interesting conversation. We agree with everything in the Hermeneutics as Foundation section (particularly the last sentence) and are thrilled Evers thinks that what we propose is "cool"!

[2] Thomas adds here that some angels sometimes assume "sensible" bodies, which would make them, by definition, empirical realities.

References

Berger, Peter L. 2008. "Secularization Falsified." *First Things First*, February 2008. https://www.firstthings.com/article/2008/02/secularization-falsified.

Boeré, Robyn. "Empirical Realities? A Question of Basic Principle." *Zygon: Journal of Religion and Science* 59 (3): 893–900.

Evers, Dirk. 2024. "Good Fences Make Engaging Neighbors." *Zygon: Journal of Religion and Science* 59 (3): 871–83.

Harrison, Peter. 2019. "Conflict, Complexity, and the Secularization in the History of Science and Religion." In *Rethinking History, Science, and Religion: An Exploration of Conflict and The Complexity Principle*, edited by Bernard Lightman, 221–34. Pittsburgh, PA: University of Pittsburgh Press.

Leidenhag, Joanna. 2021. *Minding Creation: Theological Panpsychism and the Doctrine of Creation*. London: Bloomsbury/T & T Clark.

Leidenhag, Mikael. 2024. "Considering the Place of Apophaticism within Science-Engaged Theology." *Zygon: Journal of Religion and Science* 59 (3): 901–11.

Leung, King-Ho. 2024. "Theology and Policing." *Zygon: Journal of Religion and Science* 59 (3): 884–92.

McLeish, Tom. 2023. "Complementary Causation and Emergence: A Substrate for Conjunctive Explanations." In *Conjunctive Explanations in Science and Religion*, edited by Diarmid A. Finnegan, David H. Glass, Mikael Leidenhag, and David N. Livingstone, 149–66. London: Routledge.

Milbank, John. 2000. "The Conflict of the Faculties: Theology and the Economy of the Sciences." In *Faithfulness and Fortitude: Conversations with the Theological Ethics of Stanley Hauerwas*, edited by Mark Theisson Nation and Samuel Wells, 39–58. Edinburgh: T & T Clark.

Numbers, Ronald L. 2019. "Revisiting the Battlefields of Science and Religion." In *Rethinking History, Science, and Religion: An Exploration of Conflict and The Complexity Principle*, edited by Bernard Lightman, 183–90. Pittsburgh, PA: University of Pittsburgh Press.

Perry, John, and Joanna Leidenhag. 2023. *Science-Engaged Theology*. Cambridge: Cambridge University Press.

Taylor, Charles. 2018. *A Secular Age*. Cambridge, MA: Belknap Press.

IN THIS ISSUE

ARTICLES

Mark Harris — Quantum Fundamentalism and Theological Liberty

Gordon L. Miller — The Soul of the Frog for a Postmodern Age: Updating a Classic Debate

Mitchell Mallary — Scientific Methods and Sacred Matters: On Methodological Naturalism, Divine Agency, and Christians in Science

RELIGIOUS ROBOTS: AI AS RELIGIOUS SUBJECT AND OBJECT

Pim Haselager — From Angels to Artificial Agents? AI as a Mirror for Human (Im)perfections

Fraser Watts
Yorick Alexander Wilks — Spiritual Conversation with a Companion Machine

Max Tretter — What Is the Afterlife Like for Robots? An Experimental Eschatological Sneak Peek

William F. Clocksin — Religion and the Android

Robert M. Geraci — Religion among Robots: An If/When of Future Machine Intelligence

Daekyung Jung — Are Religious Machines Possible? Embodied Cognition, AI, and Religious Behavior

Marius Dorobantu — Could Robots Become Religious? Theological, Evolutionary, and Cognitive Perspectives

SCHAEFER'S *WILD EXPERIMENT*

Mathew Arthur — Science and Technology Studies and Affect in the Wild

Anne Pollock — Cultivating Calm Passions

Josh A. Reeves — Kuhn, Conspiracy Theories, and Scientific Progress

Esha Shah — Love Affair with an Idea: Commentary on *Wild Experiment: Feeling Science and Secularism after Darwin* by Donovan Schaefer

Mari van Emmerik — Changing Minds: Cogency Theory and Secular Bodies

Donovan O. Schaefer — "They Made up Quiet Songs": Classification, Feeling, and Technique in Science and Religion

PERRY AND LEIDENHAG'S *SCIENCE-ENGAGED THEOLOGY*

Dirk Evers — Good Fences Make Engaging Neighbors

King Ho-Leung — Theology and Policing

Robyn Boeré — Empirical Realities? A Question of Basic Principle

Mikael Leidenhag — Considering the Place of Apophaticism within Science-Engaged Theology

John Perry
Joanna Leidenhag — Replies to Some of Our Friendlier Critics