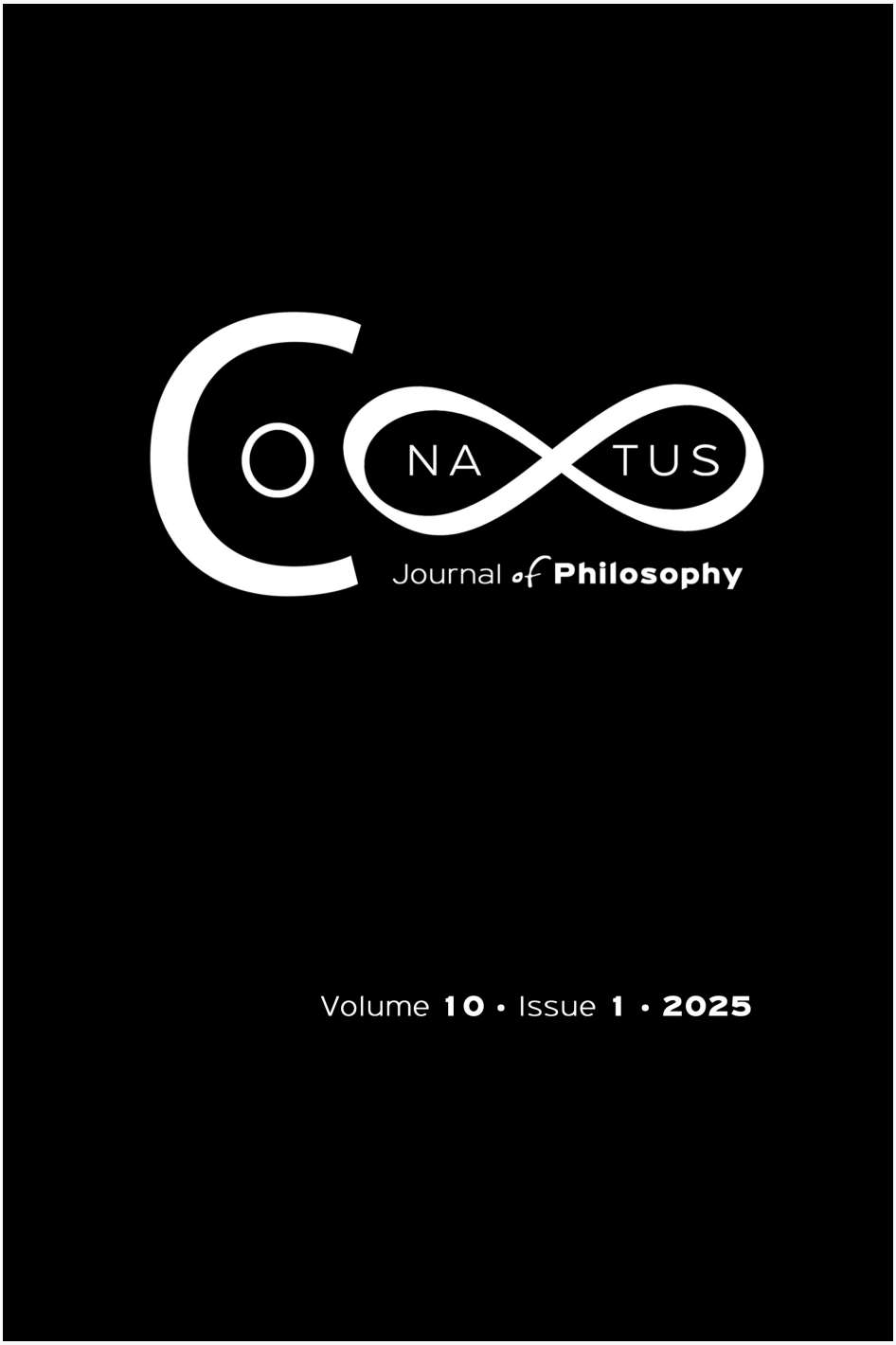
Conatus - Journal of Philosophy

Vol 10, No 1 (2025)

Conatus - Journal of Philosophy



Volume 10 • Issue 1 • 2025

Volume **10,** Issue **1 • 2025**

ρ-ISSN: 2653-9373
e-ISSN: 2459-3842

p-ISSN: 2653-9373
e-ISSN: 2459-3842

Contact information
School of Philosophy
7th floor, Office 746
University Campus, 15703 Zografos, Athens, Greece
e-mail: conatus@philosophy.uoa.gr
http://conatus.philosophy.uoa.gr
https://ejournals.epublishing.ekt.gr/index.php/Conatus

contents

articles

articles

K. I. Alkhouri · *Conatus* 10, no. 1 (2025): 9-56
doi: https://doi.org/10.12681/cjp.38734

Neuralink's Brain-Computer Interfaces and the Reshaping of Religious-Psychological Experience

Khader I. Alkhouri
National and Kapodistrian University of Athens, Greece
E-mail address: kalkhour@soctheol.uoa.gr
ORCID iD: https://orcid.org/0000-0002-9623-2468

Abstract

This research article examines the profound implications of Neuralink's brain-computer interface (BCI) technology on religious and psychological experiences. As BCIs advance toward direct neural interfacing, they present unprecedented opportunities and challenges for human spirituality, cognition, and self-understanding. Drawing on interdisciplinary research, we investigate the potential for technologically-mediated spiritual experiences and their impact on traditional religious practices and institutions. The study explores ethical considerations surrounding cognitive liberty, mental privacy, and the authenticity of BCI-induced experiences. Key findings indicate that BCIs could potentially induce or enhance altered states of consciousness associated with spiritual experiences, augment meditation practices, and redefine religious rituals. However, these capabilities raise significant ethical concerns, including issues of cognitive manipulation and equitable access. The research also highlights potential shifts in religious authority structures and the emergence of new techno-spiritual philosophies. By analyzing the societal and cultural impacts of widespread BCI adoption, this study provides a nuanced understanding of how Neuralink's technology may reshape the landscape of human consciousness and spirituality. The article contributes to the critical dialogue on the future of religious and psychological experiences in an era of advancing neurotechnology, balancing the transformative potential of BCIs with careful consideration of their ethical implications and philosophical ramifications.

Keywords: *Neuralink; brain-computer interfaces; religious experience; consciousness; neurotheology; cognitive enhancement; neuroethics; transhumanism*

I. Introduction

The rapid advancement of brain-computer interface (BCI) technology, spearheaded by companies like Neuralink, stands poised to revolutionize not only medical treatments and human-computer interaction but also the very nature of human psycho-

logical and spiritual experiences.[1] Founded by entrepreneur Elon Musk in 2016, Neuralink aims to develop high-bandwidth brain-machine interfaces that could potentially alter the fundamental ways in which humans perceive reality, process information, and engage with concepts of consciousness and spirituality.[2] As these technologies progress from treating neurological disorders to potentially enhancing human cognitive capabilities, they raise profound questions about the nature of religious experience,[3] the boundaries of human consciousness, and the ethical implications of technologically mediated spiritual states.[4]

a. The potential impact on religious and psychological experiences

The intersection of neurotechnology and religious experience represents a frontier that challenges long-held beliefs about the nature of consciousness, free will, and the human soul.[5] As Neuralink and similar companies push the boundaries of what is possible in brain-machine symbiosis, we are compelled to consider how these advancements might reshape our understanding of transcendent experiences, alter traditional religious practices, and potentially give rise to entirely new forms of techno-spirituality.[6] This article explores the potential impacts of Neuralink's BCI technology on religious and psychological experiences, examining both the promising possibilities and the ethical concerns that arise when the realm of the sacred intersects with cutting-edge neurotechnology.

b. Importance and relevance of the topic

The implications of this technology extend far beyond the medical realm, potentially transforming how individuals experience altered states of consciousness, engage in spiritual practices, and conceptualize their rela-

---

[1] Ajit Venniyoor, "Neuralink and Brain-Computer Interface – Exciting Times for Artificial Intelligence," *South Asian Journal of Cancer* 13, no. 1 (2024): 63-65.

[2] Elon Musk and Neuralink, "An Integrated Brain-Machine Interface Platform with Thousands of Channels," *Journal of Medical Internet Research* 21, no. 10 (2019): 1-14.

[3] See Richard Swinburne and Vasileios Meichanetsidis, "Proofs for the Existence of God: A Discussion with Richard Swinburne," *Conatus – Journal of Philosophy* 9, no. 2 (2024): 307ff.

[4] Khader I. Alkhouri, "The Role of Artificial Intelligence in the Study of the Psychology of Religion," *Religions* 15, no. 3 (2024): 1-27.

[5] Christoph Bublitz, "Neurotechnologies and Human Rights: Restating and Reaffirming the Multi-Layered Protection of the Person," *The International Journal of Human Rights* 28, no. 5 (2024): 782-807.

[6] A. Newberg and E. D'Aquili, "The Neuropsychology of Religious and Spiritual Experience," *Journal of Consciousness Studies* 7, no. 11-12 (2000): 251-266.

tionship with the divine or transcendent.[7] As we stand on the brink of this neurotechnological revolution, it becomes crucial to engage in thoughtful dialogue about the philosophical, ethical, and societal ramifications of BCIs in the context of religious and psychological experiences.[8]

II. Background on Neuralink's BCI technology

Neuralink's brain-computer interface technology represents a significant leap forward in the field of neurotechnology.[9] At its core, the Neuralink system consists of ultra-thin, flexible electrode "threads" that can be surgically implanted into the brain to record and stimulate neural activity. These threads, each thinner than a human hair, are connected to a small implantable device that processes and wirelessly transmits neural signals to external devices.[10]

a. Recent developments and human trials

Recent years have seen significant advancements in the field of Brain-Computer Interfaces (BCIs) and their integration with Artificial Intelligence (AI), bringing us closer to realizing the potential of this groundbreaking technology. Several key developments and human trials have marked important milestones in this rapidly evolving field.

Neuralink, the company founded by Elon Musk, has been at the forefront of BCI development. In 2022, Neuralink announced that it had submitted paperwork to the U.S. Food and Drug Administration (FDA) to begin human trials of its brain implant.[11] This move followed successful animal trials, including a demonstration of a monkey playing the video game Pong using only its mind.[12]

In parallel, other research groups have made significant strides. In 2021, researchers at Stanford University reported a breakthrough in

---

[7] Gabriel Fernandez Borsot, "Spirituality and Technology: A Threefold Philosophical Reflection," *Journal of Religion & Science* 58, no. 1 (2023): 6-22.

[8] Rafael Yuste et al., "Four Ethical Priorities for Neurotechnologies and AI," *Nature* 551 (2017): 159-163.

[9] Mujiba Shaima et al., "Elon Musk's Neuralink Brain Chip: A Review on 'Brain-Reading' Device," *Journal of Computer Science and Technology Studies* 6, no. 1 (2024): 200-203.

[10] Musk and Neuralink.

[11] Ashley Capoot, "Elon Musk Shows Off Updates to His Brain Chips and Says He's Going to Install One in Himself When They Are Ready," *CNBC*, December 1, 2022, https://www.cnbc.com/2022/12/01/elon-musks-neuralink-makes-big-claims-but-experts-are-skeptical-.html.

[12] Neuralink, "*Monkey MindPong*," YouTube, April 8, 2021, https://www.youtube.com/ watch?v=rsCul1sp4hQ.

BCI technology, demonstrating a system that allowed a paralyzed individual to write on a computer screen by imagining handwriting movements.[13] This study showcased the potential of AI in decoding complex neural signals and translating them into meaningful outputs.

Another notable advancement came from a team at the University of California, San Francisco. They developed a neuroprosthesis that successfully translated attempted speech into text in real-time for a paralyzed individual.[14] This study highlighted the potential of BCIs in restoring communication abilities to those with severe motor impairments.

In the realm of non-invasive BCIs, researchers at Carnegie Mellon University demonstrated a system that allowed users to mentally control a robotic arm to perform complex manipulation tasks.[15] This study underscored the potential of non-invasive BCIs in providing intuitive control of external devices.

As research continues to progress, the ethical and regulatory landscape is also evolving. In 2021, Chile became the first country to pass a "neurorights" law, aimed at protecting mental privacy and integrity in the face of advancing neurotechnology.[16] This legislative action underscores the growing recognition of the potential impacts of BCI and AI technologies on fundamental human rights and privacy.

While these advancements are promising, challenges remain in improving the longevity of implanted devices, enhancing signal resolution, and developing more sophisticated AI algorithms for interpreting neural signals. Nevertheless, the rapid pace of progress in this field suggests that the convergence of BCIs and AI will continue to push the boundaries of human-machine interaction in the coming years.

b. Current capabilities and medical applications

The current iteration of Neuralink's device, known as the Link, contains over 3,000 electrodes distributed across 96 threads, allowing for high-resolution recording of brain activity across large areas of the cor-

---

[13] Francis R Willett et al., "High-Performance Brain-to-Text Communication via Handwriting," *Nature* 593 (2021): 249-254.

[14] David A. Moses et al., "Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria," *The New England Journal of Medicine* 385, no. 3 (2021): 217-227.

[15] Andrew B. Schwartz, "Movement: How the Brain Communicates with the World," *Cell* 164, no. 6 (2016): 1122-1135.

[16] Alejandra Zúñiga Fajuri et al., "Chapter Seven - Neurorights in Chile: Between Neuroscience and Legal Science," in *Developments in Neuroethics and Bioethics*, vol. 4, ed. Martín Hevia, 165-179 (Academic Press, 2021).

tex. This represents a dramatic increase in the number of neural channels that can be simultaneously monitored compared to existing BCI systems. The implantation process is performed by a custom-built surgical robot, designed to insert the electrode threads with micron-level precision while minimizing tissue damage.[17]

While Neuralink's initial focus has been on medical applications, such as treating neurological disorders and restoring sensorimotor function in individuals with paralysis, the company's long-term vision is far more ambitious.[18] Elon Musk has articulated a future where BCIs could enhance human cognitive abilities, enable direct brain-to-brain communication, and even achieve a form of "symbiosis with artificial intelligence."[19] This vision aligns with broader trends in the field of neural interfaces, which have already demonstrated remarkable successes in medical contexts, such as treating Parkinson's disease and restoring hearing function through cochlear implants.[20]

The data processing capabilities of Neuralink's system are equally impressive. The implanted chip contains custom-designed application-specific integrated circuits (ASICs) that can process neural signals in real-time, using advanced signal processing and machine learning algorithms to decode complex patterns of brain activity.[21] This on-board processing capability is crucial for enabling high-bandwidth, low-latency communication between the brain and external devices.[22]

c. Recent developments and human trials

In early 2024, Neuralink announced a significant milestone with the successful implantation of its device in the first human subject. According to reports, this individual was able to control a computer cursor and play online chess using only their thoughts, demonstrating the potential for BCIs to restore communication and control capabilities

---

[17] Musk and Neuralink.

[18] Leigh R. Hochberg et al., "Neuronal Ensemble Control of Prosthetic Devices by a Human with Tetraplegia," *Nature* 442 (2006): 164-171.

[19] Musk and Neuralink.

[20] Matthew D. Johnson et al., "Neuromodulation for Brain Disorders: Challenges and Opportunities," *IEEE Transactions on Bio-Medical Engineering* 60, no. 3 (2013): 610-624.

[21] Musk and Neuralink.

[22] Bingzhao Zhu et al., "Closed-Loop Neural Prostheses with On-Chip Intelligence: A Review and a Low-Latency Machine Learning Model for Brain State Detection," *IEEE Transactions on Biomedical Circuits and Systems* 15, no. 5 (2021): 877-897.

in people with severe motor impairments.[23] While this achievement is promising, it is important to note that brain-controlled cursor movement is not unprecedented in the field of BCIs, and the full capabilities and long-term effects of Neuralink's technology remain to be seen.[24]

d. Potential future applications and Neuralink's long-term vision

As Neuralink continues to refine its technology, the potential applications extend far beyond medical treatments. The high-bandwidth, bidirectional communication between the brain and external devices opens up possibilities for enhancing memory, accelerating learning, and even expanding sensory perception beyond normal human ranges.[25] These capabilities, if realized, could have profound implications for how humans experience consciousness, process information, and engage with spiritual and religious concepts.[26]

The development of Neuralink's BCI technology raises important ethical and philosophical questions that intersect with religious and spiritual domains.[27] As we move towards a future where the boundaries between mind and machine become increasingly blurred, we must carefully consider the implications for human identity, free will, and the nature of religious and transcendent experiences.[28]

The capabilities of Neuralink's BCI system extend beyond simple recording and stimulation of neural activity. The company's ambitious goals include developing a "whole brain interface" capable of more closely connecting biological and artificial intelligence.[29] This high-bandwidth neural interface aims to achieve simultaneous recording from millions of neurons, potentially allowing for unprecedented

---

[23] Neuralink, "PRIME Study Progress Update," April 12, 2024, https://neuralink.com/blog/prime-study-progress-update/.

[24] Liam Drew, "Neuralink Brain Chip: Advance Sparks Safety and Secrecy Concerns," *Nature* 627, no. 8002 (2024): 19-20.

[25] Jobair Hossain Faruka et al., "An Investigation on Non-Invasive Brain-Computer Interfaces: Emotiv Epoc+ Neuroheadset and Its Effectiveness," ın *IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, 580-589 (Madrid, 2021).

[26] Dalia Fahmy, "Highly Religious Americans More Skeptical of Human Enhancements Such as Brain Implants, Gene Editing," May 4, 2022, https://pewrsr.ch/3kD3SGW. See also Julian Savulescu and Evangelos D. Protopapadakis, "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu," *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

[27] Allen Coin, "Ethical Aspects of BCI Technology: What is the State of the Art?" *Philosophies* 5, no. 4 (2020): 1-9.

[28] Soonkwan Hong, "Transcendence Up for Sale: Cracking the Onto-Existential Codes for Übermensch," *Consumption Markets & Culture* 27, no. 2 (2024): 152-177.

[29] Neuralink.

insights into brain function and the ability to modulate neural activity with remarkable precision.[30]

One of the key innovations in Neuralink's approach is the use of flexible polymer probes that can be inserted into the brain with minimal tissue damage.[31] These probes are designed to move with the brain, potentially reducing the risk of long-term inflammation and signal degradation that has plagued other invasive BCI systems.[32] The flexibility of these probes also allows for a greater number of electrodes to be implanted, increasing the spatial resolution and overall information bandwidth of the system.[33]

Neuralink's implantation procedure is performed by a custom-designed neurosurgical robot, which uses advanced imaging technology and precision control to insert the electrode threads into specific brain regions while avoiding blood vessels. This robotic approach aims to make the implantation process faster, safer, and more consistent than traditional neurosurgical techniques.[34] The company envisions that in the future, the implantation procedure could be performed on an outpatient basis, potentially making BCI technology more accessible to a wider population.[35]

While Neuralink's initial focus has been on medical applications, the potential uses of their BCI technology are far-reaching. In addition to restoring sensory and motor function in individuals with neurological disorders, future iterations of the technology could potentially enhance cognitive abilities in healthy individuals.[36] Elon Musk has speculated about the possibility of "consensual telepathy," where individuals could share thoughts and experiences directly through brain-to-brain interfaces.[37] Such capabilities, if realized, would have profound

---

[30] Gian Nicola Angotzi et al., "SiNAPS: An Implantable Active Pixel Sensor CMOS-Probe for Simultaneous Large-Scale Neural Recordings," *Biosensors and Bioelectronics* 126 (2019): 355-364.

[31] Jason E. Chung et al., "High-Density, Long-Lasting, and Multi-Region Electrophysiological Recordings Using Polymer Electrode Arrays," *Neuron* 101, no. 1 (2019): 21-31.

[32] Musk and Neuralink.

[33] Flavia Vitale et al., "Fluidic Microactuation of Flexible Electrodes for Neural Recording," *Nano Letters* 18, no. 1 (2017): 326-335.

[34] Musk and Neuralink.

[35] Annalisa Colucci et al., "Brain-Computer Interface-Controlled Exoskeletons in Clinical Neurorehabilitation: Ready or Not?" *Neurorehabilitation and Neural Repair* 36, no. 12 (2022): 747-756.

[36] Brian Fiani et al., "An Examination of Prospective Uses and Future Directions of Neuralink: The Brain-Machine Interface," *Cureus* 13, no. 3 (2021): 1-4.

[37] Edd Gent, "Brain-Computer Interfaces Are Coming: 'Consensual Telepathy', Anyone?" *Washington Post*, June 11, 2017, https://www.washingtonpost.com/national/health-science/brain-computer-interfaces-are-coming-consensual-telepathy-anyone/2017/06/09/9345c682-46ef-11e7-98cd-af64b4fe2dfc_story.html.

implications for human communication, learning, and even the nature of consciousness itself.[38]

However, it is important to note that many of these proposed applications remain speculative and face significant technical and biological challenges. The human brain is an incredibly complex system, and our understanding of how it encodes information and generates consciousness is still limited.[39] Critics have pointed out that some of Neuralink's more ambitious goals, such as "downloading" memories or achieving true mind-reading capabilities, may be overly optimistic given our current scientific understanding.[40]

Despite these challenges, the progress made by Neuralink and other companies in the field of BCIs is undeniable. The successful implantation of Neuralink's device in a human subject in early 2024 marked a significant milestone, demonstrating the potential for high-bandwidth neural interfaces to restore function in individuals with severe motor impairments.[41] As the technology continues to advance, it is likely to have far-reaching implications not only for medicine and human-computer interaction[42] but also for our understanding of consciousness, cognition, and even spirituality.[43]

The development of Neuralink's BCI technology raises important ethical considerations, particularly in the context of cognitive enhancement and potential non-medical applications.[44] As these devices become more sophisticated and potentially more widespread, questions of mental privacy, cognitive liberty, and the nature of human identity

---

[38] Tim Urban, "Neuralink and the Brain's Magical Future," *Wait But Why*, April 20, 2017, https://waitbutwhy.com/2017/04/neuralink.html.

[39] Birgitta Langley, "Consciousness as the Final Beacon of Humanity," *Recent Research Advances in Arts and Social Studies* 8 (2024): 118-154.

[40] William Armstrong and Katina Michael, "The Implications of Neuralink and Brain Machine Interface Technologies," in *2020 IEEE International Symposium on Technology and Society (ISTAS)*, 201-203 (Tempe, AZ, USA, 2020). Also, Evangelos D. Protopapadakis, "Messing with Autobiographical Memory: Identity, and Moral Status," *International Dialogue East-West* 4 (2022): 175-181, and Panagiotis Kormas et al., "Implications of Neuroplasticity to the Philosophical Debate of Free Will and Determinism," in *Handbook of Computational Neurodegeneration*, eds. P. Vlamos, I. S. Kotsireas, I. Tarnanas, 453-471 (Spriner, 2023).

[41] Neuralink.

[42] Ujwal Chaudhary et al., "Brain-Computer Interfaces for Communication and Rehabilitation," *Nature Reviews Neurology* 12 (2016): 513-525.

[43] Francis R. Willett et al., "High-Performance Brain-to-Text Communication via Handwriting," *Nature* 593 (2021): 249-254.

[44] Ethan Waisberg et al., "Correction: Ethical Considerations of Neuralink and Brain-Computer Interfaces," *Annals of Biomedical Engineering* 52 (2024): 1937-1939.

will become increasingly pressing.[45] These ethical concerns are particularly relevant when considering the potential impact of BCIs on religious and psychological experiences, as we will explore in subsequent sections of this article.[46]

Neuralink's brain-computer interface (BCI) technology represents a potential paradigm shift in how humans interact with their own minds and the digital world, with far-reaching implications for religious and psychological experiences. The case of Nolan Arbaugh, the first human recipient of a Neuralink implant, provides a glimpse into this transformative potential. Nolan's ability to control a computer cursor with his thoughts not only restored a degree of independence but also reinforced his faith, illustrating how BCIs could enhance feelings of gratitude, purpose, and divine connection. His experience highlights the complex interplay between cutting-edge neurotechnology and deeply held spiritual beliefs, suggesting that as BCIs advance, they may become a new frontier for exploring consciousness and transcendent experiences.[47]

As BCI technology progresses, it could fundamentally reshape our understanding of spirituality and consciousness. The ability to directly interface with the brain opens up possibilities for novel spiritual practices, such as digitally-mediated prayer or technologically-induced mystical states. This convergence of neuroscience and spirituality may lead to new fields of study, like neurotheology, which explores the neural correlates of religious experiences. Moreover, the capacity to stimulate specific brain regions or alter neural patterns could potentially allow for the modulation of emotional states, personality traits, or even core beliefs, raising profound questions about the nature of free will, personal identity, and the authenticity of technologically-mediated spiritual experiences.[48]

However, the integration of BCIs into spiritual and psychological realms also presents significant ethical challenges and societal implications. The power to directly influence brain function could be misused to manipulate beliefs or experiences, potentially infringing on cogni-

---

[45] Marcello Ienca and Roberto Andorno, "Towards New Human Rights in the Age of Neuroscience and Neurotechnology," *Life Sciences, Society and Policy* 13, no. 1 (2017): 1-27.

[46] Alexander N. Pisarchik, "From Novel Technology to Novel Applications: Comment on 'An Integrated Brain-Machine Interface Platform with Thousands of Channels' by Elon Musk and Neuralink," *Journal of Medical Internet Research* 21, no. 10 (2019): 1-7.

[47] Lex Fridman and Elon Musk. "Neuralink Human Trial: Nolan Arbaugh & Elon Musk | Lex Fridman Podcast," YouTube, August 2, 2024, https://youtu.be/Kbk9BiPhm7o?feature=shared.

[48] Ibid.

tive liberty and religious freedom. As BCIs become more prevalent, society will need to grapple with questions of data privacy, mental autonomy, and the potential for creating new forms of inequality based on neural enhancement.[49] Additionally, the blurring of lines between human cognition and artificial intelligence may necessitate a reevaluation of traditional concepts of personhood and consciousness. As we stand on the brink of this neurotechnological revolution, it is crucial to foster interdisciplinary dialogue between neuroscientists, ethicists, theologians, and philosophers to navigate the complex landscape of BCI-mediated religious and psychological experiences, ensuring that these technologies are developed and implemented in ways that respect human dignity and diversity of belief.[50]

III. Potential impacts on religious experience

The integration of Neuralink's brain-computer interface technology into human cognition has the potential to profoundly reshape religious and spiritual experiences.[51] By enabling direct modulation of neural activity and potentially expanding human perceptual and cognitive capabilities, BCIs may alter how individuals engage with transcendent states, religious practices, and spiritual concepts.[52]

a. Altered states of consciousness and mystical experiences

One of the most significant potential impacts of BCI technology on religious experience is the ability to induce or enhance altered states of consciousness often associated with spiritual and mystical experiences.[53] Research has shown that religious and mystical experiences involve distinct patterns of brain activity, particularly in regions like the prefrontal cortex, temporal lobes, and limbic system.[54]

Neuralink's high-resolution neural interfaces could potentially allow for precise stimulation or inhibition of these brain regions, potentially inducing states of consciousness traditionally associated with

---

[49] John R. Hamilton et al., "Adding External Artificial Intelligence (AI) into Internal Firm-Wide Smart Dynamic Warehousing Solutions," *Sustainability* 16, no. 10 (2024): 1-23.

[50] Fridman and Musk.

[51] Alkhouri.

[52] Jordan Grafman et al., "The Neural Basis of Religious Cognition," *Current Directions in Psychological Science* 29, no. 2 (2020): 126-133.

[53] Ibid.

[54] A. Newberg and E. D'Aquili, "The Neuropsychology of Religious and Spiritual Experience," *Journal of Consciousness Studies* 7, no. 11-12 (2000): 251-266.

deep meditation, prayer, or even mystical encounters. As Newberg and Waldman (2009) suggest, "If we can stimulate these regions, is it possible to artificially induce spiritual experiences? And if so, are these experiences as authentic as those that occur naturally?"[55]

This capability raises intriguing possibilities for both religious practice and scientific study of spiritual phenomena. On one hand, it could democratize access to profound spiritual states that typically require years of dedicated practice to achieve reliably.[56] On the other hand, it challenges traditional notions of the authenticity and value of spiritual experiences, potentially blurring the lines between naturally occurring and technologically mediated transcendent states.[57]

b. Enhanced meditation and contemplative practices

BCIs could potentially enhance meditation and other contemplative practices by providing real-time neurofeedback and even direct neural entrainment. Studies have shown that experienced meditators exhibit distinct patterns of brain activity, such as increased gamma wave synchronization, associated with states of heightened awareness and blissful consciousness.[58]

Neuralink's technology could potentially allow practitioners to more easily achieve and maintain these desired brain states, accelerating the development of meditative skills and potentially allowing for deeper or more sustained contemplative experiences.[59] This could be seen as a powerful new tool for spiritual development, analogous to how psychedelic substances have been used in some spiritual traditions to catalyze mystical experiences and insights.

However, the use of BCIs in this context also raises questions about the role of effort and discipline in spiritual practice.[60] Many religious

---

[55] Andrew Newberg and Mark Robert Waldman, *How God Changes Your Brain: Breakthrough Findings from a Leading Neuroscientist* (Random House Publishing Group, 2009), 164-165.

[56] Michael Inzlicht et al., "Neural Markers of Religious Conviction," *Psychological Science* 20, no. 3 (2009): 385-392.

[57] Gabriel Fernandez Borsot, "Spirituality and Technology: A Threefold Philosophical Reflection," *Journal of Religion & Science* 58, no. 1 (2023): 6-22.

[58] Antoine Lutz et al., "Long-Term Meditators Self-Induce High-Amplitude Gamma Synchrony During Mental Practice," *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 46 (2004): 16369-16373.

[59] Xiao-yu Sun and Bin Ye, "The Functional Differentiation of Brain-Computer Interfaces (BCIs) and Its Ethical Implications," *Humanities and Social Sciences Communications* 10 (2023): 1-9.

[60] Zhi-Ping Zhao et al., "Modulating Brain Activity with Invasive Brain-Computer Interface: A Narrative Review," *Brain Sciences* 13, no. 1 (2023): 1-14.

traditions emphasize the importance of dedicated practice and gradual spiritual development. The ability to rapidly induce meditative states through technological means may be viewed by some as a shortcut that bypasses important aspects of the spiritual journey.[61] To illustrate how BCI technology might be integrated into religious practices, consider the following hypothetical scenario:

Imagine a future where a Buddhist temple offers BCI-enhanced meditation sessions. Practitioners wear non-invasive BCI headsets that provide real-time feedback on their brain activity, guiding them towards states associated with deep meditation. The temple's meditation instructor uses a dashboard to monitor the collective brain states of the group, adjusting the guided meditation accordingly. This technologically-assisted approach could potentially accelerate the development of meditation skills, especially for beginners, while also offering experienced practitioners new insights into their practice.

This scenario highlights both the potential benefits and challenges of integrating BCI technology into traditional spiritual practices. While it could make advanced meditative states more accessible, it also raises questions about the authenticity of the experience and the role of personal effort in spiritual growth. Religious institutions adopting such technologies would need to carefully consider how to balance technological assistance with the traditional values and methods of their spiritual traditions.

c. Technologically-mediated divine communication

Some individuals may interpret the enhanced cognitive and perceptual capabilities enabled by BCIs as a form of divine or spiritual communication. The ability to access vast stores of information instantly or to perceive aspects of reality beyond normal human sensory ranges could be seen as a technologically-mediated form of revelation or insight.[62]

This possibility aligns with what some scholars have termed "techno-spirituality" or "cybernetic spirituality," where advanced technologies are integrated into religious and spiritual practices and beliefs.[63] As Max Hodak, co-founder of Neuralink, speculated, there may be an "oppor-

[61] Fazale Rana, "A Christian Perspective on Living Electrodes," January 13, 2021, https://reasons.org/explore/blogs/the-cells-design/a-christian-perspective-on-living-electrodes.

[62] Ranganatha Sitaram et al., "Brain-Computer Interfaces and Neurofeedback for Enhancing Human Performance," in *Human Performance Optimization: The Science and Ethics of Enhancing Human Capabilities*, eds. Michael D. Matthews and David M. Schnyer, 125-141 (Oxford Academic, 2019).

[63] Robert M. Geraci, "Spiritual Robots: Religion and Our Scientific View of the Natural World," *Theology and Science* 4, no. 3 (2006): 229-246.

tunity for a new religion" that embraces scientific understanding while facilitating profound altered states through technological means.[64]

d. Shared religious experiences and collective consciousness

Neuralink's long-term vision includes the possibility of direct brain-to-brain communication, which could enable unprecedented forms of shared religious experiences.[65] Imagine, for instance, the ability to directly share the subjective experience of a profound spiritual moment with others, or to engage in a form of technologically-mediated collective prayer or meditation.[66]

This capability could transform how religious communities form and practice, potentially intensifying feelings of unity and shared spiritual consciousness. However, it also raises questions about the boundaries of individual spiritual experiences and the potential for technological mediation to homogenize or standardize what have traditionally been deeply personal and subjective encounters with the divine.[67]

e. Redefinition of religious rituals and practices

As BCI technology becomes more advanced and widespread, it may lead to the evolution of new forms of religious rituals and practices that incorporate direct neural interfaces.[68] Traditional practices like prayer, meditation, or participation in religious ceremonies could be augmented or even replaced by technologically-mediated experiences.

For example, instead of reading sacred texts, future religious practitioners might directly download or experience the emotional and cognitive states associated with religious narratives. Or, religious communities might engage in synchronized neural entrainment as a form of collective worship.[69]

---

[64] Victor Tangermann, "Neuralink Co-Founder Has an Idea for a New Religion," March 26, 2021, https://futurism.com/neuralink-co-founder-new-religion-drugs-experience-god.

[65] Neuralink, "Neuralink Progress Update," YouTube, August 28, 2020, https://www.youtube.com/live/DVvmgjBL74w?feature=shared.

[66] Michael Muller et al., "Spiritual Life and Information Technology," *Communications of the ACM* 44, no. 3 (2001): 82-83.

[67] Jim Denison, "Elon Musk's Neuralink Implants Brain Chip in Human: Four Biblical Responses," *DenisonForum*, February 1, 2024, https://www.denisonforum.org/daily-article/elon-musks-neuralink-implants-brain-chip-in-human/.

[68] Abdullah Ayub Khan et al., "A Blockchain Security Module for Brain-Computer Interface (BCI) with Multimedia Life Cycle Framework (MLCF)," *Neuroscience Informatics* 2, no. 1 (2022): 1-14.

[69] Gemma Perry et al., "How Chanting Relates to Cognitive Function, Altered States and Qual-

While these possibilities may seem far-fetched, they highlight the potential for BCI technology to fundamentally alter how humans engage with religious concepts and practices.[70] As Neuralink and similar technologies continue to advance, religious institutions and individuals will need to grapple with how to integrate or respond to these new capabilities in ways that remain true to their spiritual values and traditions.[71]

IV. Philosophical implications

Neuralink's brain-computer interface (BCI) technology, spearheaded by Elon Musk, represents a paradigm shift in human-machine interaction, raising profound philosophical questions about consciousness, identity, and the nature of humanity itself.[72] At its core, this technology challenges our understanding of the mind-body problem, a centuries-old philosophical debate about the relationship between mental phenomena and physical processes.[73] By potentially enabling direct communication between the brain and external devices, Neuralink's BCI blurs the line between biological cognition and artificial systems, echoing Andy Clark's concept of humans as "natural-born cyborgs."[74] This integration prompts us to reconsider the boundaries of personal identity and the self, particularly as our thoughts and memories become increasingly intertwined with technology. Furthermore, the prospect of cognitive enhancement through BCIs raises ethical concerns about fairness and equality, reminiscent of debates surrounding human enhancement technologies.[75] Questions of free will and autonomy also

---

ity of Life," *Brain Sciences* 12, no. 11 (2022): 1-22.

[70] Evelyn Karikari and Konstantin A. Koshechkin, "Review on Brain-Computer Interface Technologies in Healthcare," *Biophysical Reviews* 15, no. 5 (2023): 1351-1358.

[71] Ian M. Giatti, "Healing the Lame, Bringing Sight to the Blind? Elon Musk's Ambitions for Neuralink Raise 'Deep, Serious' Questions (Part 1)," *The Christian Post*, December 26, 2022, https://www.christianpost.com/news/elon-musk-ambitions-for-neuralink-raise-deep-serious-questions.html.

[72] Musk. On the fragility of identity see also Gerard Elfstrom, "The Theft: An Analysis of Moral Agency," *Conatus – Journal of Philosophy* 5, no. 1 (2020): 27-53, as well as David Menč ik, "Identity Theft: A Thought Experiment on the Fragility of Identity," *Conatus – Journal of Philosophy* 5, no. 1 (2020): 71-83.

[73] David J. Chalmers, "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2, no. 3 (1995): 200-219.

[74] Andy Clark, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford University Press, 2003).

[75] Michael J. Sandel, *The Case against Perfection: Ethics in the Age of Genetic Engineering* (Harvard University Press, 2007).

come to the forefront, as the ability to influence or generate thoughts through external systems challenges traditional notions of agency and responsibility.[76]

The implications of Neuralink's technology extend beyond individual identity to broader societal and existential questions. The intimate nature of BCIs introduces unprecedented challenges to mental privacy and cognitive liberty, echoing concerns raised by philosophers and ethicists about the right to mental sovereignty in the digital age.[77] This technology also represents a significant step towards transhumanism, a philosophical movement that advocates for the use of technology to enhance human physical and cognitive capacities.[78] As we approach the possibility of a "post-human" future, we must grapple with fundamental questions about what it means to be human and how our relationship with technology might evolve. The potential for BCIs to augment or even replicate aspects of human consciousness also raises profound questions about the nature of subjective experience and the possibility of artificial consciousness.[79] Moreover, the development of BCIs intersects with ongoing debates in the philosophy of mind about the computational theory of consciousness and whether the mind can be fully explained in terms of information processing.[80] As Neuralink and similar technologies progress, they not only push the boundaries of neuroscience and engineering but also challenge us to reevaluate our philosophical understanding of mind, consciousness, and the human condition.

V. Psychological implications

The potential impact of Neuralink's brain-computer interface technology extends far beyond religious experiences, touching on fundamental aspects of human psychology and cognition. As these devices become more sophisticated in their ability to read, write, and modulate neural activity, they may profoundly alter how we perceive, think, and experience consciousness itself.[81]

---

[76] Robert A. Kane, *Contemporary Introduction to Free Will* (Oxford University Press, 2005).

[77] Ienca and Andorno, 5.

[78] Nick Bostrom, "A History of Transhumanist Thought," *Journal of Evolution and Technology* 14, no. 1 (2005): 1-25.

[79] Giulio Tononi and Christof Koch, "Consciousness: Here, There and Everywhere?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, no. 1668 (2015): 20140167.

[80] Gualtiero Piccinini and Sonya Bahar, "Neural Computation and the Computational Theory of Cognition," *Cognitive Science* 37, no. 3 (2013): 453-488.

[81] Ray Kurzweil, *How to Create a Mind: The Secret of Human Thought Revealed* (Penguin, 2012).

a. Changes in perception and cognition

Neuralink's BCI technology has the potential to dramatically enhance and alter human perceptual and cognitive capabilities. By interfacing directly with sensory processing regions of the brain, BCIs could potentially expand the range of human perception beyond our biological limitations. This might include the ability to perceive electromagnetic fields, infrared light, or even abstract data streams.[82]

Moreover, the potential for BCIs to enhance cognitive functions like memory, attention, and information processing could lead to significant changes in how individuals engage with complex ideas, including religious and philosophical concepts. The ability to rapidly access and process vast amounts of information could transform religious scholarship and philosophical inquiry.[83]

b. Impact on sense of self and personal identity

One of the most profound psychological implications of BCI technology is its potential to alter our sense of self and personal identity.[84] As BCIs enable more direct connections between our brains and external devices or even other brains, the boundaries of what we consider to be "self" may become increasingly blurred.[85]

This aligns with ongoing debates in cognitive science and philosophy of mind about the nature of consciousness and the self. Some researchers, like philosopher Andy Clark, argue for an "extended mind" thesis, where our cognitive processes already extend beyond the boundaries of our skulls to include external tools and technologies.[86] Brain-Computer Interfaces (BCIs) exemplify this by creating direct neural-technological connections, offering new perspectives on classic philosophical problems. While BCIs demonstrate physical-mental bridges relevant to the mind-body problem, they don't fully resolve the "hard problem" of consciousness.[87] BCI technology could take this

---

[82] Musk.

[83] Kirk A. Bingaman, "Religion in the Digital Age: An Irreversible Process," *Religions* 14, no. 1 (2023): 1-14.

[84] F. Gilbert et al., "Embodiment and Estrangement: Results from a First-in-Human 'Intelligent BCI' Trial," *Science and Engineering Ethics* 25, no. 1 (2019): 83-96.

[85] Barış Serim et al., "Revisiting Embodiment for Brain-Computer Interfaces," *Human-Computer Interaction* 39, nos. 5-6 (2023): 1-27.

[86] Andy Clark and David Chalmers, "The Extended Mind," *Analysis* 58, no. 1 (1998): 7-19.

[87] David J. Chalmers, "The Puzzle of Conscious Experience," *Scientific American* 273, no. 6

extension to a new level, potentially leading to a more fluid and expansive sense of self.[88]

The implications for religious and spiritual conceptions of the soul or essential self are significant.[89] Many religious traditions have specific beliefs about the nature of the soul and its relationship to the body and brain.[90] As BCI technology blurs the lines between mind and machine, these beliefs may need to be reevaluated or reinterpreted.[91]

c. Emotional regulation and mood enhancement

Neuralink's technology could potentially offer unprecedented control over emotional states and mood. By modulating activity in brain regions associated with emotion, BCIs might allow individuals to regulate their affective states more directly than ever before.[92]

This capability could have profound implications for mental health treatment, potentially offering new approaches to managing conditions like depression, anxiety, and PTSD.[93] The potential of BCIs for treating mental health conditions intersects with philosophical perspectives on personal growth through adversity. Nietzsche's concept of "what does not kill me makes me stronger" emphasizes psychological resilience and self-overcoming.[94] Similarly, Existentialist philosophers like Viktor Frankl argue that finding meaning through struggle is essential for mental health, as detailed in his work on logotherapy.[95] The Stoic tradition, particularly through Epictetus and Marcus Aurelius, emphasizes personal development through confronting difficulties, viewing obstacles as opportunities for

(1995): 80-86.

[88] Baraka Maiseli et al., "Brain-Computer Interface: Trend, Challenges, and Threats," *Brain Informatics* 10, no. 1 (2023): 1-16.

[89] Sam Harris et al., "The Neural Correlates of Religious and Nonreligious Belief," *PLoS One* 5, no. 1 (2010): 1-10.

[90] Mario Beauregard and Vincent Paquette, "Neural Correlates of a Mystical Experience in Carmelite Nuns," *Neuroscience Letters* 405, no. 3 (2006): 186-190.

[91] Ienca and Andorno.

[92] Sara Goering and Eran Klein, "Neurotechnologies and Justice by, with, and for Disabled People," in *The Oxford Handbook of Philosophy and Disability*, eds. D. T. Wasserman and A. Cureton, 616-632 (Oxford University Press, 2019).

[93] David Bergeron, "Use of Invasive Brain-Computer Interfaces in Pediatric Neurosurgery: Technical and Ethical Considerations," *Journal of Child Neurology* 38, nos. 3-4 (2023): 223-238.

[94] Friedrich Nietzsche, *The Twilight of the Idols and the Anti-Christ: or How to Philosophize with a Hammer* (National Geographic Books, 1990).

[95] Victor E. Frankl, *Man's Search for Meaning* (Beacon Press, 2006), 96-97, 133.

growth.[96] However, this raises ethical questions about whether technological interventions that potentially bypass struggle might impact personal development opportunities.[97] However, it also raises questions about the nature of authentic emotional experiences and the role of emotional struggle in personal growth and spiritual development.[98]

Various philosophical and religious traditions have long emphasized the value of struggle and adversity in personal and spiritual growth, which could be challenged by Neuralink's emotion-modulating capabilities. Western philosophers like Friedrich Nietzsche promoted the concept of "amor fati" (love of fate), encouraging the embrace of life's challenges as opportunities for growth.[99] Existentialist thinkers such as Jean-Paul Sartre and Albert Camus viewed struggle as essential to finding meaning in life.[100] Eastern philosophies, particularly Buddhism, teach that understanding and overcoming suffering is crucial for spiritual development.[101] Christian theology, as articulated by thinkers like C.S. Lewis, often emphasizes the redemptive power of suffering.[102] In psychology, the concept of post-traumatic growth suggests that struggling with adversity can lead to positive psychological changes.[103] These diverse perspectives highlight the potential tension between traditional views on the value of emotional and psychological struggles and the unprecedented ability of brain-computer interfaces to alleviate such experiences.

Many religious and spiritual traditions emphasize the importance of grappling with difficult emotions as part of the path to enlightenment or spiritual maturity.[104] The ability to technologically regulate emotions could be seen as short-circuiting this process, potentially diminishing opportunities for spiritual growth through emotional challenges.[105]

---

[96] Pierre Hadot, *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault* (Blackwell, 1995), 83-85, 207-208.

[97] Francis Fukuyama, *Our Posthuman Future: Consequences of the Biotechnology Revolution* (Farrar, Straus and Giroux, 2002), 172-173.

[98] Goering.

[99] Friedrich Nietzsche, *The Gay Science*, trans. Walter Kaufmann (Vintage Books, 1974).

[100] Albert Camus, *The Myth of Sisyphus and Other Essays*, trans. Justin O'Brien (Vintage Books, 1991).

[101] Walpola Rahula, *What the Buddha Taught* (Grove Press, 1974).

[102] C. S. Lewis, *The Problem of Pain* (HarperOne, 2001).

[103] Richard G. Tedeschi and Lawrence G. Calhoun, "Posttraumatic Growth: Conceptual Foundations and Empirical Evidence," *Psychological Inquiry* 15, no. 1 (2004): 1-18.

[104] Carol D. Ryff, "Spirituality and Well-Being: Theory, Science, and the Nature Connection," *Religions* 12, no. 11 (2021): 1-19.

[105] Alexandra H. Bettis et al., "Digital Technologies for Emotion-Regulation Assessment and Intervention: A Conceptual Review," *Clinical Psychological Science: A Journal of the Associa-*

d. Memory enhancement and its implications for religious knowledge

BCIs could potentially enhance both the formation and recall of memories, with significant implications for religious education and the transmission of spiritual knowledge.[106] Imagine being able to perfectly recall religious texts, historical details, or the nuances of complex theological arguments.[107]

While this enhanced recall could greatly facilitate religious scholarship and practice, it also raises questions about the value of effortful study and contemplation in spiritual development.[108] Many religious traditions emphasize the importance of grappling with sacred texts and gradually internalizing spiritual teachings. If this knowledge can be rapidly "downloaded" or accessed through a BCI, how might it change the nature of religious learning and wisdom?

e. Potential for treating mental health conditions

Brain-computer interfaces (BCIs) like those being developed by Neuralink hold significant promise for treating mental health conditions. This potential application extends beyond the initial focus on motor function restoration and could revolutionize our approach to psychological disorders.[109]

BCIs could potentially offer new approaches to treating conditions such as depression, anxiety, addiction, and even existential distress. By allowing for more precise and personalized neuromodulation therapies, these technologies might provide relief for symptoms of mental illness that can interfere with overall well-being, including spiritual well-being.[110]

The ability to directly modulate neural activity in specific brain regions associated with mood and emotional regulation could offer new

---

tion for Psychological Science 10, no. 1 (2022): 3-26.

[106] John F. Burke et al., "Brain-Computer Interface to Enhance Episodic Memory in Human Participants," *Frontiers in Human Neuroscience* 8 (2015): 1-10.

[107] Alkhouri.

[108] Kirk A. Bingaman, "Religion in the Digital Age: An Irreversible Process," *Religions* 14, no. 1 (2023): 1-14.

[109] Imane El. Atillah, "Neuralink's Brain Chip: How Brain-Computer Interfaces May Revolutionise Treatment for Disabilities," *Euronews*, June 8, 2024, https://www.euronews.com/health/2024/06/06/neuralinks-brain-chip-how-brain-computer-interfaces-may-revolutionise-treatment-for-disabi.

[110] Elon Musk, "Neuralink Update," X, July 10, 2024, https://x.com/neuralink/status/1811095113281720722.

therapeutic options for individuals who have not responded well to traditional treatments. This aligns with the broader goals of Neuralink's technology, which aims to create high-bandwidth connections between the brain and external devices.[111]

However, the use of BCIs in mental health treatment also raises ethical questions about the nature of the self and the role of struggle in personal growth. As these technologies develop, it will be crucial to carefully consider their implications for human identity, autonomy, and the authentic experience of emotions.

Brain-computer interfaces (BCIs) in mental health treatment show promise but remain largely theoretical, requiring careful consideration through the lens of biomedical ethics principles: autonomy, beneficence, non-maleficence, and justice.[112] While companies like Neuralink advance BCI development, comprehensive clinical trials are essential to validate their safety and efficacy, particularly given the significant knowledge gaps regarding long-term neural effects.[113] Critical concerns include safety protocols and equitable access across socioeconomic groups, highlighting the need for careful implementation and ethical oversight.[114]

To address these concerns, it's valuable to consider the Four Principles of Biomedical Ethics in the context of BCI applications for mental health.[115] These principles – autonomy, beneficence, non-maleficence, and justice – provide a framework for evaluating the ethical implications of new medical technologies such as BCIs.

Autonomy in this context refers to ensuring that patients have the right to make informed decisions about their treatment, including the use of BCIs. This involves providing comprehensive information about the potential benefits and risks of the technology. Beneficence involves striving to maximize the potential benefits of BCI technology for mental health patients, such as improved symptom management and quality of life.[116]

---

[111] Musk, "An Integrated Brain-Machine Interface."

[112] Tom L. Beauchamp and James F. Childress, *Principles of Biomedical Ethics* (Oxford University Press, 2019).

[113] Brandon J. King et al., "Prospectively Identifying Risks and Controls for Advanced Brain-Computer Interfaces: A Networked Hazard Analysis and Risk Management System (Net-HARMS) Approach," *Applied Ergonomics* 122 (2025): 104382.

[114] Rafael Yuste et al., "Four Ethical Priorities for Neurotechnologies and AI," *Nature* 551, no. 7679 (2017): 159-163.

[115] Beauchamp and Childress.

[116] Walter Glannon, "Ethical Issues with Brain-Computer Interfaces," *Frontiers in Systems Neu-*

Non-maleficence, the principle of "do no harm," is particularly relevant given the speculative nature of BCI technology in mental health treatment. This principle underscores the need for thorough testing and long-term studies before widespread implementation of BCI technology in mental health treatment.[117] The principle of justice ensures fair and equal access to BCI applications in mental health care, raising questions about the equitable distribution of these technologies and the potential for socioeconomic stratification in access to treatment.[118]

f. BCIs and philosophical questions of mind

The advancement of brain-computer interfaces (BCIs) raises intriguing possibilities for addressing long-standing philosophical questions, particularly the Mind-Body problem and the Problem of Other Minds. By creating a direct interface between the brain and external devices, BCIs offer a unique perspective on the relationship between mental phenomena and physical processes.[119] They may provide new insights into how mental states correspond to neural activity, potentially bridging the explanatory gap between subjective experience and objective brain function.[120] However, while BCIs may offer valuable data about the neural correlates of consciousness, they may not necessarily resolve the hard problem of consciousness – explaining why and how we have qualitative subjective experiences.[121]

Regarding the Problem of Other Minds, advanced BCIs could potentially offer new approaches by allowing for more direct communication of mental states between individuals.[122] If BCIs could transmit not just information but also subjective experiences or emotional states directly from one brain to another, it might provide a more immediate understanding of another's mental state.[123] However, it's crucial to ap-

roscience 8 (2014): 1-3.

[117] Yuste et al.

[118] Eran Klein et al., "Engineering the Brain: Ethical Issues and the Introduction of Neural Devices," *Hastings Center Report* 45, no. 6 (2015): 26-35.

[119] Jaegwon Kim, "Mind-body Problem," in *The Oxford Companion to Philosophy*, ed. Ted Honderich (Oxford University Press, 2005).

[120] Gualtiero Piccinini, *Neurocognitive Mechanisms: Explaining Biological Cognition* (Oxford University Press, 2020).

[121] David J. Chalmers, "Facing up to the Problem of Consciousness," *Journal of Consciousness Studies* 2, no. 3 (1995): 200-219.

[122] Alec Hyslop, *Other Minds* (Springer eBooks, 1995).

[123] Christopher Grau et al., "Conscious Brain-to-brain Communication in Humans Using Non-in-

proach these potential insights with caution. The ability to observe or manipulate neural activity does not necessarily equate to a complete understanding of consciousness or subjective experience.[124] Moreover, the interpretation of data from BCIs will always be filtered through our existing conceptual frameworks and scientific paradigms. As such, while BCIs may provide new tools for investigating these philosophical problems, they are unlikely to offer definitive solutions on their own.[125]

VI. Ethical considerations in BCI use for religious and psychological experiences

As BCI technology advances and its potential applications in religious and psychological contexts become more apparent, several critical ethical considerations emerge:

a. Cognitive liberty and mental privacy

The use of BCIs in religious contexts raises significant questions about cognitive liberty. If a religious organization encourages or requires the use of BCIs for certain practices, it could infringe on an individual's right to mental privacy and freedom of thought. There's also the risk of subtle coercion, where individuals feel pressured to use BCIs to fully participate in religious activities. This pressure could compromise the voluntary nature of religious practice and potentially violate the principle of freedom of religion.[126]

b. Potential for coercion or manipulation

BCIs open up unprecedented possibilities for direct influence on an individual's cognitive processes. In a religious context, this raises concerns about the potential for manipulation of beliefs or experiences. For instance, a BCI could theoretically be used to enhance feelings of spiritual conviction or alter perceptions during religious rituals. The line between facilitation of spiritual experiences and unethical manipulation could become blurred, necessitating careful ethical guidelines and oversight.[127] Furthermore, some argue that machines should not repli-

---

vasive Technologies," *PloS One* 9, no. 8 (2014): e105225.

[124] Thomas W. Polger and Lawrence A. Shapiro, *The Multiple Realization Book* (Oxford University Press, 2016).

[125] Michael S. A. Graziano, *Consciousness and the Social Brain* (Oxford University Press, 2013).

[126] Ienca and Andorno.

[127] Allen Coin et al., "Ethical Aspects of BCI Technology: What Is the State of the Art?" *Philos-*

cate human ethical behavior since humans are imperfect. This consideration becomes particularly relevant when BCIs are used to influence or modify religious and moral decision-making processes. The imperfect nature of human ethical reasoning suggests that technological systems should perhaps aim for different or higher ethical standards rather than simply mimicking human moral cognition.[128]

c. Equity and access issues

As with many emerging technologies, the development and distribution of BCIs raise questions of equity and access. If BCI-enhanced spiritual or psychological experiences become widely adopted, individuals or communities without access to these technologies might be at a disadvantage. This could potentially create new forms of spiritual or psychological inequality based on technological access. Religious and therapeutic institutions would need to consider how to ensure fair access and prevent the creation of 'technological elite' within their communities.[129]

d. Data privacy and security

The use of BCIs in religious or therapeutic contexts would involve the collection and processing of highly sensitive neural data. Ensuring the privacy and security of this data is crucial to protect individuals from potential misuse or unauthorized access. Religious organizations and mental health providers would need to develop robust data protection protocols and be transparent about how neural data is collected, used, and stored.[130]

e. Authenticity of experience

There are philosophical and ethical questions about the authenticity of BCI-mediated spiritual or psychological experiences. If a profound religious experience or psychological insight is facilitated or enhanced by technology, does this diminish its validity or significance? This consideration touches on fundamental questions about the nature of consciousness, spirituality, and human experience.[131]

---

ophies 5, no. 4 (2020): 1-9.

[128] Michael Anderson et al., "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson," *Conatus - Journal of Philosophy*, 6 no. 1, (2021): 177-202.

[129] Sara Goering and Rafael Yuste, "On the Necessity of Ethical Guidelines for Novel Neurotechnologies," *Cell* 167, no. 4 (2016): 882-885.

[130] Ethan Waisberg et al., "Correction: Ethical Considerations of Neuralink and Brain-Computer Interfaces," *Annals of Biomedical Engineering* 52 (2024): 1937-1939.

[131] Gabriel Fernandez Borsot, "Spirituality and Technology: A Threefold Philosophical Reflec-

These ethical considerations highlight the need for ongoing dialogue between neuroscientists, ethicists, religious leaders, and policymakers as BCI technology continues to develop. Establishing clear ethical guidelines and regulatory frameworks will be crucial to ensure that the integration of BCIs into religious and psychological domains respects individual rights, promotes equitable access, and preserves the integrity of spiritual and therapeutic practices.

VII. Implications for religious institutions and practices

The potential widespread adoption of brain-computer interfaces like those being developed by Neuralink could have profound implications for religious institutions and traditional spiritual practices. As these technologies reshape individual religious experiences and cognition, religious organizations will need to adapt to remain relevant and address new ethical and theological challenges.[132]

a. Challenges to traditional religious authority

One of the most significant implications of BCI technology for religious institutions is the potential challenge to traditional sources of religious authority. If individuals can access profound spiritual experiences or vast repositories of religious knowledge directly through neural interfaces, the role of religious leaders as mediators of divine wisdom or interpreters of sacred texts may be diminished.

Moreover, the potential for BCIs to induce mystical or transcendent states that have traditionally been the domain of advanced spiritual practitioners raises questions about the value and meaning of long-term spiritual discipline. Religious institutions may need to articulate new understandings of spiritual growth and attainment in a world where profound religious experiences can be technologically mediated.

b. Adaptation of religious teachings and practices

Religious institutions will likely need to adapt their teachings and practices to address the ethical and theological implications of BCI technology. This may involve developing new interpretations of sacred

tion," *Journal of Religion & Science* 58, no. 1 (2023): 6-22.

[132] Michael Inzlicht et al., "Neural Markers of Religious Conviction," *Psychological Science* 20, no. 3 (2009): 385-392.

texts and doctrines to accommodate the possibilities and challenges presented by direct neural interfaces.[133]

For instance, religious teachings on the nature of the soul, free will, and divine communication may need to be reexamined in light of the capabilities of BCI technology. As theologian Ted Peters (2015) suggests, "religious traditions will need to engage in serious theological reflection to determine how their core beliefs can be understood and expressed in a world of enhanced human-machine symbiosis."[134]

Additionally, traditional religious practices may evolve to incorporate BCI technology. We might see the emergence of new forms of technologically-mediated prayer, meditation, or communal worship. For example, future religious services could involve synchronized neural entrainment among congregants, or spiritual retreats might offer BCI-enhanced contemplative experiences.

For instance, in Christianity, the concept of humans being created in God's image (imago dei) is central to understanding human nature and dignity. Some theologians argue that cognitive enhancement through BCIs could be seen as a continuation of God-given abilities to improve ourselves, while others view it as potentially distorting the divine image.[135]

Some researchers have used neuroimaging techniques to study the brain activity of individuals during prayer, meditation, and other spiritual practices.[136] As BCI technology advances, it may offer even more detailed insights into these experiences, potentially challenging or affirming religious beliefs about the nature of spiritual encounters.[137]

c. Potential for new religious movements or techno-spiritual philosophies

The development of BCI technology may give rise to entirely new religious movements or techno-spiritual philosophies that fully embrace the possibilities of human-machine integration.[138] These new spiritual frameworks

---

[133] J. R. Schmid et al., "Thoughts Unlocked by Technology – A Survey in Germany About Brain-Computer Interfaces," *NanoEthics* 15 (2021): 303-313.

[134] Ted Peters, "Theologians Testing Transhumanism," *Theology and Science* 13, no. 2 (2015): 130-149.

[135] Ted Peters, "Imago Dei, DNA, and the Transhuman Way," *Theology and Science* 16, no. 3 (2018): 353-362.

[136] Kevin L Ladd and Meleah L. Ladd, "How God Changes Your Brain: Breakthrough Findings from a Leading Neuroscientist. By Andrew Newberg and Mark Robert Waldman," *The International Journal for the Psychology of Religion* 20, no. 3 (2010): 219-222.

[137] Ronald Cole Turner, *Transhumanism and Transcendence: Christian Hope in an Age of Technological Enhancement* (Georgetown University Press, 2011).

[138] Robert M. Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and*

might seek to reconcile scientific understanding of the brain with experiences of transcendence and meaning facilitated by neural interfaces.

Max Hodak, co-founder of Neuralink, has speculated about the potential for a "new religion" that combines scientific knowledge with technologically-induced profound experiences.[139] Such movements might view BCI technology as a tool for expanding consciousness and achieving new forms of spiritual insight or collective awareness.

These emerging techno-spiritual movements could pose both opportunities and challenges for traditional religious institutions. On one hand, they might attract individuals seeking a more scientifically-aligned approach to spirituality. On the other hand, they could be seen as competing with or undermining established religious traditions.

d. Legal and ethical challenges for religious organizations

The integration of BCI technology into religious practices will likely present novel legal and ethical challenges for religious organizations. Issues of informed consent, mental privacy, and cognitive liberty will need to be carefully navigated, particularly when it comes to the use of BCIs in religious education or spiritual counseling.[140]

Religious institutions may need to develop new ethical guidelines and policies regarding the use of BCI technology in spiritual contexts. This might include considerations around the appropriate use of neurofeedback in religious practices, the protection of individuals' mental privacy during technologically-mediated spiritual experiences, and safeguards against coercive uses of BCI technology in religious settings.[141]

e. Impact on religious education and transmission of tradition

BCI technology could dramatically transform approaches to religious education and the transmission of spiritual traditions. The ability to directly access or "download" religious knowledge could accelerate learning processes, potentially allowing for more rapid and comprehensive religious education.[142]

Virtual Reality (Oxford University Press, 2012).

[139] Tangermann.

[140] Rutger J. Vlek et al., "Ethical Issues in Brain-Computer Interface Research, Development, and Dissemination," *Journal of Neurologic Physical Therapy: JNPT* 36, no. 2 (2012): 94-99.

[141] Anwar Almofleh et al., "Brain Computer Interfaces: The Future of Communication Between the Brain and the External World," *Science, Engineering and Technology* 3, no. 2 (2023): 106-118.

[142] Christopher Wegemer, "Brain-Computer Interfaces and Education: The State of Technology and Imperatives for the Future," *International Journal of Learning Technology* 4, no. 2 (2019): 141-161.

However, this capability also raises questions about the value of traditional methods of study, contemplation, and gradual internalization of religious teachings. Many spiritual traditions emphasize the importance of long-term engagement with sacred texts and practices as a means of deepening understanding and fostering spiritual growth.

Religious educators and institutions will need to grapple with how to integrate the possibilities offered by BCI technology while preserving the formative aspects of traditional religious education. This may involve developing new pedagogical approaches that combine technologically-enhanced learning with traditional contemplative practices.[143]

VIII. Societal and cultural impact

The integration of Neuralink's brain-computer interfaces into society could profoundly shift our cultural landscape, reshaping collective attitudes toward faith, science, and human consciousness, particularly in relation to religion, spirituality, and psychology may reshape our collective attitudes towards faith, science, and the nature of human consciousness.[144]

a. Reshaping of cultural attitudes towards religion and spirituality

The widespread adoption of BCI technology could lead to a significant shift in how society at large views religion and spirituality.[145] As neuroscientific explanations for religious experiences become more prevalent and accessible through BCI-mediated insights, we may see a growing tension between materialist and spiritual interpretations of these phenomena.[146]

On one hand, the ability to induce or enhance spiritual experiences through technological means might lead to a form of "techno-spirituality" that blends scientific understanding with transcendent experienc-

[143] Bert Roebben and Klaus von Stosch, "Religious Education and Comparative Theology: Creating Common Ground for Intercultural Encounters," *Religions* 13, no. 11 (2022): 1-13.

[144] Tony Davenport, "Warnings About Brain Chip Technology," *Vision*, February 3, 2024, https://vision.org.au/news/warnings-about-brain-chip-technology/.

[145] Patrick McNamara, "Religion and the Brain: Jordan Grafman's Contributions to Religion and Brain Research and the Special Case of Religious Language," *Cortex* 169 (2023): 374-379.

[146] Manar Alohaly, "The Brain Computer Interface Market Is Growing – But What Are the Risks?" *World Economic Forum*, June 14, 2024, https://www.weforum.org/agenda/2024/06/the-brain-computer-interface-market-is-growing-but-what-are-the-risks/.

es.[147] This could potentially bridge the perceived gap between science and spirituality, leading to new syncretic worldviews.

Conversely, the technological mediation of spiritual experiences might lead to increased skepticism about the authenticity or value of religious experiences in general.[148] This could potentially accelerate trends of secularization in some societies, as traditionally mystical or transcendent experiences become more readily explainable in neurological terms.

b. Potential changes in the relationship between science and religion

The development of BCI technology may catalyse new dialogues and potential collaborations between scientific and religious communities.[149] As these technologies begin to touch on questions of consciousness, free will, and the nature of transcendent experiences, interdisciplinary exchanges between neuroscientists, philosophers, and theologians may become increasingly important and frequent.

This increased interaction could lead to what sociologist of religion Eileen Barker (2020) calls a "neuro-theological turn" in both scientific and religious discourse. We might see the emergence of new fields of study that attempt to reconcile neuroscientific insights with religious and spiritual perspectives on human nature and consciousness.[150]

However, this convergence of science and spirituality through BCI technology might also exacerbate existing tensions between scientific and religious worldviews.[151] Some religious communities may view the technological manipulation of spiritual experiences as a threat to traditional beliefs and practices, potentially leading to new forms of religious resistance to scientific and technological advancement.[152]

---

[147] Elizabeth Buie, "Let Us Say What We Mean: Towards Operational Definitions for Techno-Spirituality Research," in *CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-10 (Association for Computing Machinery, 2019).

[148] Iddo Landau, "The Case for Technological Mysticism," *Journal of Posthuman Studies* 2, no. 1 (2018): 67-85.

[149] Janis Peksa and Dmytro Mamchur, "State-of-the-Art on Brain-Computer Interface Technology," *Sensors* 23, no. 13 (2023): 1-28.

[150] Joyce Ann Konigsburg, "Scientific Wonder, Artificial Intelligence, and Awe of the Divine," *Religions* 15, no. 4 (2024): 1-12.

[151] Massimo Leone, "Technology and Sacrifice," *Religions* 15, no. 6 (2024): 1-17.

[152] Jennifer Karns Alexander, "Introduction: The Entanglement of Technology and Religion," *History and Technology* 36, no. 2 (2020): 165-186.

c. Impact on social cohesion and religious diversity

The potential for BCIs to enable shared or collective spiritual experiences could have significant implications for social cohesion and religious diversity.[153] However, it's important to note that the complexity of spiritual experiences may not be fully replicable through technological means alone. Critics argue that artificially induced states may lack the depth, context, and personal significance of naturally occurring spiritual experiences. This tension between technologically mediated and traditional spiritual experiences raises important questions about authenticity and the nature of religious experience in a world of advancing neurotechnology.[154] On one hand, the ability to directly share subjective spiritual states might foster greater empathy and understanding between individuals from different faith traditions.[155] As religious studies scholar William Grassie (2010) suggests, "technologically-mediated shared spiritual experiences could potentially transcend traditional religious boundaries, fostering a more universal sense of spirituality."[156] This perspective aligns with the concept of "neurotheology," which explores the neural correlates of religious and spiritual experiences.[157]

However, this technology could also lead to new forms of religious polarization or extremism.[158] The ability to create immersive, shared ideological experiences through BCIs might be used to reinforce group identities and beliefs, potentially exacerbating religious divides.[159] There are also concerns about the potential for BCI technology to be used for religious indoctrination or thought control, raising significant ethical and societal concerns.

d. Legal and regulatory challenges

The development and deployment of brain-computer interface (BCI)

---

[153] Michael N. Tennison, "Moral Transhumanism: The Next Step," *The Journal of Medicine and Philosophy* 37, no. 4 (2012): 405-416.

[154] Alkhouri.

[155] M. Andersen et al., "Mystical Experience in the Lab," *Method & Theory in the Study of Religion* 26, no. 3 (2014): 217-245.

[156] William Grassie, *The New Sciences of Religion: Exploring Spirituality from the Outside In and Bottom Up* (Palgrave Macmillan, 2010), 215.

[157] Andrew Newberg, *Neurotheology: How Science Can Enlighten Us about Spirituality* (Columbia University Press, 2018).

[158] Ienca and Andorno.

[159] Robert M. Sapolsky, *Behave: The Biology of Humans at Our Best and Worst* (Penguin, 2017).

technology, such as that being developed by Neuralink, raises significant legal and regulatory challenges. These challenges are particularly complex when considering the use of BCIs in contexts related to religious and psychological experiences.[160]

One key area of concern is the regulation of BCIs as medical devices.[161] The European Union has taken steps to incorporate some non-medical BCIs into its regulatory framework through Annex XVI of the Medical Devices Regulation (MDR).[162] This regulation covers "equipment intended for brain stimulation that apply electrical currents or magnetic or electromagnetic fields that penetrate the cranium to modify neuronal activity in the brain." However, it's important to note that this only applies to non-invasive, transcranial devices, not invasive BCIs like those being developed by Neuralink.[163]

There is an ongoing debate about whether new "neurorights" are needed to protect cognitive liberty and mental privacy in the age of BCIs.[164] This suggests that existing legal frameworks may be insufficient to address the unique challenges posed by direct brain-computer interfaces.[165]

Issues of mental privacy, cognitive liberty, and freedom of religion will need to be carefully considered in light of these new technological capabilities.[166] There may be a need for new legal protections to safeguard individuals' right to control their own mental processes and to prevent unwanted interference or manipulation through BCI technology.[167]

Furthermore, the potential use of BCIs in religious or therapeutic contexts may require the development of specific regulatory guidelines

---

[160] Colin Conrad and Carla Heggie, "Legal and Ethical Challenges Raised by Advances in Brain-Computer Interface Technology," *SSRN* (2024): 1-18.

[161] Xue-Qin Wang et al., "Challenges and Suggestions of Ethical Review on Clinical Research Involving Brain-Computer Interfaces," *Chinese Medical Sciences Journal* 39, no. 2 (2024): 131-139.

[162] Albert Manero et al., "Emerging Medical Technologies and Their Use in Bionic Repair and Human Augmentation," *Bioengineering* 11, no. 7 (2024): 1-36.

[163] European Parliament and Council, "Regulation (EU) 2017/745 on Medical Devices," *Official Journal of the European Union*, L117 (2017): 1-175.

[164] Christoph Bublitz, "Neurotechnologies and Human Rights: Restating and Reaffirming the Multi-Layered Protection of the Person," *The International Journal of Human Rights* 28, no. 5 (2024): 782-807.

[165] Stephen Rainey et al., "Brain Recording, Mind-Reading, and Neurotechnology: Ethical Issues from Consumer Devices to Brain-Based Speech Decoding," *Science and Engineering Ethics* 26, no. 4 (2020): 2295-2311.

[166] Tong-Kuo Zhang, "Perspective and Boundary Exploration of Privacy Transfer Dilemma in Brain-Computer Interface – Dimension Based on Ethical Matrix," *Philosophies* 9, no. 1 (2024): 1-9.

[167] Ienca and Andorno.

to prevent potential abuses or exploitation. This could involve considerations around informed consent, especially when it comes to the use of BCIs in religious education or spiritual counseling.[168]

As BCI technology continues to advance, it will be crucial for policymakers, ethicists, and legal experts to work together to develop comprehensive and nuanced regulatory frameworks that can keep pace with these rapidly evolving technologies.[169]

e. Economic and social implications

The development and potential widespread adoption of BCI technology could have significant economic and social implications. If these technologies provide substantial cognitive or experiential enhancements, they could create new forms of social and economic stratification based on access to neural augmentation.[170]

In the context of religion and spirituality, this could lead to what some scholars have termed "neuro-spiritual inequality," where access to technologically-mediated transcendent experiences becomes a new marker of privilege. Religious institutions and society at large will need to grapple with how to ensure equitable access to these technologies and prevent the exacerbation of existing social disparities.[171]

f. Shifting notions of human nature and identity

Finally, the integration of BCI technology into religious and psychological domains may lead to fundamental shifts in how we conceive of human nature and identity.[172] As the boundaries between mind and machine become increasingly blurred, traditional notions of the self, consciousness, and even the soul may need to be reconsidered.[173]

---

[168] Sasha Burwell et al., "Ethical Aspects of Brain Computer Interfaces: A Scoping Review," *BMC Medical Ethics* 18, no. 1 (2017): 1-11; Sasha Burwell, "Ethical Aspects of Brain Computer Interfaces: A Scoping Review," *BMC Medical Ethics* 18, no. 1 (2017): 1-11.

[169] Sedat Sonko et al., "Neural Interfaces and Human-Computer Interaction: A U.S. Review: Delving into the Developments, Ethical Considerations, and Future Prospects of Brain-Computer Interfaces," *International Journal of Science and Research Archive* 11, no. 1 (2024): 702-717.

[170] Allen Coin et al., "Ethical Aspects of BCI Technology: What Is the State of the Art?" *Philosophies* 5, no. 4 (2020): 1-9.

[171] E. Mohandas, "Neurobiology of Spirituality," *Mens Sana Monographs* 6, no. 1 (2008): 63-80.

[172] Sonja C. Kleih and Andrea Kübler, "Psychological Factors Influencing Brain-Computer Interface (BCI) Performance," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 3192-3196 (IEEE, 2015).

[173] Dongyang Li, "Blurring Human and Machine Boundary: The Post-Humanist Metaphor of Cyborg-Body in Artificial Intelligence and Minority Report," in *Proceedings of the 2020 Inter-*

This evolving understanding of human nature could have far-reaching implications for our social, legal, and ethical systems, many of which are grounded in particular conceptions of human agency and identity. As philosopher Andy Clark (2003) argues, "the integration of neural interfaces into human cognition may require us to develop new, more fluid conceptions of personhood and identity that can accommodate our increasingly hybrid nature."[174]

IX. Research results

Our comprehensive investigation into Neuralink's brain-computer interfaces and their potential impact on religious-psychological experiences has yielded a range of significant findings. These results span multiple domains, including neuroscience, religious studies, psychology, and ethics. The following points summarize the key outcomes of our research, offering insights into the complex interplay between advanced neurotechnology and human spirituality. These findings not only shed light on the current state of BCI technology and its implications but also point towards future developments and challenges in this rapidly evolving field.

The research results from this study on Neuralink's brain-computer interfaces (BCIs) and their impact on religious-psychological experiences reveal a complex landscape of potential transformations in human spirituality, cognition, and social structures. Key findings include:

a. Altered states of consciousness: BCIs show potential to induce and enhance altered states of consciousness traditionally associated with spiritual and mystical experiences. This capability could democratize access to profound spiritual states, but also raises questions about the authenticity and value of technologically-mediated experiences compared to naturally occurring ones.
b. Enhanced meditation and contemplative practices: BCI technology could significantly augment meditation and other contemplative practices through real-time neurofeedback and neural entrainment. While this may accelerate the development of meditative skills, it also challenges traditional notions of spiritual discipline and effort.

---

national Conference on Language, Art and Cultural Exchange (ICLACE 2020), 47-50 (Springer Nature, 2020).

[174] Andy Clark, *Natural-born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford University Press, 2003).

c. Redefinition of religious rituals and practices: The integration of BCIs into religious contexts could lead to novel forms of rituals and practices, potentially transforming how individuals and communities engage with spiritual concepts and experiences. This may necessitate a reevaluation of traditional religious frameworks and doctrines.
d. Psychological and cognitive implications: BCIs have the potential to profoundly alter perception, cognition, and emotional regulation. This could lead to enhanced cognitive abilities and mood management, but also raises concerns about cognitive liberty and the nature of authentic emotional and spiritual experiences.
e. Challenges to religious institutions: The widespread adoption of BCI technology could challenge traditional religious authorities and structures, potentially democratizing spiritual experiences and knowledge. This may require religious institutions to adapt their roles and teachings to remain relevant in a technologically-enhanced spiritual landscape.
f. Ethical considerations: The research highlights significant ethical challenges, including issues of cognitive liberty, mental privacy, potential for manipulation, and equitable access to BCI technology in religious and spiritual contexts. These concerns underscore the need for robust ethical frameworks and guidelines.
g. Societal and cultural impact: The integration of BCIs into religious and spiritual practices could have far-reaching societal implications, potentially reshaping cultural attitudes towards religion, spirituality, and the relationship between science and faith. This may lead to new forms of techno-spiritual philosophies and movements.
h. Neurotheological insights: The study contributes to the emerging field of neurotheology, offering new perspectives on the neural correlates of religious experiences and the potential for technology to interact with and possibly enhance spiritual states.
i. Identity and consciousness: BCIs challenge traditional notions of self, consciousness, and human identity, particularly in the context of religious and spiritual beliefs about the soul or essential self. This may necessitate a philosophical reevaluation of what it means to be human in an era of brain-machine symbiosis.
j. Future trajectories: The research points to several critical areas for future investigation, including long-term studies on the psy-

chological effects of BCI use in spiritual practices, comparative analyses of natural versus BCI-induced spiritual experiences, and explorations of how BCI technology might impact religious belief systems and institutions over time.

These findings collectively underscore the transformative potential of BCI technology in the realm of religious and psychological experiences, while also highlighting the complex ethical, philosophical, and societal challenges that accompany these advancements. The research suggests that as BCI technology continues to evolve, it will likely play an increasingly significant role in shaping the future landscape of human spirituality and consciousness, necessitating ongoing interdisciplinary dialogue and careful consideration of its implications.

X. Future research directions

As our study has revealed, the intersection of brain-computer interfaces and religious-psychological experiences is a rich and complex area that warrants further investigation. The following research directions emerge as particularly promising avenues for future study. These proposed areas of research aim to address critical questions raised by our findings, explore emerging phenomena, and contribute to the development of ethical frameworks for the responsible advancement of BCI technology in spiritual and psychological contexts. By pursuing these lines of inquiry, researchers can continue to expand our understanding of how neurotechnology may reshape human consciousness and spiritual experiences in the coming years.

As the field of brain-computer interfaces and their applications in religious and psychological contexts continues to evolve, several key areas emerge as priorities for future research:

a. Long-term Psychological Effects of BCI Use in Spiritual Practices: Future research could explore the long-term psychological effects of regular BCI use in spiritual practices. Longitudinal studies tracking individuals over several years could provide insights into how technologically mediated spiritual experiences might shape religious beliefs and practices over time. Such studies could examine changes in religious conviction, spiritual well-being, and overall psychological health among regular users of BCI-enhanced spiritual practices.

b. Neurological Differences Between Natural and BCI-Induced Spiritual Experiences: Investigations into the neurological differ-

ences between naturally occurring and BCI-induced spiritual experiences are crucial. Comparative studies using advanced neuroimaging techniques could help elucidate whether technologically mediated experiences activate the same neural pathways as spontaneous spiritual experiences. This research could shed light on questions of authenticity and the nature of religious experiences.
c. Impact of BCI Technology on Religious Belief Systems and Institutions: Research on how the adoption of BCI technology affects religious belief systems and institutions is needed. This could include sociological studies on how religious communities adapt to and incorporate BCI technology, as well as examinations of potential changes in religious doctrine or practice in response to these technological advancements.
d. Ethical Frameworks for BCI Use in Religious Contexts: Development of comprehensive ethical frameworks specifically addressing the use of BCIs in religious and spiritual contexts is an important area for future work. This could involve interdisciplinary collaborations between ethicists, religious scholars, neuroscientists, and legal experts to establish guidelines for the responsible use of this technology in spiritual practices.
e. Cross-Cultural Studies on BCI Acceptance in Religious Practices: Given the global diversity of religious traditions, cross-cultural studies examining the acceptance and integration of BCI technology in various religious contexts would be valuable. This research could explore how different cultural and religious backgrounds influence attitudes towards and adoption of BCI-enhanced spiritual practices.
f. Potential Therapeutic Applications of BCI-Enhanced Spiritual Experiences: Investigation into the potential therapeutic benefits of BCI-enhanced spiritual experiences in treating mental health conditions like depression, anxiety, or addiction could be a fruitful area of research. This could build on existing research on the mental health benefits of spiritual practices, exploring how BCI technology might enhance these effects.

These research directions highlight the complex and multifaceted nature of the intersection between BCI technology, spirituality, and psychology. As this field continues to develop, ongoing research will be crucial in understanding the full implications of these technologies and guiding their responsible development and use.

XI. Conclusion

The advent of Neuralink's brain-computer interface technology stands poised to profoundly reshape the landscape of religious and psychological experiences. This research has illuminated the complex interplay between cutting-edge neurotechnology and human spirituality, revealing both transformative potential and significant ethical challenges.

Our findings suggest that BCIs could dramatically alter how individuals engage with transcendent states, potentially democratizing access to profound spiritual experiences while simultaneously raising questions about their authenticity and value. The ability to technologically mediate or enhance religious practices, from meditation to collective worship, may lead to a paradigm shift in how spirituality is experienced and expressed.

However, these advancements do not come without concerns. The ethical implications of BCIs in religious contexts are far-reaching, touching on issues of cognitive liberty, mental privacy, and the potential for manipulation of beliefs. As these technologies progress, it becomes increasingly crucial to develop robust ethical frameworks and guidelines to ensure their responsible use.

Furthermore, the potential societal impacts of widespread BCI adoption in spiritual domains are profound. We may witness the emergence of new techno-spiritual philosophies, shifts in religious authority structures, and evolving notions of human consciousness and identity. These changes could reshape the relationship between science and religion, potentially bridging long-standing divides or creating new points of tension.

As we stand at the threshold of this neurotechnological revolution, it is clear that the implications extend far beyond the realm of medical applications. The future of human spirituality and consciousness may be intimately intertwined with our ability to interface directly with our neural processes. This research underscores the need for ongoing interdisciplinary dialogue and careful consideration as we navigate this uncharted territory.

In conclusion, while Neuralink's BCI technology offers unprecedented opportunities for enhancing and exploring human spiritual and psychological experiences, it also presents us with profound ethical and philosophical challenges. As we move forward, it is imperative that we approach these advancements with both excitement for their potential and mindfulness of their implications, ensuring that the future of human-machine symbiosis respects the depth and diversity of human spiritual experience.

References

Alexander, Jennifer Karns. "Introduction: The Entanglement of Technology and Religion." *History and Technology* 36, no. 2 (2020): 165-186.

Alkhouri, Khader I. "The Role of Artificial Intelligence in the Study of the Psychology of Religion." *Religions* 15, no. 3 (2024): 1-27.

Almofleh, Anwar, Mohamed Alseddiqi, Osama Najam, Leena Albalooshi, Abdulla Alheddi, and Ahmed Alshaimi. "Brain Computer Interfaces: The Future of Communication Between the Brain and the External World." *Science, Engineering and Technology* 3, no. 2 (2023): 106-118.

Alohaly, Manar. "The Brain Computer Interface Market Is Growing – But What Are the Risks?" *World Economic Forum*, June 14, 2024. https://www.weforum.org/agenda/2024/06/the-brain-computer-interface-market-is-growing-but-what-are-the-risks/.

Andersen, M., U. Schjoedt, K. L. Nielbo, and J. Sorensen. "Mystical Experience in the Lab." *Method & Theory in the Study of Religion* 26, no. 3 (2014): 217-245.

Anderson, Michael, Susan Leigh Anderson, Alkis Gounaris, and George Kosteletos. "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 177-202.

Angotzi, Gian Nicola, Fabio Boi, Aziliz Lecomte, Ermanno Miele, Mario Malerba, Stefano Zucca, Antonino Casile, and Luca Berdondini. "SiNAPS: An Implantable Active Pixel Sensor CMOS-Probe for Simultaneous Large-Scale Neural Recordings." *Biosensors and Bioelectronics* 126 (2019): 355-364.

Armstrong, William, and Katina Michael. "The Implications of Neuralink and Brain Machine Interface Technologies." In *2020 IEEE International Symposium on Technology and Society (ISTAS)*, 201-203. USA, 2020.

Atillah, Imane El. "Neuralink's Brain Chip: How Brain-Computer Interfaces May Revolutionise Treatment for Disabilities." *Euronews*, June 8, 2024. https://www.euronews.com/health/2024/06/06/neuralinks-brain-chip-how-brain-computer-interfaces-may-revolutionise-treatment-for-disabi.

Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*. Oxford University Press, 2019.

Beauregard, Mario, and Vincent Paquette. "Neural Correlates of a Mystical Experience in Carmelite Nuns." *Neuroscience Letters* 405, no. 3 (2006): 186-190.

Bergeron, David, Christian Iorio Morin, Marco Bonizzato, Guillaume Lajoie, Nathalie Orr Gaucher, Eric Rachine, and Alexander G. Weil. "Use of Invasive Brain-Computer Interfaces in Pediatric Neurosurgery: Technical and Ethical Considerations." *Journal of Child Neurology* 38, no. 3-4 (2023): 223-238.

Bettis, Alexandra H., Taylor A. Burke, Jacqueline Nesi, and Richard T. Liu. "Digital Technologies for Emotion-Regulation Assessment and Intervention: A Conceptual Review." *Clinical Psychological Science: A Journal of the Association for Psychological Science* 10, no. 1 (2022): 3-26.

Bingaman, Kirk A. "Religion in the Digital Age: An Irreversible Process." *Religions* 14, no. 1 (2023): 1-14.

Borsot, Gabriel Fernandez. "Spirituality and Technology: A Threefold Philosophical Reflection." *Journal of Religion & Science* 58, no. 1 (2023): 6-22.

Bostrom, Nick. "A History of Transhumanist Thought." *Journal of Evolution and Technology* 14, no. 1 (2005): 1-25.

Bublitz, Christoph. "Neurotechnologies and Human Rights: Restating and Reaffirming the Multi-Layered Protection of the Person." *The International Journal of Human Rights* 28, no. 5 (2024): 782-807.

Buie, Elizabeth. "Let Us Say What We Mean: Towards Operational Definitions for Techno-Spirituality Research." In *CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-10. Association for Computing Machinery, 2019.

Burke, John F., Maxwell B. Merkow, Joshua Jacobs, Michael J. Kahana, and Kareem A. Zaghloul. "Brain-Computer Interface to Enhance Episodic Memory in Human Participants." *Frontiers in Human Neuroscience* 8 (2015): 1-10.

Burwell, Sasha, Matthew Sample, and Eric Racine. "Ethical Aspects of Brain Computer Interfaces: A Scoping Review." *BMC Medical Ethics* 18, no. 1 (2017): 1-11.

Camus, Albert. *The Myth of Sisyphus and Other Essays*. Translated by Justin O'Brien. Vintage Books, 1991.

Capoot, Ashley. "Elon Musk Shows Off Updates to His Brain Chips and Says He's Going to Install One in Himself When They Are Ready." *CNBC*, December 1, 2022. https://www.cnbc.com/2022/12/01/elon-musks-neuralink-makes-big-claims-but-experts-are-skeptical-.html.

Chalmers, David J. "The Puzzle of Conscious Experience." *Scientific American* 273, no. 6 (1995): 80-86.

Chalmers, David J. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2, no. 3 (1995): 200-219.

Chalmers, David J. *The Character of Consciousness*. Oxford University Press, 2010.

Chaudhary, Ujwal, Niels Birbaumer, and Ander Ramos Murguialday. "Brain-Computer Interfaces for Communication and Rehabilitation." *Nature Reviews Neurology* 12 (2016): 513-525.

Chung, Jason E., Hannah R. Joo, Jiang Lan Fan, Daniel F. Liu, Alex H. Barnett, Supin Chen, Charlotte Geaghan-Breiner, Mattias P. Karlsson, Magnus Karlsson, Kye Y. Lee, Hexin Liang, Jeremy F. Magland, Jeanine A. Pebbles, Angela C. Tooker, Leslie F. Greengard, Vanessa M. Tolosa, and Loren M. Frank. "High-Density, Long-Lasting, and Multi-Region Electrophysiological Recordings Using Polymer Electrode Arrays." *Neuron* 101, no. 1 (2019): 21-31.

Clark, Andy, and David Chalmers. "The Extended Mind." *Analysis* 58, no. 1 (1998): 7-19.

Clark, Andy. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press, 2003.

Coin, Allen, Megan Mulder, and Veljko Dubljevič . "Ethical Aspects of BCI Technology: What Is the State of the Art?" *Philosophies* 5, no. 4 (2020): 1-9.

Colucc, Annalisa, Mareike Vermehren, Alessia Cavallo, Cornelius Angerhöfer, Niels Peekhaus, Loredana Zollo, Won-Seok Kim, Nam-Jong Paik, and Surjo R. Soekadar. "Brain-Computer Interface-Controlled Exoskeletons in Clinical Neurorehabilitation: Ready or Not?" *Neurorehabilitation and Neural Repair* 36, no. 12 (2022): 747-756.

Conrad, Colin, and Carla Heggie. "Legal and Ethical Challenges Raised by Advances in Brain-Computer Interface Technology." *SSRN* (2024): 1-18.

Davenport, Tony. "Warnings About Brain Chip Technology." *Vision*, February 3, 2024. https://vision.org.au/news/warnings-about-brain-chip-technology/.

Denison, Jim. "Elon Musk's Neuralink Implants Brain Chip in Human: Four Biblical Responses." *DenisonForum*, February 1, 2024. https://www.denisonforum.org/daily-article/elon-musks-neuralink-implants-brain-chip-in-human/.

Drew, Liam. "Neuralink Brain Chip: Advance Sparks Safety and Secrecy Concerns." *Nature* 627, no. 8002 (2024): 19-20.

Eagleman, David. *Livewired: The Inside Story of the Ever-Changing Brain*. Knopf Doubleday Publishing Group, 2021.

Elfstrom, Gerard. "The Theft: An Analysis of Moral Agency." *Conatus – Journal of Philosophy* 5, no. 1 (2020): 27-53.

European Parliament and Council. "Regulation (EU) 2017/745 on Medical Devices." *Official Journal of the European Union*, L117 (2017): 1-175.

Fahmy, Dalia. "Highly Religious Americans More Skeptical of Human Enhancements Such as Brain Implants, Gene Editing." *Pew Research Center*, May 4, 2022. https://pewrsr.ch/3kD3SGW.

Fajuri, Alejandra Zúñiga, Luis Villavicencio Miranda, Danielle Zaror Miralles, and Ricardo Salas Venegas. "Chapter Seven – Neurorights in Chile: Between Neuroscience and Legal Science." In *Developments in Neuroethics and Bioethics*, vol. 4, edited by Martín Hevia, 165-179. Academic Press, 2021.

Faruka, Jobair Hossain, Maria Valero, and Hossain Shahriar. "An Investigation on Non-Invasive Brain-Computer Interfaces: Emotiv Epoc+ Neuroheadset and Its Effectiveness." In *IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, 580-589. Madrid, 2021.

Fiani, Brian, Taylor Reardon, Benjamin Ayres, David Cline, and Sarah R. Sitto. "An Examination of Prospective Uses and Future Directions of Neuralink: The Brain-Machine Interface." *Cureus* 13, no. 3 (2021): 1-4.

Frankl, Victor E. *Man's Search for Meaning*. Beacon Press, 2006.

Fridman, Lex, and Elon Musk. "Neuralink Human Trial: Nolan Arbaugh & Elon Musk | Lex Fridman Podcast." YouTube, August 2, 2024. https://youtu.be/Kbk9BiPhm7o?feature=shared.

Fukuyama, Francis. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. Farrar, Straus and Giroux, 2002.

Gent, Edd. "Brain-Computer Interfaces Are Coming: 'Consensual Telepathy', Anyone?" *Washington Post*, June 11, 2017. https://www.washingtonpost.com/national/health-science/brain-computer-interfaces-are-coming-consensual-telepathy-anyone/2017/06/09/9345c682-46ef-11e7-98cd-af64b4fe2dfc_story.html.

Geraci, Robert M. "Spiritual Robots: Religion and Our Scientific View of the Natural World." *Theology and Science* 4, no. 3 (2006): 229-246.

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

Giatti, Ian M. "Healing the Lame, Bringing Sight to the Blind? Elon Musk's Ambitions for Neuralink Raise 'Deep, Serious' Questions (Part 1)." *The Christian Post*, December 26, 2022. https://www.christianpost.com/news/elon-musk-ambitions-for-neuralink-raise-deep-serious-questions.html.

Gilbert, F., M. Cook, T. O'Brien, and J. Illes. "Embodiment and Estrangement: Results from a First-in-Human 'Intelligent BCI' Trial." *Science and Engineering Ethics* 25, no. 1 (2019): 83-96.

Glannon, Walter. "Ethical Issues with Brain-Computer Interfaces." *Frontiers in Systems Neuroscience* 8 (2014): 1-3.

Goering, Sara, and Eran Klein. "Neurotechnologies and Justice by, with, and for Disabled People." In *The Oxford Handbook of Philosophy and Disability*, edited by D. T. Wasserman and A. Cureton, 616-632. Oxford University Press, 2019.

Goering, Sara, and Rafael Yuste. "On the Necessity of Ethical Guidelines for Novel Neurotechnologies." *Cell* 167, no. 4 (2016): 882-885.

Grafman, Jordan, Irene Cristofori, Wanting Zhong, and Joseph Bulbulia. "The Neural Basis of Religious Cognition." *Current Directions in Psychological Science* 29, no. 2 (2020): 126-133.

Grassie, William. *The New Sciences of Religion: Exploring Spirituality from the Outside in and Bottom up*. Palgrave Macmillan, 2010.

Grau, Christopher, Romuald Ginhoux, Alejandro Riera, Thanh Lam Nguyen, Hubert Chauvat, Michel Berg, Julià L. Amengual, Alvaro Pascual-Leone, and Giulio Ruffini. "Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies." *PloS One* 9, no. 8 (2014): e105225.

Graziano, Michael S. A. *Consciousness and the Social Brain*. Oxford University Press, 2013.

Hadot, Pierre. *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault*. Blackwell, 1995.

Hamilton, John R., Stephen J. Maxwell, Syeda Arfa Ali, and Singwhat Tee. "Adding External Artificial Intelligence (AI) into Internal Firm-Wide Smart Dynamic Warehousing Solutions." *Sustainability* 16, no. 10 (2024): 1-23.

Harris, Sam, Jonas T. Kaplan, Ashley Curiel, Susan Y. Bookheimer, Marco Iacoboni, and Mark S. Cohen. "The Neural Correlates of Religious and Nonreligious Belief." *PLoS One* 5, no. 1 (2010): 1-10.

Hochberg, Leigh R., Mijail D. Serruya, Gerhard M. Friehs, Jon A. Mukand, Maryam Saleh, Abraham H. Caplan, Almut Branner, David Chen, Richard D. Penn, and John P. Donoghue. "Neuronal Ensemble Control of Prosthetic Devices by a Human with Tetraplegia." *Nature* 442 (2006): 164-171.

Hong, Soonkwan. "Transcendence Up for Sale: Cracking the Onto-Existential Codes for Übermensch." *Consumption Markets & Culture* 27, no. 2 (2024): 152-177.

Hyslop, Alec. *Other Minds*. Springer, 1995.

Ienca, Marcello, and Roberto Andorno. "Towards New Human Rights in the Age of Neuroscience and Neurotechnology." *Life Sciences, Society and Policy* 13, no. 1 (2017): 1-27.

Inzlicht, Michael, Ian McGregor, Jacob B. Hirsh, and Kyle Nash. "Neural Markers of Religious Conviction." *Psychological Science* 20, no. 3 (2009): 385-392.

Johnson, Matthew D., Hubert H. Lim, Theoden I. Netoff, Allison T. Connolly, Nessa Johnson, Abhrajeet Roy, Abbey Holt, Kelvin O. Lim, James R. Carey, Jerrold L. Vitek, and Bin He. "Neuromodulation for Brain Disorders: Challenges and Opportunities." *IEEE Transactions on Bio-Medical Engineering* 60, no. 3 (2013): 610-624.

Kane, Robert. *A Contemporary Introduction to Free Will*. Oxford University Press, 2005.

Karikari, Evelyn, and Konstantin A. Koshechkin. "Review on Brain-Computer Interface Technologies in Healthcare." *Biophysical Reviews* 15, no. 5 (2023): 1351-1358.

Khan, Abdullah Ayub, Asif Ali Laghari, Aftab Ahmed Shaikh, Mazhar Ali Dootio, Vania V. Estrela, and Ricardo Tadeu Lopes. "A Blockchain Security Module for Brain-Computer Interface (BCI) with Multimedia Life Cycle Framework (MLCF)." *Neuroscience Informatics* 2, no. 1 (2022): 1-14.

Kim, Jaegwon. "Mind-body problem." In *The Oxford Companion to Philosophy*, edited by Ted Honderich. Oxford University Press, 2005.

King, Brandon J., Gemma J.M. Read, and Paul M. Salmon. "Prospectively Identifying Risks and Controls for Advanced Brain-Computer Interfaces: A Networked Hazard Analysis and Risk Management System (Net-HARMS) Approach." *Applied Ergonomics* 122 (2025): 104382.

Kleih, Sonja C., and Andrea Kübler. "Psychological Factors Influencing Brain-Computer Interface (BCI) Performance." In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 3192-3196. IEEE, 2015.

Klein, Eran, Tim Brown, Matthew Sample, Anjali R. Truitt, and Sara Goering. "Engineering the Brain: Ethical Issues and the Introduction of Neural Devices." *Hastings Center Report* 45, no. 6 (2015): 26-35.

Konigsburg, Joyce Ann. "Scientific Wonder, Artificial Intelligence, and Awe of the Divine." *Religions* 15, no. 4 (2024): 1-12.

Kormas, Panagiotis, Antonia Moutzouri, and Evangelos D. Protopapadakis. "Implications of Neuroplasticity to the Philosophical Debate of Free Will and Determinism." In *Handbook of Computational Neurodegeneration*, edited by P. Vlamos, I. S. Kotsireas, I. Tarnanas, 453-471. Spriner, 2023.

Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. Penguin, 2012.

Ladd, Kevin L., and Meleah L. Ladd. "How God Changes Your Brain: Breakthrough Findings from a Leading Neuroscientist. By Andrew Newberg and Mark Robert Waldman." *The International Journal for the Psychology of Religion* 20, no. 3 (2010): 219-222.

Landau, Iddo. "The Case for Technological Mysticism." *Journal of Posthuman Studies* 2, no. 1 (2018): 67-85.

Langley, Birgitta Dresp. "Consciousness as the Final Beacon of Humanity." *Recent Research Advances in Arts and Social Studies* 8 (2024): 118-154.

Leone, Massimo. "Technology and Sacrifice." *Religions* 15, no. 6 (2024): 1-17.

Lewis, C.S. *The Problem of Pain*. HarperOne, 2001.

Li, Dongyang. "Blurring Human and Machine Boundary--The Post-Humanist Metaphor of Cyborg-Body in Artificial Intelligence and Minority Report." In *Proceedings of the 2020 International Conference on Language, Art and Cultural Exchange (ICLACE 2020)*, 47-50. Springer Nature, 2020.

Lutz, Antoine, Lawrence L. Greischar, Nancy B. Rawlings, Matthieu Ricard, and Richard J. Davidson. "Long-Term Meditators Self-Induce High-Amplitude Gamma Synchrony During Mental Practice." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 46 (2004): 16369-16373.

Maiseli, Baraka, Abdi T. Abdalla, Libe V. Massawe, Mercy Mbise, Khadija Mkocha, Nassor Ally Nassor, Moses Ismail, James Michael, and Samwel Kimambo. "Brain-Computer Interface: Trend, Challenges, and Threats." *Brain Informatics* 10, no. 1 (2023): 1-16.

Manero, Albert, Viviana Rivera, Qiushi Fu, Jonathan D. Schwartzman, Hannah Prock-Gibbs, Neel Shah, Deep Gandhi, Evan White, Kaitlyn E. Crawford, and Melanie J. Coathup. "Emerging Medical Technologies and Their Use in Bionic Repair and Human Augmentation." *Bioengineering* 11, no. 7 (2024): 1-36.

McNamara, Patrick. "Religion and the Brain: Jordan Grafman's Contributions to Religion and Brain Research and the Special Case of Religious Language." *Cortex* 169 (2023): 374-379.

Menčik, David. "Identity Theft: A Thought Experiment on the Fragility of Identity." *Conatus – Journal of Philosophy* 5, no. 1 (2020): 71-83.

Mohandas, E. "Neurobiology of Spirituality." *Mens Sana Monographs* 6, no. 1 (2008): 63-80.

Moses, David A., Sean L. Metzger, Jessie R. Liu, Gopala K. Anumanchipalli, Joseph G. Makin, Pengfei F. Sun, Josh Chartier, and Edward F. Chang. "Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria." *The New England Journal of Medicine* 385, no. 3 (2021): 217-227.

Muller, Michael, Ellen Christiansen, Bonnie Nardi, and Susan Dray. "Spiritual Life and Information Technology." *Communications of the ACM* 44, no. 3 (2001): 82-83.

Musk, Elon, and Neuralink. "An Integrated Brain-Machine Interface Platform with Thousands of Channels." *Journal of Medical Internet Research* 21, no. 10 (2019): 1-14.

Musk, Elon. "Neuralink Update." X, July 10, 2024. https://x.com/neuralink/status/1811095113281720722

Neuralink. "Neuralink Progress Update." YouTube August 28, 2020. https://www.youtube.com/live/DVvmgjBL74w?feature=shared.

Neuralink. *Monkey MindPong*. YouTube, April 8, 2021. https://www.youtube.com/watch?v=rsCul1sp4hQ.

Neuralink. "PRIME Study Progress Update." May 8, 2024. https://neuralink.com/blog/prime-study-progress-update/.

Newberg, A., and E. D'Aquili. "The Neuropsychology of Religious and Spiritual Experience." *Journal of Consciousness Studies* 7, no. 11-12 (2000): 251-266.

Newberg, Andrew, and Mark Robert Waldman. *How God Changes Your Brain: Breakthrough Findings from a Leading Neuroscientist*. Random House Publishing Group, 2009.

Newberg, Andrew. *Neurotheology: How Science Can Enlighten Us About Spirituality*. Columbia University Press, 2018.

Nietzsche, Friedrich. *The Twilight of the Idols and the Anti-Christ: or How to Philosophize with a Hammer*. National Geographic Books, 1990.

Nietzsche, Friedrich. *The Gay Science*. Translated by Walter Kaufmann. Vintage Books, 1974.

Peksa, Janis, and Dmytro Mamchur. "State-of-the-Art on Brain-Computer Interface Technology." *Sensors* 23, no. 13 (2023): 1-28.

Perry, Gemma, Vince Polito, Narayan Sankaran, and William Forde Thompson. "How Chanting Relates to Cognitive Function, Altered States and Quality of Life." *Brain Sciences* 12, no. 11 (2022): 1-22.

Peters, Ted. "Imago Dei, DNA, and the Transhuman Way." *Theology and Science* 16, no. 3 (2018): 353-362.

Piccinini, Gualtiero, and Sonya Bahar. "Neural Computation and the Computational Theory of Cognition." *Cognitive Science* 37, no. 3 (2013): 453-488.

Piccinini, Gualtiero. *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford University Press, 2020.

Pisarchik, Alexander N., Vladimir A. Maksimenko, and Alexander E. Hramov. "From Novel Technology to Novel Applications: Comment on 'An Integrated Brain-Machine Interface Platform With Thousands of Channels' by Elon Musk and Neuralink." *Journal of Medical Internet Research* 21, no. 10 (2019): 1-7.

Polger, Thomas W., and Lawrence A. Shapiro. *The Multiple Realization Book*. Oxford University Press, 2016.

Protopapadakis, Evangelos D. "Messing with Autobiographical Memory: Identity, and Moral Status." *International Dialogue East-West* 4 (2022): 175-181.

Rahula, Walpola. *What the Buddha Taught*. Grove Press, 1974.

Rainey, Stephen, Stéphanie Martin, Andy Christen, Pierre Mégevand, and Eric Fourneret. "Brain Recording, Mind-Reading, and Neurotechnology: Ethical Issues from Consumer Devices to Brain-Based Speech Decoding." *Science and Engineering Ethics* 26, no. 4 (2020): 2295-2311.

Rana, Fazale. "A Christian Perspective on Living Electrodes." *Reasons to Believe*, January 13, 2021. https://reasons.org/explore/blogs/the-cells-design/a-christian-perspective-on-living-electrodes.

Roebben, Bert, and Klaus von Stosch. "Religious Education and Comparative Theology: Creating Common Ground for Intercultural Encounters." *Religions* 13, no. 11 (2022): 1-13.

Ryff, Carol D. "Spirituality and Well-Being: Theory, Science, and the Nature Connection." *Religions* 12, no. 11 (2021): 1-19.

Sandel, Michael J. *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Harvard University Press, 2007.

Sapolsky, Robert M. *Behave: The Biology of Humans at Our Best and Worst*. Penguin, 2017.

Savulescu, Julian, and Evangelos D. Protopapadakis. "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu." *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

Schmid, J. R., O. Friedric, S. Kessner, and R. J. Jox. "Thoughts Unlocked by Technology – a Survey in Germany About Brain-Computer Interfaces." *NanoEthics* 15 (2021): 303-313.

Schwartz, Andrew B. "Movement: How the Brain Communicates with the World." *Cell* 164, no. 6 (2016): 1122-1135.

Serim, Barış, Michiel Spapé, and Giulio Jacucci. "Revisiting Embodiment for Brain-Computer Interfaces." *Human-Computer Interaction* 39, nos. 5-6 (2023): 1-27.

Shaima, Mujiba, Nasir Uddin Rana, Estak Ahmed, Mousumi Hasan Mukti, Norun Nabi, Tanvir Islam, Mazharul Islam Tusher, and Quazi Saad-Ul-Mosaher. "Elon Musk's Neuralink Brain Chip: A Review on 'Brain-Reading' Device." *Journal of Computer Science and Technology Studies* 6, no. 1 (2024): 200-203.

Sitaram, Ranganatha, Andrea Sánchez Corzo, Mariana Zurita, Constanza Levican, Daniela Huepe-Artigas, Juan Andrés Mucarquer, and Matías Ramírez. "Brain-Computer Interfaces and Neurofeedback for Enhancing Human Performance." In *Human Performance Optimization: The Science and Ethics of Enhancing Human Capabilities*, edited by Michael D. Matthews and David M. Schnyer, 125-141. Oxford Academic, 2019.

Sonko, Sedat, Adefunke Fabuyide, Kenneth Ifeanyi Ibekwe, Emmanuel Augustine Etukudoh, and Valentine Ikenna Ilojianya. "Neural Interfaces and Human-Computer Interaction: A U.S. Review: Delving into the Developments, Ethical Considerations, and Future Prospects of Brain-Computer Interfaces." *International Journal of Science and Research Archive* 11, no. 1 (2024): 702-717.

Sun, Xiao-yu, and Bin Ye. "The Functional Differentiation of Brain-Computer Interfaces (BCIs) and Its Ethical Implications." *Humanities and Social Sciences Communications* 10 (2023): 1-9.

Swinburne, Richard, and Vasileios Meichanetsidis. "Proofs for the Existence of God: A Discussion with Richard Swinburne." *Conatus – Journal of Philosophy* 9, no. 2 (2024): 305–314.

Tangermann, Victor. "Neuralink Co-Founder Has an Idea for a New Religion." *Futurism*, March 26, 2021. https://futurism.com/neuralink-co-founder-new-religion-drugs-experience-god.

Tedeschi, Richard G., and Lawrence G. Calhoun. "Posttraumatic Growth: Conceptual Foundations and Empirical Evidence." *Psychological Inquiry* 15, no. 1 (2004): 1-18.

Tennison, Michael N. "Moral Transhumanism: The Next Step." *The Journal of Medicine and Philosophy* 37, no. 4 (2012): 405-416.

Tononi, Giulio, and Christof Koch. "Consciousness: Here, There and Everywhere?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, no. 1668 (2015): 20140167.

Turner, Ronald Cole. *Transhumanism and Transcendence: Christian Hope in an Age of Technological Enhancement*. Georgetown University Press, 2011.

Urban, Tim. "Neuralink and the Brain's Magical Future." *Wait But Why*, April 20, 2017. https://waitbutwhy.com/2017/04/neuralink.html.

Venniyoor, Ajit. "Neuralink and Brain-Computer Interface – Exciting Times for Artificial Intelligence." *South Asian Journal of Cancer* 13, no. 1 (2024): 63-65.

Vitale, Flavia, Daniel G. Vercosa, Alexander V. Rodriguez, Sushma Sri Pamulapati, Frederik Seibt, Eric Lewis, J. Stephen Yan, Krishna Badhiwala, Mohammed Adnan, Gianni Royer-Carfagni, Michael Beierlein, Caleb Kemere, Matteo Pasquali, and Jacob T. Robinson. "Fluidic Microactuation of Flexible Electrodes for Neural Recording." *Nano Letters* 18, no. 1 (2017): 326-335.

Vlek, Rutger J., David Steines, Dyana Szibbo, Andrea Kübler, Mary-Jane Schneider, Pim Haselager, and FemkeNijboer. "Ethical Issues in Brain-Computer Interface Research, Development, and Dissemination." *Journal of Neurologic Physical Therapy: JNPT* 36, no. 2 (2012): 94-99.

Waisberg, Ethan, Joshua Ong, and Andrew G. Lee. "Correction: Ethical Considerations of Neuralink and Brain-Computer Interfaces." *Annals of Biomedical Engineering* 52 (2024): 1937-1939.

Wang, Xue-Qin, Hong-Qiang Sun, Jia-Yue Si, Zi-Yan Lin, Xiao-Mei Zhai, and Lin Lu. "Challenges and Suggestions of Ethical Review on Clinical Research Involving Brain-Computer Interfaces." *Chinese Medical Sciences Journal* 39, no. 2 (2024): 131-139.

Wegemer, Christopher. "Brain-Computer Interfaces and Education: The State of Technology and Imperatives for the Future." *International Journal of Learning Technology* 4, no. 2 (2019): 141-161.

Willett, Francis R., Donald T. Avansino, Leigh R. Hochberg, Jaimie M. Henderson, and Krishna V. Shenoy. "High-Performance Brain-to-Text Communication via Handwriting." *Nature* 593 (2021): 249-254.

Yuste, Rafael, Sara Goering, Blaise Agüera y Arcas, Guoqiang Bi, Jose M. Carmena, Adrian Carter, Joseph J. Fins, Phoebe Friesen, Jack Gallant, Jane E. Huggins, Judy Illes, Philipp Kellmeyer, Eran Klein, Adam Marblestone, Christine Mitchell, Erik Parens, Michelle Pham, Alan Rubel, Norihiro Sadato, Laura Specker Sullivan, Mina Teicher, David Wasserman, Anna Wexler, Meredith Whittaker, and Jonathan Wolpaw. "Four Ethical Priorities for Neurotechnologies and AI," *Nature* 551, no. 7679 (2017): 159-163.

Zhang, Tong-Kuo. "Perspective and Boundary Exploration of Privacy Transfer Dilemma in Brain-Computer Interface – Dimension Based on Ethical Matrix." *Philosophies* 9, no. 1 (2024): 1-9.

Zhao, Zhi-Ping, Chuang Nie, Cheng-Teng Jiang, Sheng-Hao Cao, Kai-Xi Tian, Shan Yu, and Jian-Wen Gu. "Modulating Brain Activity with Invasive Brain-Computer Interface: A Narrative Review." *Brain Sciences* 13, no. 1 (2023): 1-14.

Zhu, Bingzhao, Uisub Shin, and Mahsa Shoaran. "Closed-Loop Neural Prostheses with On-Chip Intelligence: A Review and a Low-Latency Machine Learning Model for Brain State Detection." *IEEE Transactions on Biomedical Circuits and Systems* 15, no. 5 (2021): 877-897.

T. Basilaia · *Conatus* 10, no. 1 (2025): 57-71
doi: https://doi.org/10.12681/cjp.38547

Liberalism and Aristotelianism: Reflecting on Alasdair MacIntyre's After Virtue

Tatia Basilaia
Charles University, Czech Republic
E-mail address: tatia.basilaia545@student.cuni.cz
ORCID iD: https://orcid.org/0009-0007-0702-3183

Abstract

Alasdair MacIntyre marks liberalism as a key opponent standing in opposition to an Aristotelian virtue ethics framework, and of the ability of communities to base a way of life around virtues centered upon man's telos and what makes a good human life. This paper will argue that this does not need to be the case by citing how classical liberal political aims of decentralization of power and federalism can promote the efforts of communities attempting to build a culture with a focus on inculcating virtue through the lens of an Aristotelian sense of telos. MacIntyre himself acknowledges the vast differences in definitions of the virtues across cultures throughout history, and how there is unlikely to be any moral consensus. This paper will look at examples from the United States of America's early history, as well as the modern example of the European Union, to illustrate samples of societies inculcating and guarding a traditional worldview within a decentralized political environment. The liberal political aim of decentralization of power provides more autonomy to local communities, including allowing those communities to build their own culture with a focus on forming a society interested in answering the question of what a good human life consists of. This paper will argue that it is precisely the liberal individualism that MacIntyre decries as a foe to Aristotelian teleology that provides an avenue for those interested in restoring Aristotelian virtue ethics to thrive.

Keywords: *virtue; liberalism; Aristotelian; federalism; human good*

I. Introduction

Alasdair MacIntyre is a leading philosopher in the revival of virtue ethics. His book *After Virtue* is a major critique of modern philosophical discussion, particularly that stemming from the Enlightenment. The liberal individualism that emerged out of the Enlightenment is cited as a

constant foe to the Aristotelianism that MacIntyre favors. MacIntyre writes the following in the prologue to *After Virtue*:

> My own critique of liberalism derives from a judgment that the best type of human life, that in which the tradition of the virtues is most adequately embodied, is lived by those engaged in constructing and sustaining forms of community directed towards the shared achievement of those common goods without which the ultimate human good cannot be achieved. Liberal political societies are characteristically committed to denying any place for a determinate conception of the human good in their public discourse, let alone allowing that their common life should be grounded in such a conception. On the dominant liberal view, government is to be neutral as between rival conceptions of the human good, yet in fact what liberalism promotes is a kind of institutional order that is inimical to the construction and sustaining of the types of communal relationship required for the best kind of human life.[1]

However, it does not necessarily need to be the case that liberalism should stand as a nemesis to the realization of a community centered around inculcating Aristotelian conceptions of what a good human life is. According to Bagby, genuine happiness develops gradually via experience, education, friendships, and the development of virtue rather than being something we just happen to find. We must live, struggle, and develop before we can act wisely. Our internal motivation to persevere is what propels that growth. Furthermore, pleasure can foster the growth of virtue by keeping us dedicated to doing the right thing, rather than merely serving as a temptation.[2] MacIntyre overlooks the aspects of political liberalism conducive to building communities interested in reviving an Aristotelian virtue ethics framework. There is an entire political program in the liberal tradition with a focus on decentralization of power and federalism that would create the conditions for smaller-scale political units to emerge with more autonomy, which would be more in line with the idea of the *polis* as envisioned by Aristotle. This paper will argue that contrary to being a foe to Aristotelian teleology, political liberalism should be seen as a potential ally for those who want to build communities with a focus on inculcating Aristotelian conceptions of the good human life in the public square.

---

[1] Alasdair MacIntyre, *After Virtue: A Study in Moral Theory*, 3rd ed. (University of Notre Dame Press, 2007), xiv-xv.

[2] John R. Bagby, "Aristotle and Aristoxenus on Effort," *Conatus – Journal of Philosophy* 6, no. 2 (2021): 68.

II. MacIntyre's critique of liberal political order

A brief discussion of classical Aristotelianism is necessary to understand MacIntyre's critique of liberalism's incompatibility with an Aristotelian worldview. MacIntyre's view is blunt. He writes,

> My own conclusion is very clear. It is that on the one hand we still, in spite of the efforts of three centuries of moral philosophy and one of sociology, lack any coherent rationally defensible statement of a liberal individualist point of view; and that, on the other hand, the Aristotelian tradition can be restated in a way that restores intelligibility and rationality to our moral and social attitudes and commitments.[3]

According to MacIntyre, the Aristotelian worldview is one where it is understood that every practice human beings engage in aspires to some good, or some goal. Human beings have a particular nature, "and that nature is such that they have certain aims and goals, such that they move by nature towards a specific telos."[4] He calls the ultimate end towards which human beings seek *eudaimonia*, or what can perhaps be thought of as happiness or human flourishing. MacIntyre elaborates that "It is the state of being well and doing well in being well, of a man's being well-favored himself and in relation to the divine."[5] MacIntyre describes the virtues as "qualities the possession of which will enable an individual to achieve eudaimonia and the lack of which will frustrate his movement towards that *telos*,"[6] and he notes the surprising lack of attention Aristotle gives to rules in his work on ethics.[7] Aristotle's philosophy of ethics is more focused on the building of character through constant exercise of the virtues, as opposed to devising specific rules for people to follow. Although MacIntyre notes that "Aristotle...recognizes that his account of the virtues has to be supplemented by some account, even if a brief one, of those types of action which are absolutely prohibited."[8]

Liberal individualism on the other hand sidelines Aristotelian notions of human beings having a telos that we can discover through the use of reason, and places the following of rules as the highest moral good as opposed to

[3] Ibid., 259.

[4] Ibid., 148.

[5] Ibid., 148.

[6] Ibid.

[7] Ibid., 150.

[8] Ibid., 152.

any Aristotelian idea of building one's character through active practice of virtue. As MacIntyre writes about modernity,

> Rules become the primary concept of the moral life." Qualities of character then generally come to be prized only because they will lead us to follow the right set of rules...Hence on the modern view the justification of the virtues depends upon some prior justification of rules and principles; and if the latter become radically problematic, as they have, so also must the former.[9]

MacIntyre echoes the thought of legal philosopher Ronald Dworkin that

> the central doctrine of modern liberalism is the thesis that questions about the *good life for man* or the ends of human life are to be regarded from the public standpoint as systematically unsettlable...[And therefore,] The rules of morality and law hence are not to be derived from or justified in terms of some more fundamental conception of the good for man.[10]

MacIntyre has reasons to be skeptical that an Aristotelian focus on man's teleology can be fostered in a liberal institutional arrangement where questions about the good human life are placed to the side. The liberal individualism characteristic of much of the West today is in conflict with Aristotelian notions of man having a telos. MacIntyre acknowledges the challenge in Aristotle's time for a city of tens of thousands of Athenian men to share a common vision of what is good for man.[11] If it was difficult enough for the comparatively smaller city-state of Athens to maintain a shared vision of what is good for man among its people, then there is little hope that the massive cities of today composed of millions of people (who are not even autonomous since they are merely part of a wider nation-state) can hope to come together with a shared vision of how man ought to live. The adage "It takes a village to raise a child" reflects an Aristotelian sensibility of educating youth needing to be a common endeavor, rather than a merely private one. MacIntyre even uses education (along with hospitals and philanthropic organizations) as an example of an area of life that is occasionally viewed as a common project in the way that Aristotle would envision the community

[9] Ibid., 119.

[10] Ibid.

[11] Ibid., 156-157.

as a polis interested in a holistic view of life.[12] The direction of education, including the more all-encompassing educational aspect of raising one's family within an environment of similar values to one's own, will struggle to be formed by any Aristotelian sense of man's telos within a culture of liberal individualism that places questions of man's telos into the arena of subjective opinion. In fact, a liberal individualist mindset is prone to think of any group of people who attempt to form tightly-knit communities separate from others, while actively discouraging outside values opposed to the group's values, as "cultish."[13] MacIntyre himself realizes that given the state of moral discourse where moral concepts mean different things to different people based on one's own subjective opinions, that "It follows that our society cannot hope to achieve moral consensus."[14] However, another significant contributor to such questions of what human good is being placed to the side in favor of a liberalism as described by Ronald Dworkin is that modern states are behemoths by the standards of Aristotle's time. While it might be a valid Aristotelian critique of liberalism to point out how liberalism hampers the ability of a culture to focus on the fundamental issue of what a good human life consists of in favor of leaving that problem to subjective individual opinion, it may also be unfair to expect any other outcome given the sheer size of states today. The Athenian city-state of Aristotle's time is minuscule compared with the size of jurisdictions today, with MacIntyre mentioning that the number of Athenian men[15] numbered somewhere in the tens of thousands.[16] Meanwhile, non-autonomous cities in the West today have populations running into the millions. MacIntyre makes clear his discontent in how the field of ethics moved away from an Aristotelian focus on developing character through

---

[12] Ibid., 156.

[13] Jeff Zeleny, "Prominent Pastor Calls Romney's Church a Cult," *The New York Times,* October 8, 2011, https://www.nytimes.com/2011/10/08/us/politics/prominent-pastor-calls-romneys-church-a-cult.html. Consider the case of religious groups. Members of the Church of Jesus Christ of Latter-Day Saints have traditionally been known for their strong sense of community, including a preference of supporting one another's businesses, education, and other endeavors as opposed to those of non-members. They are also encouraged to avoid material inconsistent with their church's values. However, this sense of community has led to accusations of them being a cult. This debate entered the American political sphere prominently when Mitt Romney, a Mormon, won the Republican Party's nomination for president in 2012. Any group of people who try to inculcate a particular moral outlook into their community while trying to keep out opposing moral outlooks hostile to one's own values is going to be at odds with a liberal individualist framework.

[14] MacIntyre, 252.

[15] Ibid., 159. MacIntyre also acknowledges Aristotle's blind spot when it comes to his assessment that groups such as non-Greeks and slaves are incapable of political relationships. His blind spot towards the role of women must be acknowledged too.

[16] Ibid., 156.

discipline and practice of the virtues, and towards the Enlightenment focus on following universal rules, such as in a Kantian mold. But such a move is unavoidable given the scale of modern nation-states. MacIntyre even admits that "different and rival lists of virtues, different and rival attitudes toward the virtues and different and rival definitions of individual virtues are at home in fifth-century Athens"[17] even though he ultimately thinks that "nonetheless the city-state and the *agōn* (ἀγών) [or contest] provide the shared contexts in which the virtues are to be exercised."[18] And consider also that despite the diversity of lived experiences endured by the various ancient Athenians, this diversity of lived experience has only increased with populations continually growing for centuries and economies becoming far more complex. It should be no surprise that a growing population in a rapidly changing economy will lead to the creation of different groups of people with dissimilar values, interests, and goals. MacIntyre describes the Aristotelian notion of friendship as requiring "a shared recognition of and pursuit of a good,"[19] and that "We are to think then of friendship as being the sharing of all in the common project of creating and sustaining the life of the city."[20] But such a notion of friendship is not sustainable in ever-growing communities as a practical matter. Given the size of political states today, the most efficient way to keep such disparate people together is to place questions about human goods on the side and instead adopt basic rules for everyone to follow. Attempts in a large nation-state (or even just a large city today) at trying to nurture a particular moral outlook based on an Aristotelian sense of telos are going to lead to discontent among those who do not share that vision. Likewise, those in support of traditional ancient and medieval virtue might be at risk of having a moral outlook imposed on them that they do not support. MacIntyre runs into the problem that the current political entities are far too big for any notion of a shared, common project in an Aristotelian sense to thrive. Attempts at trying to inculcate a shared moral vision are going to alienate at least one group of people, and likely many more.

III. Liberalism as a political program

The way forward for a culture to develop with an openness to Aristotelian ideas of humans having a telos, and to reexamining the assumptions of liberal individualism is, ironically, to embrace political liberalism in the form of political decentralization and federalism. Perhaps liberalism itself is a loaded

[17] Ibid., 138.

[18] Ibid.

[19] Ibid., 155.

[20] Ibid., 156.

term, and it would be helpful to think of liberalism in a couple different ways. What MacIntyre is objecting to is only one aspect of philosophical liberalism. This would be a form of liberal individualism with roots in the Enlightenment that discards notions of teleology, and instead assigns moral agency to the individual, which eventually dilutes morality to a meaningless subjective opinion. MacIntyre even applauds Nietzsche's dismantling of any notion of objective morality developed by Enlightenment philosophers. MacIntyre writes the following in praise of his intellectual foe:

> In a famous passage in *The Gay Science* (section 335) Nietzsche jeers at the notion of basing morality on inner moral sentiments, on conscience, on the one hand, or on the Kantian categorical imperative, on universalizability, on the other. In five swift, witty and cogent paragraphs he disposes of both what I have called the Enlightenment project to discover rational foundations for an objective morality and of the confidence of the everyday moral agent in post-Enlightenment culture that his moral practice and utterance are in good order.[21]

MacIntyre is in stark disagreement with Friedrich Nietzsche, but is of the mindset that Nietzsche is a far more logical alternative to Aristotelianism than anything produced out of the Enlightenment. He even calls Nietzsche's moral philosophy "one of the two genuine theoretical alternatives confronting anyone trying to analyze the moral condition of our culture,"[22] with the other alternative of course being Aristotelianism. MacIntyre considers liberalism to be an inconsistent and muddled moral philosophy, as well as inferior to the Aristotelianism that it dethroned. MacIntyre writes:

> I take it then that both the utilitarianism of the middle and late nineteenth century and the analytical moral philosophy of the middle and late twentieth century are alike unsuccessful attempts to rescue the autonomous moral agent from the predicament in which the failure of the Enlightenment project of providing him with a secular, rational justification for his moral allegiances had left him. I have already characterized that predicament as one in which the price paid for liberation from what appeared to be the external authority of traditional morality was the loss of any authoritative content from the would-be moral utterances of the newly autonomous agent. Each moral agent now

[21] Ibid., 113.

[22] Ibid., 110.

> spoke unconstrained by the externalities of divine law, natural teleology or hierarchical authority; but why should anyone else now listen to him?[23]

As MacIntyre summarizes in his conclusion,

> ...ever since belief in Aristotelian teleology was discredited moral philosophers have attempted to provide some alternative rational secular account of the nature and status of morality, but...all these attempts, various and variously impressive as they have been, have in fact failed, a failure perceived most clearly by Nietzsche.[24]

MacIntyre makes clear that individualism with its "modern liberal distinction between law and morality"[25] is antithetical to the Aristotelian notion of a shared moral vision among a community. He writes,

> There is of course a crucial difference between the way in which the relationship between moral character and political community is envisaged from the standpoint of liberal individualist modernity and the way in which that relationship was envisaged from the standpoint of the type of ancient and medieval tradition of the virtues which I have sketched. For liberal individualism a community is simply an arena in which individuals each pursue their own self-chosen conception of the good life, and political institutions exist to provide that degree of order which makes such self-determined activity possible. Government and law are, or ought to be, neutral between rival conceptions of the good life for man, and hence, although it is the task of government to promote law-abidingness, it is on the liberal view no part of the legitimate function of government to inculcate any one moral outlook.[26]

This is certainly a revolution away from an Aristotelianism focused on virtue and man's ultimate good that MacIntyre describes in his book. MacIntyre is correct to be wary of this style of liberal individualism that ignores the

[23] Ibid., 68.

[24] Ibid., 256.

[25] Ibid., 172.

[26] Ibid., 195.

fundamental question of what makes a good man and puts little emphasis on the development of virtue and character. However, this does not mean that liberalism is entirely at odds with Aristotelianism. Alasdair MacIntyre argues that liberalism's focus on individualism undermines the communal pursuit of virtue, which is central to an Aristotelian vision of the good life. In *After Virtue*, MacIntyre critiques liberalism for its inability to sustain a shared moral framework, asserting that it fragments society into competing moral claims without a common telos.[27]

However, Aristotle's own political theory, as presented in *Politics*, offers a more nuanced view. Aristotle recognizes the importance of local communities, or *polis*, in cultivating virtue, but he does not impose strict limits on the size of political entities. Unlike Plato's rigid and idealized state model, Aristotle acknowledges that larger political structures, such as empires, can function effectively if they operate through subsidiarity granting local units' autonomy to address their unique needs.[28] This insight challenges MacIntyre's skepticism about larger liberal political frameworks, suggesting that liberal federalism could in principle, support the cultivation of Aristotelian virtues at the community level. Despite MacIntyre's belief that the history of political and moral action cannot be separated from the history of political and moral theorizing,[29] there are elements of liberal political action compatible with strengthening communities interested in pursuing questions of human good in the public square from an Aristotelian perspective. The reality of politics is messier than the world of pure theory, meaning that the philosophy of liberal individualism and the political program of liberalism are not necessarily the same thing. In fact, political liberalism may be used to push for illiberal aims when tools such as decentralization of political power through federalism are employed to specific ends. For example, the early American republic is often thought of as being engaged in a program of political liberalism, and that is true to an extent. However, part of the political program of liberalism in the American context was the idea of states' rights and federalism, which were often employed to protect the traditional features of life for each of the various states that shaped the United States of America. Consider liberalism's political history as it has been advanced in the United States. One major concern at the American Constitutional Convention was that the new general government was going to eventually supplant the authority of the state governments that initially formed the United States government with the constitution. The various state governments all developed their own unique cultures from their colonial days that citizens were interested in protecting,

[27] Ibid., xiv-xv.

[28] Aristotle, *Politics*, trans. C. D. C. Reeve (Hackett Publishing Company, 1998), 1252a1-10.

[29] MacIntyre, 61.

including a number of states maintaining official state churches supported with taxpayer money.[30] None of the original thirteen British colonies in North America wanted to be politically dominated by the other states, particularly those with whom they shared the most disagreement with. The Bill of Rights was added to the constitution to quell the fears of Anti-Federalists[31] that the general government would overtake the states and begin regulating their internal affairs. As Akhil Reed Amar of Yale Law School observes, speech and religion were put together in the original First Amendment largely for reasons of federalism"[32] and "Congress was prohibited not only from establishing a national church, but also from disestablishing a state church."[33] Thomas Jefferson even wrote in an 1804 letter that "While we deny that [the United States of America] Congress has a right to control the freedom of the press, we have ever asserted the right of the States, and their exclusive right to do so."[34] In these cases, the liberal political tactics of federalism and decentralized political power could be used for illiberal aims, such as allowing local communities to make autonomous political decisions in the name of protecting their own set of values separate from those of other cultures. There are plenty of forms of political liberalism that are not conducive to Aristotelianism as well. If we look at the liberalism of the French Revolution, we see a movement interested in destroying French tradition. But on the other hand, the political liberalism of the American Revolution helped to preserve the traditional system of British Common Law that the representatives of

---

[30] David Hackett Fischer, *Albion's Seed: Four British Folkways in America* (Oxford University Press, 1989). Consider that three New England states – Massachusetts, New Hampshire, and Connecticut – each had official state churches when the First Amendment was ratified, and Massachusetts maintained an official state church all the way until 1833. The colonies did not always have amicable relations with one another either. David Hackett Fischer's *Albion's Seed* is a helpful book for understanding the various waves of immigration from Great Britain to North America in the colonial period, and how the colonial period led up to and informed the development of the United States of America's early years as a republic.

[31] "Anti-Federalists" was the name attached to those who were more disposed to support a decentralized government with more power in the hands of the state governments, and who also opposed the centralizing tendencies of the new American constitution. Ironically, it is the faction labeled Anti-Federalists who were advocating for federalism and decentralized political power in the new republic. Hence, another example of politics as a practice not necessarily meshing perfectly with politics as elaborated in theory.

[32] Akhil Reed Amar, "Anti-Federalists, *The Federalist Papers*, and the Big Argument for Union," *Harvard Journal of Law and Public Policy* 16, no. 1 (1993): 115. https://openyls.law.yale.edu/bitstream/handle/20.500.13051/233/Anti_Federalists__The_Federalist_Papers__and_the_Big_Argument_for_Union.pdf.

[33] Ibid., 116.

[34] Thomas Jefferson, "From Thomas Jefferson to Abigail Adams," *Founders Online*, National Archives, September 11, 1804, https://founders.archives.gov/documents/Jefferson/01-44-02-0341.

the original thirteen colonies believed they had developed through their tradition as English subjects.[35] However, both events are referred to as "revolutions" even though they were each fought with different motivations in mind. Likewise, the political program of liberalism contains a wide array of perspectives and strategies that can be used either for or against the kind of society MacIntyre desires. Liberalism is a political tactic just as much as it is a theory. It is an oversimplification to label liberalism as an unambiguous rival to Aristotelianism, and the topic must be covered with more nuance.

Or perhaps we can consider the modern-day case of the European Union. There is considerable difference in opinion within the political class of the European Union's leaders today on issues such as immigration. Political decentralization allows individual member states to enact different policies in response to immigration. Some countries such as Germany will be more open to refugee immigration, while others like Hungary will be less open to refugee immigration. But both sides are making their own autonomous decisions within the decentralized political format of the European Union. This political tactic of liberalism can be employed in ways that appeal to either the political right or the political left. Individuals in countries such as Hungary even use rhetoric of protecting their identity and sovereignty when operating within the decentralized political environment of the European Union, such as Viktor Orbč n referring to the current decade of politics as being about Hungary maintaining its sovereignty, and claiming that "Hungary remaining a sovereign country is not in the interest of the world around us, and neither is it in the interest of that world's people inside Hungary."[36] Advocates for a society

[35] Consider these two works from Edmund Burke discussing the French Revolution and American Revolution respectively. Figures like Edmund Burke felt no contradiction in their commiseration for the pleas of the American Revolution while expressing disdain for the French Revolution. Burke saw the American Revolution as a mere defense of traditional English law in the American Revolution, but saw the French Revolution as a violent destruction of tradition. Liberalism as a political program must be evaluated on a case-by-case basis to determine if it is useful for advancing Aristotelian virtue ethics. The early American republic is an example of a cause where the political program of liberalism, in the form of secession, was able to defend a society's traditional way of life from being interfered with by a stronger power. The political program of liberalism is not always at odds with community tradition, and Aristotelians should take notice of historical examples of communities protecting particular values, especially when those examples lead to outcomes as extreme as war. There should be no naivete about the potential resistance that someone advocating for a culture based on Aristotelian values could face if those values are seen to be in conflict with the wider culture. See Edmund Burke, *Reflections on the French Revolution & Other Essays* (J. M. Dent & Sons, 1951), https://archive.org/details/reflectionsonthe005907mbp/page/n5/mode/2up, and Edmund Burke, *Burke's Speech on Conciliation with the Colonies*, originally delivered March 22, 1775 (Leach, Shewell, and Sanborn, 1895), https://archive.org/details/burkespeechon00burkrich/page/42/mode/2up?view=theater&q=right.

[36] Viktor Orbč n, "Speech by Prime Minister Viktor Orbč n at the Szč zadvég Sovereignty Conference," Cabinet Office of the Prime Minister, November 13, 2023, https://miniszterelnok.hu/en/speech-by-prime-minister-viktor-orban-at-the-szazadveg-sovereignty-conference/.

based on Aristotelian grounds can take note of movements in countries like Hungary that make appeals to establish their own communities in defiance of a worldview they disapprove of. Namely, using the tools of liberalism's political program to advance a decentralized political environment allowing smaller communities to develop their own understanding of how society ought to be run. Large-scale political entities are often left with little choice but to adopt a rules-based system that puts the question concerning what a good human life is to the side, in favor of instead being a utilitarian arrangement. The European Union itself contains hundreds of millions of people from varying backgrounds, and there is a low likelihood of agreement on several issues. MacIntyre also acknowledges the vast differences across cultures in how to define virtue and what specific attributes should be considered virtues, as well as admitting that there is unlikely to be any moral consensus. He even finds common ground with Karl Marx by stating that "Marx was fundamentally right in seeing conflict and not consensus at the heart of modern social structure."[37] MacIntyre shares a view that "...modern politics cannot be a matter of genuine moral consensus...[and] Modern politics is civil war carried on by other means."[38] But what can be added to this view is that the size of modern political entities is a contributor to this experience of politics as a low-intensity civil war. Smaller-scale political units are not subject to the same challenge of rallying its people to a particular worldview, and it is far easier to form a consensus about what a good human life is when political entities are smaller. Smaller jurisdictions give local populations more say in their own local political spheres, and perhaps someone like MacIntyre could find benefit in a program of political decentralization advanced by liberalism. It is far easier to inculcate a particular worldview within a small community than a large nation-state or international union of states.

IV. Subsidiarity and universal governance

The notion of subsidiarity, which seeks self-governance or devolution rooted in liberal political thought, comes out clearly in Aristotle's features of the state and its structure. For Aristotle, every political society aims to enable various individuals and social units to exist happily, which calls for an active role of the populace in managing the affairs of the state. Although the polis is basic in Aristotle's conception of the cultivation of Virtue, he does not also lose sight of the significance of other larger political entities such as empires, which one may think he will undergird because they contain populations who

[37] MacIntyre, 253.

[38] Ibid.

are sub-sourced to the menial work of administration.[39] Many empires, such as that of Alexander the Great, which ruled over many people and places, noted the need to respect local self-rule.

Madison and other US Federalists and their contemporaries believed in retaining local sovereignty and the need for some centralized rule for effective governance, the same reason argued in this American conception of subsidiarity. It was held that various localities would remain intact with their diverse practices and beliefs as a single nation or state. Although Alasdair MacIntyre shifts the focus from the disagreement between the Federalists and the Anti-Federalists, he might hold British and other economic thoughts common among the Heineman's anti-federalist perspective.[40]

These issues reach out to mere nation-states. In supranational organizations, like the EU, similar problems are faced, where the liberal ideal of self-determination and multiplicity of views faces reality. Kant's viewpoint on these problems can be traced in Perpetual Peace. A visible trend in Kant's argument is the emphasis on a federation of states where the members subscribe to and uphold certain values and standards to sustain peace and reduce instances of war.[41] While this argument is indeed reflecting liberal ideas, it seems to contradict sharply what MacIntyre considers to be the dominant focus of emphasis, namely the primacy of specific traditions and the role of social order in the development of good character. In MacIntyre's view, it is quite likely that, by embracing Kant's global approach, there will be a loss of culture and history, which is necessary for the attainment of Virtue.

In conclusion, Aristotle's idea of subsidiarity and power distribution concerning federalism explains how liberalism can be reconciled philosophically with Aristotelian concepts if implemented correctly. Creating systems that respect certain local cultures and promote collective aims helps preserve the values MacIntyre himself would even argue liberal governance allows one to do while operating in more relevant settings of modern-day politics.[42]

V. Conclusion

Alasdair MacIntyre views liberalism as an adversary to the building of a society based on Aristotelian notions of human good and flourishing, but this does not mean liberalism must always be a foe to his preferred philosophy.

[39] Aristotle, 1253a20-25.

[40] James Madison, *The Federalist Papers*, ed. Clinton Rossiter (Penguin Classics, 2003), 45.

[41] Immanuel Kant, *Perpetual Peace: A Philosophical Sketch*, trans. H. B. Nisbet (Cambridge University Press, 1991), 41-47.

[42] Alasdair MacIntyre has not directly addressed the Federalist/Anti-Federalist debate. His writings on European unification give little sense of his position on this matter.

A more nuanced perspective is in order. The political program of liberalism can be used by a wide variety of cultures to suit specific needs. In an age when Aristotelianism is not a dominant popular or academic viewpoint, perhaps those interested in Aristotelianism should consider the benefits of traditionally liberal political initiatives, such as decentralization of political power and federalism, to advance one's own perspective. Aristotelians are unlikely to dominate the cultural mainstream any time soon, and most people are never going to hold political offices like governor, mayor, or sheriff where he or she can use one's authority to resist political initiatives hostile to the development of Aristotelian sensibilities.

According to Donev and Skalovski, the breakdown of common ethical traditions is the cause of the moral disorientation that characterizes modern liberal societies. Based on the philosophical systems of Aristotle and Alasdair MacIntyre, they suggest that virtue ethics, with its emphasis on moral character, social ties, and the development of a meaningful human life, offers a workable basis for restoring harmony and significance in a world that is ethically disjointed.[43]

However, a normal person can still work in their own local community to advocate for building a culture focused on inculcating a moral outlook in step with Aristotelianism. They can use what influence they have at their disposal to begin building the kind of culture they want, creating an attachment among one's local community to a specific place with various initiatives to form a sense of home, and supporting local political initiatives to protect one's community from values he or she thinks are harmful. Measures that can be taken at a local level to build the kind of community one wants are numerous. There are local school boards who take an active role in the education system of a local community, town council positions, and numerous ways to volunteer locally. As Aristotle recognized thousands of years ago, and MacIntyre knows today, character and virtue must be developed through active practice and participation within society. That means local purposeful action is the most readily available option for constructing a culture focused on man's telos.

Acknowledgement

This review of the Contemporary Monograph article is dedicated to the Cooperatio PHIL at Charles University, Prague, Czech Republic, for their invaluable support.

[43] Dejan Donev and Denko Skalovski, "Responsibility in the Time of Crisis," Conatus – Journal of Philosophy 8, no. 1 (2023): 101.

References

Amar, Akhil Reed. "Anti-Federalists, The Federalist Papers, and the Big Argument for Union." *Harvard Journal of Law and Public Policy* 16, no. 1 (1993): 111-118. https://openyls.law.yale.edu/bitstream/handle/20.500.13051/233/Anti_Federalists__The_Federalist_Papers__and_the_Big_Argument_for_Union.pdf.

Aristotle. *Politics*. Translated by C. D. C. Reeve. Hackett Publishing Company, 1998.

Bagby, John R. "Aristotle and Aristoxenus on Effort." Conatus – Journal of Philosophy 6, no. 2 (2021): 51-74.

Burke, Edmund. *Burke's Speech on Conciliation with the Colonies*. Originally delivered March 22, 1775. Leach, Shewell, and Sanborn, 1895. https://archive.org/details/burkespeechon00burkrich/page/42/mode/2up?view=theater&q=right.

Burke, Edmund. *Reflections on the French Revolution & Other Essays*. J. M. Dent & Sons, 1951. https://archive.org/details/reflectionsonthe005907mbp/page/n5/mode/2up.

Donev, Dejan, and Denko Skalovski. "Responsibility in the Time of Crisis." Conatus – Journal of Philosophy 8, no. 1 (2023): 87-109.

Fischer, David Hackett. *Albion's Seed: Four British Folkways in America*. Oxford University Press, 1989.

Jefferson, Thomas. "From Thomas Jefferson to Abigail Adams." *Founders Online*, National Archives, September 11, 1804. https://founders.archives.gov/documents/Jefferson/01-44-02-0341.

Kant, Immanuel. *Perpetual Peace: A Philosophical Sketch*. Translated by H. B. Nisbet. Cambridge University Press, 1991.

MacIntyre, Alasdair. *After Virtue: A Study in Moral Theory*, 3rd edition. University of Notre Dame Press, 2007.

Madison, James. *The Federalist Papers*. Edited by Clinton Rossiter. Penguin Classics, 2003.

Orbán, Viktor. "Speech by Prime Minister Viktor Orbán at the Századvég Sovereignty Conference." Cabinet Office of the Prime Minister, November 13, 2023. https://miniszterelnok.hu/en/speech-by-prime-minister-viktor-orban-at-the-szazadveg-sovereignty-conference/.

Zeleny, Jeff. "Prominent Pastor Calls Romney's Church a Cult." *The New York Times*, October 8, 2011. https://www.nytimes.com/2011/10/08/us/politics/prominent-pastor-calls-romneys-church-a-cult.html.

P. A. e Castro · *Conatus* 10, no. 1 (2025): 73-85
doi: https://doi.org/10.12681/cjp.37087

The Forgotten View of the Origin of Language: The Legacy of Herder's Philosophy

Paulo Alexandre e Castro
Universidade de Coimbra, Portugal
E-mail address: paecastro@gmail.com
ORCID iD: https://orcid.org/0000-0001-8256-1343

Abstract

The question about the origin of language marked modernity with approaches that still echo in contemporary thinkers. This is the case with Herder's Treatise on the Origin of Language. The question with which Herder opens the essay is significant and expresses well the fundamental problem that marked the philosophical intentions of the 18th century, namely: "Were human beings, left to their natural abilities, able to invent language for themselves?" Forgetting for a moment the implicit reference to God that continued to mark historically the philosophical narratives, it is important to focus the question on the appearance of language. In this sense, the philosopher's essay is not limited to hypotheses about the emergence of language, or rather, about the founding characteristics of language, but it consolidates it in the anthropological, sociological and even biological horizon from which it allows the understanding of human nature and condition. The enunciation of the four natural laws and the narrative of justification that the philosopher elaborates on reveals a strong potential to understand the phenomenon of language and the human mind. This essay seeks, first, to explain Herder's theses; second, based on this explanation, see in what sense his approach allows us to understand the phenomenon of the origin and formation of language; and finally, to understand the scope of his work with regard to language and mind, that is, to seek to determine its legacy in contemporary philosophy, namely with regard to the understanding of the human mind. Regarding the latter, it is important to mention two essential points to understand the importance of Herder's thought for the understanding of the human mind: the mention of reflection as an inner thought and hearing as a fundamental characteristic for the development of language. It is all these questions that this essay seeks to address.

Keywords: *language; origin of language; Herder; imitation; mind; sociobiology*

I. Introduction on Herder's thought

One should start by saying that the philosopher's book is not limited to proposing hypotheses about the emergence of language, or rather, about the foundational characteristics of language; it provides a new approach and understanding of human nature and the human condition. The statement that the essay represents a milestone in the establishment of the philosophy of language seems, therefore, undeniable – one can risk saying that it is effectively an ontological-metaphysical approach to the contents that will come to categorize the philosophy of language as a specific area that would define much of 20th century philosophy[1] – and in this sense, an instrument of ontological formation of the human world itself, conforming from the epistemological, aesthetic and anthropological point of view, the positioning of the man who knows and knows himself in the historicity of his creative, emotional and educational experiences.

The notion of a language that makes the world happen is already in genesis here, not in the classical sense of naming and the relationship with the truth value or with the existence of the named (of which Plato's well-known *Cratylus* and *Sophist* are good examples), but in the sense of the intimate configuration of human action with the *dictum* that dictates it. Note the following words of Herder which already suggest this reconfiguration of meaning from the "philosophy of languages" to a philosophy of language, a "first philosophy" that becomes an object of reflection for the construction of man's own philosophical thought:

> And as this important theme promises so much insight into Psychology and natural economy of the human race, into the philosophy of language and of all the sciences which have arisen with it; who would not wish to make the attempt? Besides, as men are the only creatures, with whom we are acquainted, endowed with speech, and are thus distinguished from all other animals, where can we find a more certain course for inquiry than is afforded by observation of the difference between man and brute? Condillac and Rousseau must necessarily err, as to the origin of language, because they erred so decidedly and oppositely, upon this

[1] José Justo says in the introduction of the Portuguese edition: "In other words, the philosophy of language begins to occupy the strategic place of a First Philosophy." See Johann Goottfried Herder, *Ensaio sobre a Origem da Linguagem*, introduction and trans. José M. Justo (Antígona Editores, 1987), 13.

> very distinction, the former viewing animals as men, the latter, men as animals.[2]

In another way, the phenomenological thought of a man who speaks in the world is in genesis here and, let us say, an anticipatory vision of language as the house of being, which would come to mark some of the philosophical discourses of the 20th century (namely of Heidegger), as it places man as the only entity capable of a language that challenges the discourse of being and that, therefore, allows the interpellation of the self and the reasoning that thinks it.

The relationship that Herder establishes between reason and language is not only evident but also necessary for the coherence of the discourse in his essay, that is, for the justification of the human origin of language. The constitution of this language is done through the interconnection of the elements that constitute the subjective experience, particularly in the reception carried out by the senses of mundane impressions, the reflection-reasoning operation and the consciousness of (being) in a social world, which, as the philosopher says, "nature has neither created us isolated stony rocks, nor egotistical monads!"[3]

In this approach to Herder's essay (although flying over many of his words), we cannot fail to notice that there is still a philosophical work to be done regarding his thought. Two examples can be immediately provided in this regard, one, the lack of a rigorous linguistic reading of the philosopher's considerations about signs, and second, the lack of a phenomenology of hearing (or at least a phenomenological interpretation) that would certainly integrate Herder as a main character. In other words, it is necessary to carry out a reading that would integrate the elements of a phenomenology of the inaugural event of human language – where the hearing and the occurrence of the vibratory phenomena of sounds took place as part of the way of being of the being that is-to-be-in-the-world – which would prepare and expand the understanding of language and of human nature itself. This does not mean that we want to make Herder's philosophy (or this book in particular) a phenomenology of the subject or subjectivity, but to alert to the existence of these elements in a philosophy of the (origin) of language. Such elements are not merely decorative elements in this narrative but rather an exaltation of the subject endowed with reason, sensitivity, imagination (one cannot ignore the speeches/works

---

[2] Johann Gottfried Herder, *Treatise upon the Origin of Language* (Camberwell Press, 1822), 15.

[3] Ibid., 1.

of his contemporaries that populated the philosophical culture of the time). It can, therefore, be said that Herder manages to introduce into the heart of the problematic of language, the conception of a being that is exactly the way it is (in fidelity to the spirit of the philosopher's letter, the human animal), that is, a being endowed with reason and sensitivity, who develops himself in the middle of a wild nature, manages not only to learn through his intelligence (also empirically) but also to share what he has learned.

II. Language, natural laws and the thinking of human nature according to Herder

Herder's essay is divided into two parts, the first part consisting of three chapters and the second, a shorter one, consisting of a single text (eventually, a division can be seen through the natural laws that are enumerated). In the first part, the first chapter has the function of exhorting the thought about the human origin of language[4] (mainly criticizing the view of divine creation of language proposed by Süßmilch, which Condillac and Rousseau cannot escape) following the presentation of human frailty (their nature) in contrast to the nature of animals.[5]

The second chapter, perhaps the most important for the purpose of the essay, presents a set of considerations about man as a being who, lacking animal aptitudes (such as innate abilities and instinct), is endowed with intelligence and sensitivity (taken as natural dispositions).[6]

---

[4] For example: "He commented, like his predecessor, with the outcry of nature, from which human language arose. I cannot perceive how it should ever have thus arisen, and am astonished, that the penetration of a Rousseau, should have allowed him to dwell here for a moment [...]. Diodorus and Vitruvius also, who rather believed than proved the human origin of language, increased the difficulty of the subject, by supposing men for a time to have ranged about the woods, howling as animals and afterwards, God knows why, or wherefore, to have discovered language." Ibid., 14-15.

[5] "A new born infant, with the exception of the outcry of his sensitive machine, is dumb, it utters neither ideas, nor impulses by tones, as every animal does, in its peculiar manner; placed, therefore, only among animals, it would be the most orphan child of nature, naked and exposed, weak and necessitous, timid and unarmed and, what constitutes the sum of misery, devoid of any guide through life. With such divided, enfeebled, sensitive power, such indistinct, dormant faculties, such separated and weakened impulses, evidently appointed for a thousand necessities, destined to a capacious sphere, and yet so helpless and abandoned as not even to be endowed with a language to declare his wants! No, such a contradiction does not exist in the economy of nature. Instead of instinct, there must certainly be hidden faculties dormant within him!" Ibid., 19.

[6] "However, this disposition of his powers may be termed, whether understanding, reason, or reflection etc. If these names be not considered as abstract powers, or merely degrees of elevation of the animal powers, it is the same to me. It is the total arrangement of all human powers,

Here Herder introduces the fundamental concept of "reflection" that will help him to reconcile objectively three orientations: first, to present human reason and determination for thought (for example, not being a being who only knows but who knows that he knows),[7] second, to give the possibility of developing the argument in favor of a being that can reason only because it has language and vice versa, and third, the placement of the concept of "conscious reflection" (a "state") that allows to unveil the essentiality of human nature:

> It follows, therefore, from these rules of combination, that all the words, sensitive power and instinct, phantasy and reason, are merely different determinations of one and the same power, which resolves all opposition to unity, and consequently that – if man was not intended to be an instinctive animal, he must by means of the free, active, positive power of his soul, be a thinking being [...]. Reason not being a distinct power, acting separately, but a peculiar direction of all the powers of the human species, man must possess it, in the first state in which he is man.[8]

This conception of Herder's is fundamental for the global understanding of his thought, and specifically, for the understanding of the problem to which he dedicated himself in this essay. In fact, the philosopher is not only emphasizing human rationality, which, moreover, was already a mark of thought at the time, but demanding a different tone for that rationality by placing it within the scope of an essentiality that is interior to the human mind; This means that "reflection" as a characteristic of man allows him to validate an ontological primacy in human

---

the total economy of man's sentient and intellectual, of his intellectual and volitive nature, or rather, it is the positive power of reflection, which, connected with a certain organization of body, is called reason in man, and the instinctive faculty in animals, which in the former, gives birth to freedom, and in the latter, to instinct." Ibid., 22.

[7] "More correctly speaking, the rationality of man, the characteristic of his species, is completely different, *viz* 'it is the superior tendency of his thinking powers, proportioned to his sentient faculties and impulses' [...]. If the sentient animal state, and concentration upon one point ceased, a different creature must appear, whose positive power must necessarily manifest itself in greater space, as arising from a finer organization, i.e. 'more clearly' and which, distinct and free, not only understands, wills, and acts, but is also conscious of understanding, willing, and acting. This creature is man; and this disposition of his nature, we shall, to avoid the confusion of specific rational powers, term conscious reflection." Ibid., 23-24. Note: it must be taken into account that Herder uses many language subterfuges, and this allows him to redo his speech.

[8] Ibid., 24.

nature, that is, reflection only plagues reasoning, thinking, because it is already constitutively original in the human soul, if it were not, it could not appear because as Herder says:

> If there be nothing in this faculty, by what means could it enter the soul? If in the first state, there be nothing positive of reason in the soul, how could any thing, even in millions of succeeding gradations, ever become realized? It is verbal sophistry to say, that use could convert mere possibility into reality; for if power exist not, it can neither be applied, nor exercised [...]. But the sentient state of man was nevertheless human, therefore, reflection took place, though in a minor degree; and the least sentient state of the brute, is still brutish. The greatest clearness ever attained in its thoughts, therefore, could never produce the reflection of a human idea.[9]

The philosopher thus inserts in the scope of the analysis of human nature a founding form of intellection that will be conducive to language. Reflection (see also the primary meaning of this internal awareness as what it reflects in its interiority), is this capacity for internalization that operates a recognition (of differentiating characteristics, also read with a symbolic function) and that, giving a conscious dimension of himself and of the act, will favor the appearance of language. Thus, human language goes beyond a mere identification of speaking and what is spoken, that is, it is present in the very way in which the soul has consciously inscribed in itself the recognition of the reality of what is inside and outside it.[10] Herder refers:

> Man being placed in the state of reflection peculiar to him, the first time this reflection acted freely, language was dis-

[9] Ibid., 25-26.

[10] For this reason, Herder says that "If it be incomprehensible to others, how a human soul could invent language, it is incomprehensible to me, how a human soul could be what it is, without discovering language for itself. Nothing more clearly develops this origin, than the objections of its opponents." Ibid., 30. And further on, he consolidates this idea by saying: "Thus then language is the declaration, expression or organ of the understanding, and forms, as it were, an artificial sense in the human soul; in the same manner as the sensitive soul among the ancients formed for itself the eye, and the instinct of the bee constructed its cell. How excellent that this new, this artificial sense of the mind, even at its origin, was a connecting medium and must be so! I cannot imagine the first human thought, nor the first formed judgment, without a kind of dialogue within my soul, or an endeavor to carry out one; the first human thought, therefore, from its nature, is a preparation for social intercourse." Ibid., 38.

> covered. For what is reflection? What is language? This reflection is characteristically peculiar to man, and essential to his species, so is also language and his own discovery of language. Discovery of language is, therefore, natural to him as man! [...] He shows reflection, therefore, not only by clearly and acutely observing all the properties, but by acknowledging one or several as distinguishing properties. The first act of this acknowledgment, produces a clear conception, it is the first judgment of the soul. By what means did this acknowledgment take place? By means of a sign, which he must have fixed, and which as a sign of reflection remained clear within him. Well then, let us hail him with *εὕρηκα*! (it is found!) [...]. With this human language was discovered.[11]

It is, therefore, from the understanding of this state of reflection and not from the explanation of the imitative character present in nature,[12] which must lie the logical and natural explanation for the human origin of language. This is a sensitive point in Herder's speech that, as it still happens today, raises many questions. And it raises them precisely because the idea of imitation seems to appear as something reductive or as something that refers to an instinctive competence in the human, which according to Herder does not seem acceptable.[13] One of the great defenders of the mimetic theory is Susan Blackmore, who says that imitation is precisely what makes us human, or rather, we would be "differentiated imitators,"[14] meme beings (one feels the influence of

---

[11] Ibid., 27.

[12] "Another principle has been adopted, *viz*: the imitation of nature and her sounds, as if anything like thought could ever be produced from such a blind impulse! As if the ape, even with such an inclination, or the black bird, which can so well imitate sounds, could ever have contrived a language [...]. It is not the mere utterance of feeling; for it was not a breathing machine, but a reflecting creature which invented language. It was not a principle of imitation in the soul, which is only a means to attain one single object; least of all is it agreement, or an arbitrary convention of society. The savage, the hermit of the woods, would have discovered a language for himself, even had he never uttered it. It was the intelligence of the soul with itself, an intelligence necessary to man, as man." Ibid., 29-30.

[13] Daniel Everett in his most recent book seems to meet (some of) Herder's argument, thus contradicting Chomsky's view of an innateness of language. Cf. Daniel L. Everett, *How Language Began. The Story of Humanity's Greatest Invention* (W. W. Norton & Company, 2017).

[14] "To be human is to imitate [...]. Most living beings on Earth are the product of evolution based on the copying, varying and selection of genes. However, once humans began to imitate they provided a new type of copying and so let loose an evolutionary process based on the copying, varying and selection of memes. This new evolutionary system co-evolved with the old to turn us into more than gene machines. We, alone on this planet, are also meme ma-

Richard Dawkins' work, *The Selfish Gene*), which he defines as "instructions for performing behaviors, stored in the brain (or in other objects) and passed on by imitation."[15] Let us just say that despite these theories, and even with the discovery of mirror neurons (a kind of neuronal justification for imitative processes), there are still many uncertainties about their validity, since there is no scientific agreement on that.

The third chapter reveals the genius of Herder's thought as it justifies from the internal point of view (the human soul) and from the external point of view (the social history of the languages and peoples of the world), the human invention of language.[16] To this end, the philosopher introduced a curious triangulation that operates in the game between sonorities, ear and inner language (hearing, sound and reflection), and will advance with the conclusion that he will explore in the second part saying "that man must had necessarily have invented a language himself, and state under what circumstances this could have been most easily effected."[17]

In fact, and as we have already mentioned, Herder pays special attention to hearing by attributing to it the responsibility of interconnecting the heard sound and the inner resound of meaning (as the attribution of identifying and/or differentiating characteristics of external reality), that will assent in the interiority of man to the functional unveiling of reason and language.[18] While not neglecting the importance of the other senses for the process, Herder submits them to the

---

chines. We are selective imitation devices in an evolutionary arms race with a new replicator. This is why we are so different from other creatures; this is why we alone have big brains, language and complex culture." Susan Blackmore, "A imitação faz de nós humanos," in *O que nos torna humanos?* ed. Charles Pasternack, 30-46 (Texto e Grafia, 2009), 30.

[15] Apud Susan Blackmore, *The Meme Machine* (Oxford University Press, 1999), 17.

[16] "The origin of language in the human soul, is as capable of demonstration, as any philosophical evidence whatsoever; and the external analogy between all ages, languages, and nations, has as high a degree of probability as the most established historical fact could possibly have." Herder, 74.

[17] Ibid.

[18] Using a set of practical everyday examples, Herder explains the relevance of hearing in the process: "The turtle dove coos, the dog barks, thus arise three words, because he endeavored to seize three clear ideas, the ideas he marked down in his logic, the word in his dictionary. Reason and language advanced a timid step together, and nature came half-way to meet them, with the assistance of the ear. She not only brought forward the sounding tones, but caused them to penetrate to the depth of the soul, a sound is heard, the soul catches the sound, and has thus gained a sounding word! Man, therefore, as a listening observing creature, is naturally constituted for language; and even a person born blind and dumb, must form a language, unless he were also deaf and devoid of feeling." Ibid., 40.

relevance of hearing,[19] because it is "easy to comprehend how words arose from sounds, and were stamped as signs by the understanding."[20] Meaning that in the order of sensation (all feeling) will immediately have its "sound" and since the ear is a language organ that unifies the totality of the sensations that sounded, the conduction to the plane in which a characteristic is rationally attributed will pass to exist a word for such an evocation.[21] This is exactly what is said to support the notion of the ear as a central sense for that "creature of reflection and language, of consciousness and linguistic creativity" that is man. In Herder's words:

> As man attains to speech with the aid of instructive nature; by means of the hearing, without which he could not invent language, the hearing may be termed the central sense, the portal of the soul, and the bond of connection, between the other senses.[22]

It also happens that languages evolve, prepare and develop broader concepts and, therefore, bring also more abstract concepts, which according to Herder confirm once again the human origin of language:

> As human reason cannot exist without abstraction, and as no abstraction develops itself without language, the language of every nation must contain abstract ideas, *i.e.* must convey an impression of reason, from having been its instrument. Each language only contains as much abstraction, as the nation was capable of, and has no one abstract idea, independently of the senses, which is proved by its original sentient expressions. Consequently, no other di-

[19] Herder refers to the primacy of the ear over touch and vision, because in the sensations that the world offers, it is through sounds (objects, according to Herder, always sound in some way) that they are represented internally: "Feeling approaches very near to hearing, its designations, e.g. hard, rough, soft, woolly, velvet, hairy, stiff, polished, smooth, bristly, etc., which all refer merely to surface, all sound as if they were felt [...]. The words scent, tone, sweet, bitter, sour, etc., all sound as though they were felt: for what were all the senses originally but feeling?" Ibid., 51.

[20] Ibid., 48.

[21] The philosopher justifies the importance of the ear through the clarity and distinction that allows us to say that it is felt for language: "Hearing seizes something from both sides, renders clear what is too dark, and softens what is too bright, brings more unity into the obscure variety of feeling, more unity into the too brilliant variety of vision, and this recognition of many things, by one, by a sign, gave rise to language." Ibid., 53.

[22] Ibid., 52.

> vine arrangement is perceptible, excepting this, that language is human throughout.[23]

It should be noted that Herder is throughout the essay articulating (almost) unnoticed a fundamental concept that will give an extraordinary and ingenious unity to his theory and which is, in the words of José M. Justo, the "device" of totality (or globality). It is with him and from him that it makes sense to think about the global unity of man in becoming, in the historical course of his acquisitions and, thus, the totality of man makes itself resonate in the totality of the constitutive process of language by successively improving the reflexive states (in the interiority) of his being.[24]

In the second part of the essay, and on the basis of much of the argument already developed, Herder will establish the natural laws that condense the laws of nature and of the human species, with regard to its predisposition to language.

Thus, the First natural law states that: "Man is a free thinking, active being, whose powers operate progressively, and as such is a creature formed for speech!"[25]

Man is a being that is, by his nature, predisposed to develop himself, so his first moment of internal awareness would also have to be that of the inner birth of language. For Herder, man is a man from the moment he is placed in the world, and although he may not yet be a creature of consciousness, he is already a creature of reflection (since all states of reflection are linguistic states, that is, "a chain of thoughts is a chain of words"). In this way, the formation of language is a process that develops as naturally as the formation of human nature itself.

---

[23] Ibid., 67-68.

[24] "This means that each state of this process is a condition in a double sense: a condition of what can be operated with this configuration (for example, in the initial state, a condition for the internalization of characteristics) and a condition of the transformations to which the configuration will be subjected (for example extreme, the initial state contains 'in nuce' the necessary conditions for man to slowly transform himself into what he is today and, of course, into what he will be tomorrow). From this it follows that the genesis (from the moment Herder was able to face it), far from being an additive linear path, a mechanical chain of causes and effects in which there would be no place to talk about progress because all moments would have the same value, becomes a path of increasing complexity, a sequential articulation of states in which each one collects the wealth of the previous ones to prepare the following ones and in which each state being a 'living whole' produces more than the mechanical sum of the parts." Johann Goottfried Herder, *Ensaio sobre a Origem da Linguagem*, introduction and trans. José M. Justo (Antígona Editores, 1987), 16.

[25] Herder, 75.

The second natural law establishes that "Man is by destination, a gregarious and social creature, the cultivation of language is, therefore, natural, essential and necessary to him."[26]

Just as it is natural for a creature to develop within a community, it is natural for a man to develop linguistically among men. Herder says that no man exists for himself, that is, men share a nature that prevents them from uprooting themselves from the human species. He also adds that since man is a social being by essence, it would make no sense not to have a means of communication, that is, in absurd this would contradict the very notion of being social. From this also follows the diversity of languages that the third law and fourth law come to shape, as can be seen:

The third natural law dictates that "the human race could not possibly continue only one flock, confined to one language, therefore, the formation of different national languages became necessary."[27]

The fourth natural law:

> In all probability the human race constitutes one progressive totality, from one source, and forming one vast household. The same principle refers to languages, and with them to the whole chain of cultivation.[28]

The philosopher is emphasizing what was already implied before the expression of the laws; but with them, it allows him to underline, with the richness of his anthropological thought, namely, that not only is humanity one and the same, but also that language is reproduced and develops in the proximity of humanity, or to use Herder's nomenclature, with the human race (once again the cohesion of the philosopher's discourse is felt by the constant reference to globality in the proximity of the essentiality of human nature). It can be said, therefore, that Herder's thought expands and opens doors to think about human existence with language (in what can be seen as an affirmation of the coexistence of the species and its cultural legacy):

> It may, therefore, be affirmed, that there exists no thought, no invention, no step towards perfection, which may not be extended ad infinitum. I can perform no action, can entertain no thought, which may not influence the immeasur-

[26] Ibid., 90.

[27] Ibid., 99.

[28] Ibid., 107.

> able course of my existence. So also, there is no creature of my species that does not influence the whole species, and the progressive total of the whole species. Every one impels a greater or lesser wave, every one alters the state of an individual soul, therefore the total of these states always acting upon others, therefore changing something in them. The first thought in the first human soul, stands in connection with the last in the last human soul.[29]

III. Herder's legacy

The cogitation operated by Herder around the human origin of language allowed us to understand that the human species and language are in permanent evolution.[30] If there is an effective history of progress, then it must consider language as a fundamental acquisition from which there would not be this same history. It should also be considered that there is in Herder's essay a kind of teleology (of evident Kantian background) for humanity that reveals itself precisely through the conception of an incompleteness of language and of man, but which would tend towards perfectionism (of which metaphysical languages may constitute a first sample). In accordance with this, such purpose finds meaning in the horizons that it constitutes and, therefore, expanding the experience of being to a being that reinvents language.[31] Perhaps here Herder's essay gains a new meaning by alerting us to the permanent reconstruction that man makes of himself and the knowledge he generates. It is from this inventive capacity of man (inscribed in his nature) that the blossoming of the different languages that would transform the world takes place (Castro in his book about Heidegger call it the onto- potentiality of language).[32] Such a conception also allows us to understand that not all inventions can be made with the fortune of language in their creation.

---

[29] Ibid., 108.

[30] It is said: "The divine origin is rather injurious than beneficial, it destroys all the activity of the human soul, and renders both psychology and the sciences inexplicable. For with language man must have received the seeds of all knowledge from God? Nothing, therefore, proceeds from the human soul. The commencement of every art and science, and of all knowledge, must be thus rendered inconceivable. The human origin admits of no step, without some view, or without the most useful elucidation in every branch of philosophy, in all kinds and compositions of languages." Ibid., 118.

[31] Cf. Sue Savage-Rumbaugh et al., *Apes, Language, and the Human Mind* (Oxford University Press, 1998).

[32] Cf. Paulo Alexandre E. Castro, *Ontopotencialidade da Linguagem: Breve ensaio para compreender o essencial da linguagem em Heidegger* (BonD, 2024).

So, taking all of this into account, one can see that the capacity of hearing, that is, to be aware of sounds, will represent the emergency and consolidation of reflection – which nothing less than to be aware of its own thoughts or if one prefers to be aware of the existence of consciousness – that will give rise to the appearance of abstracts words. What is not said, but is somehow implied between the lines, is that the understanding of abstract words by others allows the expansion of language and therefore its enrichment. But it is not all: from an externalist point of view, this will mean the assertion of the existence of other minds (although such a slight assertion may be contested), since they would be able to reach the understanding of these words. Now, such scope refers to an internal conceptualization (reflection) that reveals all the characteristics that are attributed to a mind. In this sense, the listening and reflection highlighted by Herder reveal the deep meaning of the human mind. Herder left us a very rich text that allows different readings; this reading is perhaps just another one that touches on a set of topics that have not yet exhausted the theme.

References

Blackmore, Susan. "A imitação faz de nós humanos." In *O que nos torna humanos?* edited by Charles Pasternack, 30-46. Texto e Grafia, 2009.

Blackmore, Susan. *The Meme Machine*. Oxford University Press, 1999.

Castro, Paulo Alexandre E. *Ontopotencialidade da Linguagem: Breve ensaio para compreender o essencial da linguagem em Heidegger*. BonD, 2024.

Everett, Daniel L. *How Language Began. The Story of Humanity's Greatest Invention*. W. W. Norton & Company, 2017.

Herder, Johann Gottfried. *Ensaio sobre a Origem da Linguagem*. Introduction and translation by José M. Justo. Antígona Editores, 1987.

Herder, Johann Gottfried. *Treatise upon the Origin of Language*. Camberwell Press, 1827.

Savage-Rumbaugh, Sue, Stuart G. Shanker, and Talbot. J. Taylor. *Apes, Language, and the Human Mind.* Oxford University Press, 1998.

S. Demirhan · *Conatus* 10, no. 1 (2025): 87-109
doi: https://doi.org/10.12681/cjp.37271

Negotiating Autonomy: Lived Experiences of Female Living Organ Donors in Turkey

Sezen Demirhan
University of Luxembourg, Luxemburg
E-mail address: sezensecgin@gmail.com
ORCID iD: https://orcid.org/0000-0002-8999-8318

Abstract

One of the most significant developments in the field of health in the past century is organ transplantation. While often regarded as a life-saving solution for patients with end-stage organ failure, the lived experiences of living organ donors – especially women – remain underexplored in the literature. This study, conducted between 2022 and 2024, employed qualitative methods and a feminist phenomenological design. The study examines how socio-cultural expectations, kinship obligations, and internalized gender norms intersect to influence women's decision-making processes in living organ donation. Among living donors, those who donate a liver or a kidney take on significant physical and psychological risks, making their perspectives particularly important for ethical reflection. Based on interviews with 18 female donors in Turkey, the findings reveal how women's lived experiences of donation are shaped by their embodied vulnerability, relational roles, and the moral weight of familial obligation. Rather than autonomous acts made in isolation, these decisions emerge within gendered landscapes marked by asymmetrical power dynamics and cultural expectations. By attending to how women articulate their experiences of bodily sacrifice, risk, and post-donation subjectivity, this phenomenological inquiry highlights the necessity of integrating a gender-sensitive lens into bioethical discourse – one that recognizes how normative frameworks and structural inequalities shape and constrain women's autonomy in living organ donation.

Keywords: *lived experience; autonomy; organ transplantation; living organ donor; Turkey*

I. Introduction

In the 21st century, with the almost complete authority of medicine over the biological body, death is being challenged, reshaping the meanings of concepts such as ethics, body, life, and death.[1]

[1] Philippe Ariès, *The Hour of Our Death* (Oxford University Press, 1981); see also Roberto Andorno and George Boutlas, "Global Bioethics in the Post-Coronavirus Era: A Discussion with Roberto Andorno," *Conatus – Journal of Philosophy* 7, no. 1 (2022): 185-200.

Changing living conditions reveal different disease models and offer new treatment methods.[2] Organ transplantation, one of the most remarkable applications in medicine, changed the body's fate and included it in an endless project. One of the ongoing challenges in organ transplantation is the scarcity of resources, particularly cadaveric donors, prompting some countries to promote living organ donation and explore alternative methods such as utilizing anencephalic newborns as donors[3] and investigating xenotransplantation.

This advancement in medicine does not affect men and women equally. Gender inequalities observed in various aspects of social life are also reflected in organ transplantation practices.[4] Gender has a considerable impact on donor availability, access to services,[5] medical biases, and post-transplant care responsibilities,[6] all of which are influenced by cultural norms and gender. The prevailing masculine hierarchy within medicine, coupled with the historical normalization of the male body as the standard, along with the underrepresentation of women in medical research (exemplified by the lack of focus on issues like breast cancer),[7] perpetuates the notion of women as potential donors. Femi-

---

[2] Zeljko Kaludjerovic, "Bioethics and Hereditary Genetic Modifications," *Conatus – Journal of Philosophy* 3, no. 1 (2019): 31-44.

[3] Charles N. Rock, "The Living Dead: Anencephaly and Organ Donation," *NYLS Journal of Human Rights* 7, no. 1 (1989): 243-277.

[4] Annika Gompers et al., "Intersectional Race and Gender Disparities in Kidney Transplant Access in the United States: A Scoping Review," *BMC Nephrol* 25, no 1. (2024): 36; Sanshriti Chauhan et al., "Nationwide Data on Gender Disparity in Solid Organ Transplantation for India in the Pre-pandemic and Pandemic Era," *Transplantation* 6, no. 9 (2022): 230; Amelie Kurnikowski et al., "Country-specific Sex Disparities in Living Kidney Donation," *Nephrology, Dialysis, Transplantation* 37, no. 3 (2022): 595-598; Michael Darden et al., "Persistent Sex Disparity in Liver Transplantation Rates," *Surgery* 169, no. 3 (2021): 694-699; Javeria Peracha et al., "Gender Disparity in Living-Donor Kidney Transplant Among Minority Ethnic Groups," *Experimental and Clinical Transplantation* 14, no. 2 (2016): 139-145; Cecilia M. Øien et al., "Gender Imbalance among Donors in Living Kidney Transplantation: The Norwegian Experience," *Nephrology Dialysis Transplantation* 20, no. 4 (2005): 783-789; Anette Melk et al., "Sex Disparities in Dialysis Initiation, Access to Waitlist, Transplantation and Transplant Outcome in German Patients With Renal Disease – A Population-Based Analysis," *PLoS ONE* 15, no. 11 (2020): e0241556; Ravikiran S. Karnam et al., "Sex Disparity in Liver Transplant and Access to Living Donation," *JAMA Surgery* 156, no. 11 (2021): 1010-1017; Francesca Puoti et al., "Organ Transplantation and Gender Differences: A Paradigmatic Example of Intertwining Between Biological and Sociocultural Determinants," *Biology of Sex Differences* 7, no. 1 (2016): 35.

[5] Jessica B. Rubin et al., "Organ Transplantation and Gender Differences: A Paradigmatic Example of Intertwining Between Biological and Sociocultural Determinants," *World Journal of Gastroenterology* 25, no. 8 (2019): 980-988.

[6] Ya-Ping Lin, "Visible Body, Invisible Care: Family, Gender Politics, and the Female Caregiver in Living Donor Liver Transplantation in Taiwan," *SSM - Qualitative Research in Health* 4 (2023): 100346.

[7] Janet R. Osuch et al., "A Historical Perspective on Breast Cancer Activism in the United

nist bioethicists delve into issues such as gender discrimination, power dynamics, imbalances/disparity in transplantation processes, and autonomy, whose voices are taken into account and whose are disregarded in transplantation decisions. They also explore how medical practices intersect with societal norms.[8] The primary objective of the study is to comprehend and interpret women's decision-making experiences of living organ donation in Turkey within the context of gender.

Living organ transplantation depends on some bodies giving up their organs to provide treatment for others. While this renunciation is justified in low-income countries as a nation-specific example of self-sacrifice,[9] it is important who will risk their body and life through organ donation, and who will benefit from this risk. When examining living organ transplants, it becomes evident that women are more frequently at risk.[10] In traditional patriarchal societies, mothers, sisters, and wives are expected to make sacrifices[11] for their country, family, and children and give up their bodies. Thus, both gender and autonomy emerge as important variables in organ transplantation as well as other inequalities in the field of health. The existing gender gap in organ transplantation underscores the need to explore the multifaceted gender perspectives within the context of organ transplantation.[12] Research has consistently revealed significant gender disparities among both organ recipients and donors, underscoring a pressing issue that warrants further investigation. Additionally, numerous studies have substantiated the presence of substantial socio-ethical and biological implications surrounding organ donation, particularly within the framework of gender dynamics. These observed disparities prompt crucial inquiries into individual autonomy and the ethical considerations inherent in organ transplantation.

This qualitative study was conducted in Turkey between 2022 and 2024, employing a phenomenological design. It is important to note

States: From Education and Support to Partnership in Scientific Research," *Journal of Women's Health* 21, no. 3 (2012): 355-362.

[8] See Darija Rupčić Kelam and Ivica Kelam, "Care and Empathy as a Crucial Quality for Social Change," *Conatus – Journal of Philosophy* 7, no. 2 (2022): 157-172.

[9] Megan Crowley-Matoka, *Domesticating Organ Transplant: Familial Sacrifice and National Aspiration in Mexico* (Duke University Press, 2016).

[10] Wendy A. Rogers et al., eds., *The Routledge Handbook of Feminist Bioethics* (Routledge, 2022); Laura Rota-Musoll et al., "An Intersectional Gender Analysis in Kidney Transplantation: Women Who Donate a Kidney," *BMC Nephrology* 22, no. 1 (2021): 59-69.

[11] Ann Mongoven, "Sharing Our Body and Blood: Organ Donation and Feminist Critiques of Sacrifice," *The Journal of Medicine and Philosophy* 28 no. 1 (2003): 89-114.

[12] Vivek Kute et al., "Act Together and Act Now to Overcome Gender Disparity in Organ Transplantation," *Experimental and Clinical Transplantation* 22, no. 1 (2024): 17-27.

that the act of contemplating donation and the experience of actual donation are two distinct phenomena. Data were collected through in-depth interviews with a purposive sample of 18 women aged 18-55 who volunteered for kidney and liver donation. The names of the interviewees were coded and changed. The participants were voluntary female donors from various hospitals in Turkey over the past decade. The youngest of the women interviewed was 22, and the oldest was 70 years old. When they became donors, the youngest was 18, and the oldest was 47 years old. 4 kidney and 14 liver donors were interviewed. The study examines the experiences of women voluntarily donating their liver and kidneys within the context of gender, utilizing descriptive analysis. The data were analyzed thematically, focusing on themes such as family relationships, patriarchal structures, criticism of disclosure, and altruism, which influence women's autonomy in living organ donation decisions. The study integrated feminist theory and the concept of autonomy to interpret these experiences. The unique perspective of women in a phenomenological study with feminist concerns is a powerful counter to all tendencies that objectify the body. Through feminist phenomenology, we can now include the experiences of women who have undergone living organ donation, allowing them to express their experiences in their own words. This unique perspective is chosen to highlight the criticism that, in organ transplantation practices, while there is ample emphasis on treatments, medical terms, and narratives, as well as the comments of renowned doctors performing numerous transplants each year, the experiences and perspectives of patients and donors are often overlooked.

II. Gendered autonomy in living organ donation

It's not just medical advancements that make organ transplantation possible, but also autonomy and consent procedures. Autonomy in medicine can be defined as the principle whereby individuals possess the ability and right to make their own decisions, emphasizing their freedom to make informed choices regarding their treatments and healthcare situations. With this signature, the medicine authority legally waives all medical operation responsibilities. This consent is obtained by filling out and signing a consent form. While the decision to participate in organ donation is often perceived as a personal choice, it is a socially influenced decision shaped by factors such as societal gender expectations as sacrifice culture, norms of selflessness, altruism, body control policies, family and community expectations (expectations of caregiving and nurturing), perceptions of risk and safety,

healthcare decision-making dynamics, religious and cultural beliefs, economic considerations, educational attainment and awareness, legal and ethical considerations. Feminist autonomy theories delve deep into how internalized and external oppression influence an individual's overall and specific autonomy.[13] There is no consensus as to which theoretical position is correct. Nevertheless, there is a substantial body of evidence to suggest that oppressive socialization and oppressive practices have a detrimental impact on autonomy, potentially leading to its complete erosion. We need to thoroughly explore the practice of living organ transplantation to support feminist autonomy theories, which aim to empower women to make independent choices about their bodies and health. Otherwise, it would not be an exaggeration to claim that in the years to come, we will have a worldwide population of women with one kidney and the health problems experienced by these women. This study does not aim at a discussion on Feminist autonomy theories, but to contribute to Feminist autonomy theories by including women's experiences and their interpretations of autonomy, which is of central importance in living organ transplantation practices as a specific example.[14] In particular, relational autonomy, claimed to be self-determination is inherently social, and "the ethics of care" put forward by Gilligan[15] may effectively guide the debate.

Autonomy is the pursuit of personal independence and the desire for dialogue and negotiation with others. However, in the context of gender inequalities, autonomy often intersects with societal expectations and power dynamics, particularly concerning the female body.[16] Understanding autonomy in living organ donation involves recognizing the interplay of medical advancements, consent procedures, and societal influences, especially those related to gender. Feminist autonomy theories emphasize how societal expectations and oppressive practices can undermine individuals' autonomy.[17] While there may be differing

[13] See Andrea Ellner, "Ethics of Conflict, Violence and Peace – Just War and a Feminist Ethic of Care," *Conatus – Journal of Philosophy* 8, no. 2 (2023), 147-173.

[14] Natalie Stoljar, "Feminist Perspectives on Autonomy," *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), eds. Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/sum2024/entries/feminism-autonomy/.

[15] Carol Gilligan, *In a Different Voice: Psychological Theory and Women's Development* (Harvard University Press, 1993).

[16] For a very interesting account of how autonomy intersects with societal dynamics in another cultural environment, see Dung Van Vo, "Four Important Characteristics of Women in Confucianism and Its Contribution to the Implementation of Gender Equality in Vietnam," *Conatus – Journal of Philosophy* 9, no. 2 (2024): 283-302.

[17] Andrea Veltman and Mark Piper, eds., *Autonomy, Oppression, and Gender*, *Studies in Feminist*

theoretical positions, evidence suggests that oppressive socialization can significantly impact autonomy, potentially eroding it completely. Therefore, it is essential to further explore living organ transplantation practices from a feminist perspective, considering power dynamics, consent processes, application differences, transparency of gender data,[18] and long-term effects.

III. Results

The analysis identified four key factors shaping autonomy in living organ donation decisions. Family relationships (constructed expectations to be accepted, cared for, and self-understanding, particularly in the context of motherhood, where the societal expectation of a mother's selflessness influences the decision to donate) emerged as pivotal, influencing donors' sense of obligation and support. Paternalistic attitudes in healthcare settings often constrain donors' decision-making agencies. Criticism of disclosure highlighted donors' challenges in navigating medical information. Altruism played a significant role, intertwining personal sacrifice with moral duty and normalization of the process.[19]

IV. Relationships

The family structure in Turkey has implications beyond affection and mediates women's pursuit of self-assurance, expression, and need for acceptance. Women's acceptance of organ donation seems to be influenced by the need to repair relations and gain acceptance within family relationships. The interviews indicate that the family is a contentious environment.

The most thought-provoking finding concerning autonomy is that some women hope to gain power and advantage in this critical situation. Their longing to take full control of their lives has been a motivating factor. One interviewee highlights how she escaped discrimination

---

Philosophy (Oxford Academic, 2014).

[18] Despite the Global Observatory on Organ Donation and Transplantation (GODT) being the preeminent repository of international data on donation and transplantation rates, gender data was not included until 2017. This information is now available for the first time in the GODT's 2017 annual report.

[19] Especially on the notion of *effective altruism*, see Iraklis Ioannidis, "Shackling the Poor, or Effective Altruism: A Critique of the Philosophical Foundation of Effective Altruism," *Conatus – Journal of Philosophy* 5, no. 2 (2020): 25-46. See also Julian Savulescu and Evangelos D. Protopapadakis, "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu," *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

experienced as a girl within her family by becoming a live organ donor. Nergiz stated, "For example, men were always in the foreground in our family. My inability to study was due to the idea that girls don't study. I just finished my new school, that is, I studied externally. When my family saw my willingness to be a donor, they wondered why we didn't do it earlier or didn't allow it. For example, I am currently going through a divorce process. If I had said this five years ago, they would have opposed it, saying it would never happen. But now they don't think that way; they say we're behind you in every decision you make. They say you're strong, you can do it, so it's something much different than before. There used to be a distinction between girls and boys in the family. But not anymore. It's like nothing happened after my surgery."

Offering her life as a bargaining chip to gain approval and acceptance from her family, she takes the risk of becoming a living organ donor for her father. She emphasizes that she has empowered herself by making a significant sacrifice and gained control over her life. The women decided voluntarily, deliberately, and without pressure, but this decision resulted from specific calculations and comparisons, suggesting that the influence of gender cannot be denied in decision-making.

The women interviewed felt that they needed to make more effort than men to gain respect and affection from society. This situation aligns with Beauvoir's construction[20] of "absolute otherness." Sinem summarized the situation by saying, "I always tried to make people love me. I always gave of myself because I thought they wouldn't love me if I didn't give; I always made concessions." Through such sacrifices, women strive to prove they are strong, brave, and valuable. However, this effort is often ignored or not sufficiently appreciated after transplantation, which can lead to both emotional and physical injuries for women in the context of organ donation. Fourteen of the respondents are receiving psychological support, mostly in the form of medication. Studies underscore the urgent need for increased support and care for living donors who often face severe psychiatric challenges.[21] The pre-transplant screening of organ donor candidates identifies psychosocial contraindications: "The donor candidate should be under pressure, any untreated psychiatric disorder that may affect decision making, active drug, substance or alcohol addiction, high suspicion of secondary gain, and the candidate should refuse to give written con-

[20] Simone de Beauvoir, *The Second Sex* (Vintage Classics, 2015).

[21] James F. Trotter et al., "Severe Psychiatric Problems in Right Hepatic Lobe Donors for Living Donor Liver Transplantation," *Transplantation* 83, no. 11 (2007): 1506-1508.

sent or be unable to give consent."[22] Some of the women interviewed shared that they had previously received psychological treatment, medication or therapeutic support. However, this was not enough to prevent their consent. The women also noted that the drug treatments were intensified after the transplant for reasons such as the fact that the issue was not discussed in the family after the transplant because it did not burden the recipient and because the donor's sacrifice was not appreciated much. This highlights the pressing need for post-transplant support and care for these individuals.

Nine of the participants were married and had children when they became donors. Despite their status as mothers, they positioned themselves as children and daughters when they decided to become donors. Women's re-evaluation of themselves as children in their perception of identity made the status of donors possible. In her work, *The Second Sex*, Simone de Beauvoir analyses the reasons for the infantilization and dependence of women by society.[23] She analyses how women are socialized from childhood and how this process renders them dependent. In this context, she focuses on positioning women as the "other" and preventing their full maturation as individuals. The decision to become an organ donor is influenced by the fact that women tend to see themselves as their parents' children first and foremost. This supports Beauvoir's view; also Nancy Chodorow's *The Reproduction of Mothering* examines how women's relationships with their mothers shape their identities and how they develop a childish dependence and passivity in this process.[24] Chodorow argues that women's close bonds with their mothers cause them to feel like children in adulthood and, therefore, assume dependent roles. This theory can explain women's positioning of themselves as children when they decide to become organ donors. However, at the same time, this action also points to a step that women choose to get rid of in childhood. One interviewee stated, "Our relationship is like I am a mother and she is a child; my mother is not my mother but my child. Psychologically, I don't have a mother; I don't have a mother figure that I can consult. I also studied at university with them. When I got married, I left for the first time and went to Istanbul. My mother was so united with me that she could not stay far away from me. In 2014, I settled in Istanbul. Three years later,

---

[22] K. Keven and S. Aktürk, *Transplantasyona Hazirlik Verici. A. Türkmen (Dü.) Inside,Transplantasyon Nefrolojisi Pratik Uygulama Önerileri (S. 9-25)* (Türk Nefroloji Derneği, 2016).

[23] Beauvoir.

[24] Nancy Chodorow, *The Reproduction of Mothering: Psychoanalysis and the Sociology of Gender* (University of California Press, 1999).

my mother became ill and could not stay away from me that much and preferred to keep a part of me (liver) with her. So she made herself psychologically ill, actually; I guess you could say she took it from me by force. Because she wanted a piece of me to stay with her. So I gave it to her, and I was liberated." While cutting off the organ, she also cuts off the relationship; at this point, the woman exhibits symbolic autonomy. It is possible to see a similar situation in the relationship of another interviewee who did not keep in touch with her mother after the transplant. Sevgi said, "Then I said, 'Girl, you should not do this much to yourself anymore', and I stopped seeing her." Another woman, Hayriye, said, "It's hard being a woman. That's what the environment expects in general! I don't think I've been a good daughter in my own opinion (she thought that was because she'd had three different marriages). I feel like I've let them down. If our parents are alive, we always tend to be good children even when we are 60 years old. But society does that to us."

In the context of autonomy, it has been observed that female donors do not perceive themselves as isolated, independent, autonomous individuals within this network of family relationships. Women are constructed within societal norms as daughters, mothers, homemakers, nurturers, providers of care, and sources of comfort, which naturalizes this social role. This finding supports Gilligan's theory.[25] Therefore, it seems appropriate to reopen the discussion on feminist care ethics[26] within the context of living organ donation. Women who donate organs to their children or other family members are influenced by this societal gender role, positioning themselves in caregiving roles based on the biological assumption of their reproductive capabilities. Piraye was the oldest of the interviewees. Today she is 70 years old. She was 47 years old when she was a kidney donor for his brother. She lost her brother 7 years after the operation. When she told her husband that she would be a kidney donor for her brother, her ex-husband and his family opposed her decision, saying that she would regret it if her children needed it in the future. This approach shows the acceptance of women as organ providers within the family. This situation also points to tensions between the nuclear family and the extended family. Although she stated that she did not give up on her decision by saying;

[25] Gilligan approached the moral development of women differently from men and argued that traditional moral theories did not adequately consider the female experience. While Kohlberg defined the male moral perspective as justice ethics, Gilligan described the moral perspective of women as care ethics.

[26] Nel Noddings, *Starting at Home: Caring and Social Policy* (University of California Press, 2002); Rosemaria Tong, *Feminist Approaches to Bioethics* (Routledge, 1997).

"Then you or your family give it to the children, I will give this one to my brother and the other one to the cats, don't interfere," the motivation for giving organs to his brother was that they grew up without a father and she accepted his brother as a father.

Some women said they didn't love the person who received their donated organ, but they still went ahead with the donation. In three cases, the most significant source of motivation was not love,[27] which is contrary to what we are used to hearing in organ donation calls. Ebru said, "My mum might be the unhappiest person I know. She is a person who does not take care of her health, who does not care about anyone, her husband, her mother, her children, and who looks like that (The mother has cared for a disabled husband for many years and continues to do so). I was thinking that maybe she could say, "Don't let anyone be a donor for me. I don't want anything from anyone like this is my life, and I'm going. She's a person who's given up on her life. I don't remember many happy times for my mum. She didn't do anything when I said I was giving my liver to her. She acted like it was normal." Nancy Chodorow[28] has particularly analyzed how women internalize maternal roles and how these roles reproduce gender norms. The interviewee seems to be deeply shaken by her mother's lack of appreciation for her sacrifice, perceiving it as normal, expected behavior. However, it is expected that a mother would not accept a practice that harms and endangers her child's life to save her own.

V. Patriarchal structures

It is important in the context of rights to argue that women may not have real autonomy over their bodies due to patriarchal norms.[29] In patriarchal societies, women are often confined to specific roles, which are generally defined as weak or secondary. Within this patriarchal system, women often have to cope with feelings of weakness and vulnerability.[30] One of the interviewees stated, "I was a bit more delicate in the family's eyes, I guess. No one thought I could be so brave; my cousins told me that. They said they didn't think I could be so brave. Of course, it has changed. I used to be very afraid of getting tattoos

[27] Kristin Zeiler, "Just Love in Live Organ Donation," *Medicine, Health Care and Philosophy* 12, no. 3 (2009): 323-331.

[28] Chodorow.

[29] Sylvia Walby, *Theorizing Patriarchy* (Basil Blackwell, 1990).

[30] Iris Marion Young, *On Female Body Experience: "Throwing Like a Girl" and Other Essays* (Oxford University Press, 2005).

or piercings, so I never thought I could do something like this... Then I said, I guess I can do it; there is this power in me." When examined from a gender perspective, women's participation in living organ donation can be considered a challenge to the attributes typically associated with men, such as heroism, bravery, strength, resilience, and determination.

I was able to interview one of the families who participated in a cross-transplant. In this case, two women, whose husbands needed kidney transplants, were tested as potential donors. Since neither was compatible with their own husband, they each donated their kidney to the other woman's husband – saving both men's lives. One of the women who donated her kidney, Zehra, lived in Central Anatolia and faced economic hardships that limited her access to communication tools such as the internet and a telephone. Because of this, she could not be reached for an interview. Additionally, the family who received her kidney chose not to share her contact information. As a result, I was unable to speak with her. However, the woman I interviewed shared the following about Zehra's experience: "We provided financial support, and Zehra came on the bus. During the medical tests, a gynecological issue was discovered, which had to be treated first. Later, we also found out that Zehra couldn't read or write – something she had been too ashamed to reveal. Because of this, all the procedures were initially cancelled. Zehra turned red with embarrassment. Then her husband spoke up: 'Hodja, she can't read or write. She was ashamed to tell you! She can't sign.' So we had to start over again. Zehra later signed the consent form at the notary." A feminist perspective underscores the importance of recognizing and addressing power dynamics in discussions of autonomy, particularly in the context of women's experiences in organ donation.

The study did not find research examining the gender factor in the structuring of ethical committees for living organ transplantation. It observed that there is no targeted gender equality in the composition of committee members from a societal gender perspective, with almost all committees consisting predominantly of men. It becomes inevitable that an ethics committee composed entirely of men will be unable to address the needs of gender inequalities. The fact that Zehra could not be recognized as illiterate after all the tests had been carried out could only have been possible under the impression of a careful and gender-conscious committee. In this way, Zehra would not have travelled all this way, would not have been embarrassed because of her economic status and illiteracy, and would not have been subjected

to a series of tests. As stated in Metin's text,[31] ethics committees are products of pluralistic liberal societies in the West and owe their existence to the bureaucratic institutional structure of modernity. Ethics committees should not be instruments of bureaucratic regulation and control. It should be freed to play a critical role within the institution, to support and develop ethical research and researchers, and given time to discuss and explore difficult ethical issues where they arise.[32] The team is exclusively male. An ethics committee will inevitably be unable to respond to the needs of her.

One of the interviewees, who is the youngest child of a family with five siblings and who had just given birth, stated that the donation process was started at the same time for all five siblings and that the tests were performed at the same time and that the doctor, with a paper in his hand and based on biological data, selected her as the most suitable donor among the candidates and informed everyone. In this context, although all five siblings were compatible donors, the possibility of the donor, who was the youngest sibling selected by the doctor, objecting to the doctor as a medical authority was weakened. As Sherwin states,

> The reality is that in a hierarchical society, most patients have much less power than most doctors. As a result, many patients are virtually incapable of making truly 'autonomous' decisions in the presence of doctors.[33]

One interviewee answered the question as follows: Who else in the family could have been a donor besides you to your uncle? Were there other people who volunteered? "My brother, one of his friends, and I went as donors. When the doctors saw them, he said, 'They have a belly; they can't be donors, let's start with you.' When that happened, and I was a match, no one else took the test." Women who are donors have been given organs because there are no volunteers around them other than themselves or because they have been deemed suitable by the hospital, doctors, or ethical committee. In this scenario, it's crucial to carefully consider both paternalism, patriarchal structures, and the doctors' autonomy.[34] Both factors play significant roles in deci-

---

[31] Sevtap Metin, *Biyo-Tıp Etiği ve Hukuk* (Betim Kitaplığı, 2019).

[32] Paul M. McNeill, "Research Ethics Review and the Bureaucracy," *Monash Bioethics Review* 21, no. 3 (2002): 72-73.

[33] Susan Sherwin, *No Longer Patient: Feminist Ethics and Health Care* (Temple University Press, 1992).

[34] For an enlightening discussion on physicians' autonomy, since it discusses autonomy with

sion-making. A doctor who has been socialized in a patriarchal family structure, where traditional gender roles are the norm, and who subscribes to the view that men subjugate women, or who perceives themselves as the natural choice to spearhead this mission due to their role as a caregiver, will, make their choice in favor of women.

The impact of patriarchal structures on women's autonomy in medical decision-making is significant. Women organ donors face societal and gender dynamics that influence their experiences. Gender-sensitive approaches in ethical decision-making are essential to address the challenges faced by women donors and ensure genuine support and respect for their contributions.

VI. Criticism of disclosure and informed consent

One of the study's important findings was that the information provided about the operation, future results, and side effects for obtaining consent was insufficient. The expressions used suggest that women lack detailed information on many important issues before undergoing surgery. For example, will the abdominal muscles be cut? Why is the gallbladder removed along with the liver? Are individuals genetically predisposed to kidney or liver diseases? Is there a risk of facing the same problem in the future? How does the process of liver regeneration occur? How many years can one live with a single kidney? What will be the duration of the kidney or liver remaining in the recipient? What are the other treatment alternatives? Which lobe of the liver, right kidney or left kidney will be taken? What are the possible complications during and after surgery? How should nutrition be managed after transplantation? Some questions remain unanswered. Some interviewees have so little information about the transplantation process that one interviewee's statement, "I entered the room and saw something written on the board. It said my name and surname and left kidney. So I learned from the board that my left kidney would be taken," clearly illustrates this situation. Could the validity of consent be questioned if given without adequate knowledge about the subject matter?[35]

It is challenging to gather complete information about the long-term effects on organ donors, as most donors are not followed up regularly after the first year once liver function stabilizes. However, recent studies

relation to an even more challenging issue, euthanasia, see Jose Luis Guerrero Quiñones, "Physicians' Role in Helping to Die," *Conatus – Journal of Philosophy* 7, no. 1 (2022): 79-101.

[35] Evangelos D. Protopapadakis, "Placebo: Deception and the Notion of Autonomy," in *Thinking in Action*, eds. Evangelos D. Protopapadakis and Georgios Arabatzis, 103-115 (The NKUA Applied Philosophy Research Lab Press, 2018).

have shown a potential link between liver damage and dementia in patients with dementia.[36] Therefore, I would like to raise the question of whether organ donors could also face a long-term risk of developing dementia. Post-operative cognitive dysfunction in living donors for liver transplantation is an area of research.[37] Studies have shown that complications such as pre-eclampsia, hypertension, and proteinuria (the presence of protein in the urine) may be more common among kidney donors. Pre-eclampsia is a serious condition that can cause high blood pressure and organ damage during pregnancy. Some reports show an increased risk of gestational hypertension and pre-eclampsia after kidney donation, based on a comparison of pre and post-donation pregnancies in donors.[38] Additionally, I would like to point out that there is no existing research on the risk of early menopause in living donors.

Due to the nature of positive science, it is necessary to act in the light of available information. Although the long-term health outcomes of donors are not known with current knowledge, donors must be thoroughly and accurately informed about known risks and potential consequences. Consequently, fully informing donors about known risks, uncertainties, and alternatives and conducting regular long-term follow-ups are critical to protecting donors' health and autonomy.[39] This action supports an ethical and safe organ transplantation process.

Feminist scholars emphasize the need to create spaces for dialogue and negotiation that enable women to exercise genuine autonomy over their bodies and healthcare decisions, free from societal pressures and gender inequalities. According to the information obtained from the participants in the study, it has been observed that there are different practices regarding organ transplantation in different hospitals. Worldwide, other studies are showing the existence of different applications.[40] Some hospitals are

---

[36] Scott Silvey et al.,"A Possible Reversible Cause of Cognitive Impairment: Undiagnosed Cirrhosis and Potential Hepatic Encephalopathy in Patients with Dementia," *The American Journal of Medicine* 137, no. 11 (2024): 1082-1087.

[37] Nizamettin Bucak et al., "Postoperative Cognitive Dysfunction in Living Liver Transplant Donors," *Experimental and Clinical Transplantation* 12, no. 1 (2014): 81-85.

[38] Anna Varberg Reisaeter et al., "Pregnancy and Birth After Kidney Donation: The Norwegian Experience," *American Journal of Transplantation* 9, no. 4 (2009): 820-824; Pratik B. Shah et al., "Preeclampsia Risks in Kidney Donors and Recipients," *Current Hypertension Reports* 20, no. 7 (2018): 59; Hassan N. Ibrahim et al., "Pregnancy Outcomes After Kidney Donation," *American Journal of Transplantation* 9, no. 4 (2009): 825-834.

[39] For a seminal discussion on informed consent, its scope and limitations, see Dejan Donev and Denko Skalovski, "Responsibility in the Time of Crisis," *Conatus – Journal of Philosophy* 8, no. 1 (2023): 87-109.

[40] Federica Avorio et al., "Neurological Screening in Elderly Liver Transplantation Candidates: A Single Center Experience," *Neurology International* 14, no. 1 (2022): 245-255.

noted to expedite the process very quickly, leaving insufficient time for potential donors to consider their decisions thoroughly, as inferred from the participants' opinions. Gözde states: "I never thought about myself because it happened so fast. I was surprised at my behavior at that moment. I tried to cheer up my mother and the people around me as if I was not going to go into surgery. I mean, I was telling and showing other good examples. I was not even scared at that stage. Because there must have been no opportunity." Even with the knowledge of the situation's urgency, it is difficult to understand why everything happened overnight. Leyla, 19 years old, gave her liver to her mother, who was in a coma. She states, "The psychological test I took was not very detailed. I think because of the urgency of the situation. A psychiatrist came to the room where I was lying and asked me how I was feeling. I said I was scared and worried. I remember it very clearly. He said it was normal for me to be scared and say such suggestive things. He signed and left anyway. It didn't even take five minutes." Although she expressed feelings of fear, the response was that this anxiety was normal. It gave the impression that it had not been addressed but rather was not taken seriously.

A further consequence of the interviews was that the women who asserted their ability to withdraw from donation at the last minute also reported the use of tranquillizers the previous night. Although the use of tranquillizers before surgery is a routine medical practice for donors, it is important to note that this may potentially impact their ability to express their concerns and withdraw from the donation. Our knowledge of why this was done is limited.

The study highlights significant deficiencies in the information provided to organ donors regarding the complexities and potential long-term impacts of organ transplantation procedures. Many critical questions remain unanswered, ranging from surgical details to post-operative health outcomes, raising concerns about the validity of consent without comprehensive knowledge. The uncertain long-term health effects on donors, such as the possible association with dementia, pregnancy, menopause and other conditions, underscore the need for rigorous and ongoing follow-up studies. Addressing these gaps is essential to ensuring ethical practices that protect donors' autonomy and well-being in organ transplantation.

VII. Altruism

Women have expressed a shared view that sacrifice is a gendered emotion, particularly emphasizing that autonomy has not been used in the sense of being completely free from everything. The most obvious reason for altruism in organ donation stems from women's repetition of socially con-

structed gender roles such as motherhood, self-sacrifice, and benevolent daughter. It can be explained by traditional male-female roles in which women feel obliged to take care of their sick family members. Mine states, "I think women give more organs in the world. I have a lot of friends I know. I have 3-4 friends who are liver donors. All of them are women. I don't know; I think it may be the character. They feel more sad if they lose their loved ones. I think men are less sad. Women are more sacrificing."

It is considered "normal" for women to donate an organ from their own body, and it is often a part of their caregiving and fertility roles. It is associated with the expectation of social sacrifice. The devotion and sacrifice expected from women cause them to regard this action as "normal." As Zeiler notes, this is also a normalization of bodily exchanges in medicine.[41] Nevertheless, when considering altruism, it can be argued that the donation of an organ from one's own body to another individual represents the pinnacle of selfless acts. I do not intend to suggest that the action in question was undertaken solely for benevolence. Women have placed themselves at risk for those they love or do not love and have undergone this operation. This decision is a voluntary and altruistic action resulting from calculations and reckonings made by women in their inner worlds. But this altruistic behavior can be seen as a bargain to free them from all contracted responsibilities. In the absence of greater environmental pressures and obligations, it is possible that the decision would have been reached differently.

Women were aware of their oppression and marginalization. Figen states: "It may be because women have a softer temperament. Alternatively, they can give up things more easily. Give an example from the crisis. Women are so accustomed to being the first to go out of favour that men cannot easily give up their jobs. Therefore, women may be approaching things conscientiously because they know this. Is this a good thing? It depends on the place. If she is going to be a mother, yes, she should be. However, if it is a professional job, no. If you make a conscientious decision in that environment, you are not taken seriously, your opinions are questioned." Living organ donation is considered an extension of social expectations, and the act of donation, in parallel with the motif of sacrificial motherhood, is highly selfless, and self-sacrificing is defined as a form of behavior. Therefore, this situation is considered a woman's innately altruistic behavior, naturalized and normalized by their tendency to be.

[41] Kristin Zeiler, "A Phenomenological Approach to the Ethics of Transplantation Medicine: Sociality and Sharing When Living-With and Dying-With Others," *Theoretical Medicine and Bioethics* 35 (2014): 369-388.

VIII. Conclusion

The analysis identified key factors shaping *autonomy*, which in the context of living organ donation refers to the individual's right to make decisions about their own body and health, including family relationships, paternalism, criticism of disclosure and informed consent, and altruism. Family relationships emerged as pivotal, with constructed expectations within family dynamics significantly influencing donors' decisions. Donation was expressed as a concept that must be endured due to necessity, not a process decided by free will. In the context of live organ donation, women perceive the moral dilemma of saving someone versus allowing oneself to die as a matter of care and responsibility rather than solely as a right to not harm oneself. Being a live organ donor also goes beyond the legitimized discourse of altruism; it involves expectations, bargaining, and negotiations. By making sacrifices, women expect recognition and respect.

Societal norms, particularly those related to motherhood, create a sense of obligation and support. Mothers, sisters, and wives often feel compelled to donate due to the societal expectation of selflessness. These expectations shape donors' sense of duty and acceptance within their families, highlighting the complex interplay between personal choice and societal pressure.

Paternalistic attitudes in healthcare settings often stifle donors' decision-making agencies. Medical professionals, assuming a directive role, often overshadow the donors' autonomy, leading to a power imbalance in decision-making. This dynamic underscores the pressing need for a more balanced and respectful approach that genuinely considers the donor's perspective and autonomy. The urgency of this shift towards more patient-centric care is evident, as it can significantly improve the donor's experience and decision-making process.

Criticism of disclosure emerged as another significant factor. Donors often find themselves at a disadvantage in understanding the full implications of their decisions, needing help to navigate complex medical information. This criticism underscores the crucial role of effective communication between donors and medical practitioners. Both parties need a comprehensive understanding of each other's perspectives to bridge this gap, which can lead to increased uncertainty and stress for the donors, further complicating their decision-making process.

Altruism, intertwined with personal sacrifice, moral duty, and the normalization of the donation process, played a crucial role. This sense of altruism is driven by cultural norms and expectations of caregiving

and nurturing, often overshadowing individual autonomy. The pressure to conform to these norms can lead donors to prioritize the needs of others over their well-being, reflecting broader societal values that valorize self-sacrifice, particularly among women.

This study comprehensively examines the gendered dimensions of autonomy in living organ donation in Turkey. Integrating feminist theory and the concept of autonomy highlights the nuanced experiences of women donors and underscores the need for a deeper understanding of how gender inequalities shape medical practices. The findings reveal significant gender disparities in donor availability, access to services, and post-transplant care responsibilities, all influenced by cultural norms and gender dynamics. These disparities are often perpetuated by 'medical biases,' which refer to the systematic favouritism or discrimination towards certain groups in medical practices, in this case, women donors.

Furthermore, this study contributes to feminist autonomy theories by providing empirical insights into women's lived experiences and interpretations of autonomy in the context of living organ transplantation. These insights include specific instances where women donors felt their autonomy was compromised or respected and how they navigated the societal and familial pressures. By including women's voices and experiences, this research challenges the tendencies that objectify the body and offers a critical perspective on the intersection of medical practices and societal norms. The insights gained from this study underscore the importance of considering gender and autonomy in discussions of organ transplantation, advocating for more equitable and inclusive approaches in healthcare.

This study offers a significant contribution to the understanding of gendered autonomy in living organ donation. It reveals the intricate ways in which societal norms, familial expectations, and medical practices intersect to shape women's experiences and decisions. By highlighting these dynamics, the study calls for a reevaluation of consent procedures and a move towards more inclusive, respectful, and gender-sensitive practices in organ transplantation. The findings emphasize the necessity of addressing gender biases and ensuring that women's autonomy is respected and supported in medical contexts.

I would like to express our deepest gratitude to the women donors who shared their meaningful stories, forming the most valuable parts of the fieldwork. Their life stories have significantly contributed to this project, adding true meaning and depth to this research. Each of them has made an important step for humanity, positively impacting the lives

of others. By being at the heart of this research, they have greatly contributed to raising awareness about organ transplantation. Their participation has not only enriched this study but also contributed to the broader awareness of organ donation. I hope this work will shed light on the lives of future donors and those awaiting transplants. I extend my sincerest thanks to them and salute the women who shared this unique experience.

References

Andorno, Roberto, and George Boutlas. "Global Bioethics in the Post-Coronavirus Era: A Discussion with Roberto Andorno." *Conatus – Journal of Philosophy* 7, no. 1 (2022): 185-200.

Ariès, Philippe. *The Hour of Our Death*. Oxford University Press, 1981.

Avorio, Federica, Gianvincenzo Sparacia, Giovanna Russelli, Aurelio Seidita, Giuseppe Mamone, Rossella Alduino, Fabio Tuzzolino, Salvatore Gruttadauria, Roberto Miraglia, Matteo Bulati, and Vincenzina Lo Re. "Neurological Screening in Elderly Liver Transplantation Candidates: A Single Center Experience." *Neurology International* 14, no. 1 (2022): 245-255.

Beauvoir, Simone de. *The Second Sex*. Vintage Classics, 2015.

Bucak, Nizamettin, Zekine Begeç, Feray Erdil, Hüseyin İlksen Toprak, Birgül Elbozan Cumurcu, Yasemin Demirtaş, Saim Yoloğlu, Mahmut Durmuş, and Mehmet Özcan Ersoy. "Postoperative Cognitive Dysfunction in Living Liver Transplant Donors." *Experimental and Clinical Transplantation* 12, no. 1 (2014): 81-85.

Chauhan, Sanshriti, Vasanthi Ramesh, and Chaitali Pal. "Nationwide Data on Gender Disparity in Solid Organ Transplantation for India in the Pre-pandemic and Pandemic Era." *Transplantation* 6, no. 9 (2022): 230.

Chodorow, Nancy. *The Reproduction of Mothering: Psychoanalysis and the Sociology of Gender*. University of California Press, 1999.

Crowley-Matoka, Megan. *Domesticating Organ Transplant: Familial Sacrifice and National Aspiration in Mexico*. Duke University Press, 2016.

Darden, Michael, Geoff Parker, Edward Anderson, and Joseph F. Buell. "Persistent Sex Disparity in Liver Transplantation Rates." *Surgery* 169, no. 3 (2021): 694-699.

Donev, Dejan, and Denko Skalovski. "Responsibility in the Time of Crisis." *Conatus – Journal of Philosophy* 8, no. 1 (2023): 87-109.

Ellner, Andrea. "Ethics of Conflict, Violence and Peace – Just War and a Feminist Ethic of Care." *Conatus – Journal of Philosophy* 8, no. 2 (2023), 147-173.

Gilligan, Carol. *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press, 1993.

Gompers, Annika, Ana Rossi, and Jessica L. Harding. "Intersectional Race and Gender Disparities in Kidney Transplant Access in the United States: A Scoping Review." *BMC Nephrol* 25, no 1. (2024): 36.

Ibrahim, Hassan N., Sanjeev K. Akkina, Erin Leister, Kristen J. Gillingham, Gretchen K. Cordner, Hongfei Guo, Robert F. Bailey, Tyson B. Rogers, and Arthur J. Matas. "Pregnancy Outcomes After Kidney Donation." *American Journal of Transplantation* 9, no. 4 (2009): 825-834.

Ioannidis, Iraklis. "Shackling the Poor, or Effective Altruism: A Critique of the Philosophical Foundation of Effective Altruism." *Conatus – Journal of Philosophy* 5, no. 2 (2020): 25-46.

Kaludjerovic, Zeljko. "Bioethics and Hereditary Genetic Modifications." *Conatus – Journal of Philosophy* 3, no. 1 (2019): 31-44.

Karnam, Ravikiran S., Shiyi Chen, Wei Xu, Catherine Chen, Praniya Elangainesan, Anand Ghanekar, Ian McGilvray, Trevor Reichman, Blayne Sayed, Markus Selzner, Gonzalo Sapisochin, Zita Galvin, Gideon Hirschfield, Sumeet K. Asrani, Nazia Selzner, Mark Cattral, Leslie Lilly, and Mamatha Bhat. "Sex Disparity in Liver Transplant and Access to Living Donation." *JAMA Surgery* 156, no. 11 (2021): 1010-1017.

Keven, K., and S. Aktürk. *Transplantasyona Hazirlik Verici. A. Türkmen (Dü.) Inside, Transplantasyon Nefrolojisi Pratik Uygulama Önerileri (S. 9-25)*. Türk Nefroloji Derneği, 2016.

Kurnikowski, Amelie, Simon Krenn, Michal J. Lewandowski, Elisabeth Schwaiger, Allison Tong, Kitty J. Jager, Juan Jesus Carrero, Manfred Hecking, and Sebastian Hödlmoser. "Country-specific Sex Disparities in Living Kidney Donation." *Nephrology, Dialysis, Transplantation* 37, no. 3 (2022): 595-598.

Kute, Vivek, Sanshriti Chauhan, and Hari Shankar Meshram. "Act Together and Act Now to Overcome Gender Disparity in Organ Transplantation." *Experimental and Clinical Transplantation* 22, no. 1 (2024): 17-27.

Lin, Ya-Ping. "Visible Body, Invisible Care: Family, Gender Politics, and the Female Caregiver in Living Donor Liver Transplantation in Taiwan." *SSM – Qualitative Research in Health* 4 (2023): 100346.

McNeill, Paul M. "Research Ethics Review and the Bureaucracy." *Monash Bioethics Review* 21, no. 3 (2002): 72-73.

Melk, Anette, Bernhard M. W. Schmidt, Siegfried Geyer, and Jelena Epping. "Sex Disparities in Dialysis Initiation, Access to Waitlist, Transplantation and Transplant Outcome in German Patients With Renal Disease – A Population-Based Analysis." *PLoS ONE* 15, no. 11 (2020): e0241556.

Metin, Sevtap. *Biyo-Tıp Etiği ve Hukuk*. Betim Kitaplığı, 2019.

Mongoven, Ann. "Sharing Our Body and Blood: Organ Donation and Feminist Critiques of Sacrifice." *The Journal of Medicine and Philosophy* 28, no. 1 (2003): 89-114.

Noddings, Nel. *Starting at Home: Caring and Social Policy*. University of California Press, 2002.

Øien, Cecilia M., Anna Varberg Reisæter, Torbjørn Leivestad, Per Pfeffer, Per Fauchald, and Ingrid Os. "Gender Imbalance among Donors in Living Kidney Transplantation: The Norwegian Experience." *Nephrology Dialysis Transplantation* 20, no. 4 (2005): 783789.

Osuch, Janet R., Kami Silk, Carole Price, Janice Barlow, Karen Miller, and Ann Hernick, and Ann Fonfa. "A Historical Perspective on Breast Cancer Activism in the United States: From Education and Support to Partnership in Scientific Research." *Journal of Women's Health* 21, no. 3 (2012): 355-362.

Peracha, Javeria, Manvir Kaur Hayer, and Adnan Sharif. "Gender Disparity in Living-Donor Kidney Transplant Among Minority Ethnic Groups." *Experimental and Clinical Transplantation* 14, no. 2 (2016): 139-145.

Protopapadakis, Evangelos D. "Placebo: Deception and the Notion of Autonomy." In *Thinking in Action*, edited by Evangelos D. Protopapadakis and Georgios Arabatzis, 103-115. The NKUA Applied Philosophy Research Lab Press, 2018.

Puoti, Francesca, Andrea Ricci, Alessandro Nanni-Costa, Walter Ricciardi, Walter Malorni, and Elena Ortona. "Organ Transplantation and Gender Differences: A Paradigmatic Example of Intertwining Between Biological and Sociocultural Determinants." *Biology of Sex Differences* 7, no. 1 (2016): 35.

Quiñones, Jose Luis Guerrero. "Physicians' Role in Helping to Die." *Conatus – Journal of Philosophy* 7, no. 1 (2022): 79-101.

Reisaeter, Anna Varberg, Jo Røislien, Tore H. Henriksen, Lorentz M. Irgens, and Anders O. Hartmann. "Pregnancy and Birth After Kidney Donation: The Norwegian Experience." *American Journal of Transplantation* 9, no. 4 (2009): 820-824.

Rock, Charles N. "The Living Dead: Anencephaly and Organ Donation." *NYLS Journal of Human Rights* 7, no. 1 (1989): 243-277.

Rota-Musoll, Laura, Serena Brigidi, Esmeralda Molina-Robles, Ester Oriol-Vila, Laureano Perez-Oller, and Mireia Subirana-Casacuberta. "An Intersectional Gender Analysis in Kidney Transplantation: Women Who Donate a Kidney." *BMC Nephrology* 22, no. 1 (2021): 59-69.

Rubin, Jessica B., Marie Sinclair, Robert S. Rahimi, Elliot B. Tapper, and Jennifer C. Lai. "Organ Transplantation and Gender Differences: A Paradigmatic Example of Intertwining Between Biological and Sociocultural Determinants." *World Journal of Gastroenterology* 25, no. 8 (2019): 980-988.

Rupčić Kelam, Darija, and Ivica Kelam. "Care and Empathy as a Crucial Quality for Social Change." *Conatus – Journal of Philosophy* 7, no. 2 (2022): 157-172.

Savulescu, Julian, and Evangelos D. Protopapadakis. "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu." *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

Shah, Pratik B., Manpreet Samra, and Michelle A. Josephson. "Preeclampsia Risks in Kidney Donors and Recipients." *Current Hypertension Reports* 20, no. 7 (2018): 59.

Sherwin, Susan. *No Longer Patient Feminist Ethics and Health Care*. Temple University Press, 1992.

Silvey, Scott, Richard K. Sterling, Evan French, Michael Godschalk, Angela Gentili, Nilang Patel, and Jasmohan S. Bajaj. "A Possible Reversible Cause of Cognitive Impairment: Undiagnosed Cirrhosis and Potential Hepatic Encephalopathy in Patients with Dementia." *The American Journal of Medicine* 137, no. 11 (2024): 1082-1087.

Stoljar, Natalie. "Feminist Perspectives on Autonomy." *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), edited by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/sum2024/entries/feminism-autonomy/.

Tong, Rosemaria. *Feminist Approaches to Bioethics*. Routledge, 1997.

Trotter, James F., Margaret Hill-Callahan, Brenda Gillespie, Carrie Nielsen, Sammy Saab, Shrestha Roshan, Michael Talamantes, and Robert M. Weinrieb. "Severe Psychiatric Problems in Right Hepatic Lobe Donors for Living Donor Liver Transplantation." *Transplantation* 83, no. 11 (2007): 1506-1508.

Van Vo, Dung. "Four Important Characteristics of Women in Confucianism and Its Contribution to the Implementation of Gender Equality in Vietnam." *Conatus – Journal of Philosophy* 9, no. 2 (2024): 283-302.

Veltman, Andrea, and Mark Piper, eds. *Autonomy, Oppression, and Gender*, *Studies in Feminist Philosophy*. Oxford Academic, 2014.

Young, Iris Marion. *On Female Body Experience: Throwing Like a Girl and Other Essays*. Oxford University Press, 2005.

Zeiler, Kristin. "A Phenomenological Approach to the Ethics of Transplantation Medicine: Sociality and Sharing When Living-With and Dying-With Others." *Theoretical Medicine and Bioethics* 35 (2014): 369-388.

Zeiler, Kristin. "Just Love in Live Organ Donation." *Medicine, Health Care and Philosophy* 12, no. 3 (2009): 323-331.

O. A. Gabrielyan & I. E. Suleimenov · *Conatus* 10, no. 1 (2025): 111-125
doi: https://doi.org/10.12681/cjp.33973

Objective Foundations of Ethics and Prospects for Its Development: Information and Communication Approach

Oleg Arshavirovich Gabrielyan
V. I. Vernadsky Crimean Federal University
E-mail address: gabroleg@mail.ru
ORCID iD: https://orcid.org/0000-0003-0302-0229

Ibragim Esenovich Suleimenov
National Engineering Academy, Kazakhstan
E-mail address: esenych@yandex.ru
ORCID iD: https://orcid.org/0000-0002-7274-029X

Abstract

The paper considers the objective foundations of ethics as a system providing self-regulation of society. An information and communication approach to the study of the laws of the formation of ethical norms and the interpretation of ethics as a specific transpersonal information structure is proposed. It is shown that the ethical system can be analyzed on the basis of a neural network model of society, which proves the possibility of the emergence of transpersonal information objects of various types, including those responsible for the functioning of ethics. Moreover, the thesis is put forward and substantiated that human evolution will take place in the direction of the formation of ethics as a third signaling system. There are no other regulators of socio-natural human behavior. In this direction, humanity will need to reach the level of macrocivilizational subjects and later to the level of self-identification as a human civilization with a single ethical system or coordinated at the level of intercivilizational integration.

Keywords: *ethical system; neural network model; transpersonal structure; model; signal system; information and communication environment*

I. An ethical system and the prospects for its objective research

The problems of ethics have attracted the closest attention of philosophers, moreover, in modern conditions, in connection with the transformations of the world order, they acquire a new sound. On the other hand, challenges to the very nature of humans acquire an essential character. In particular, this applies to the problems of transhumanism,[1] immortalism,[2] gender identity,[3] etc. The emerging problems are fundamental, and the task of establishing objective patterns that determine the emergence and evolution of ethical systems, as well as the task of formalizing such patterns, is very acute.

An ethical system develops in the process of interaction of people in society. Like any of the natural languages, it ontologically functions in the information and communicative social environment as a system integrity.

An ethical system means a system of historically determined moral norms and principles that have been formed in a particular society. In society, they regulate relations between people, although ethical norms are not always codified. It is obvious that these principles have quite objective grounds, although they are formed in the process of social interaction of specific people (various subjects) for a long time. It is in this general and simplified sense that we will use the concept of "ethical system" further in the article. As a generalization, it is quite theoretical, but as an objective entity it has a practical and applied character. The specification of the varieties of ethical systems in line with the research conducted in this article seems unjustified, since the very foundations of the emergence and existence of ethics as such are considered. This, among other things, determines the validity of using the term "ethical system" as the one with an appropriate degree of generality.

Ethics as a philosophical reflection considered, first of all, the following fundamental existential problems: the criteria of good and evil; the meaning of life and the purpose of humankind; freedom of will; due and its relationship with the natural desire for happiness.

---

[1] Egidijus Juozelis, "Religious Dimensions in Transhumanist and Posthumanist Philosophies of Science," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 125-133.

[2] Donovan Van der Haak, "Death Anxiety, Immortality Projects and Happiness: A Utilitarian Argument Against the Legalization of Euthanasia," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 159-174. See also Akhat Bakirov et al., "To the Question of the Practical Implementation of 'Digital Immortality' Technologies: New Approaches to the Creation of AI," in *Proceedings of the Future Technologies Conference (FTC) 2022, Volume 1*, ed. Kohei Arai, vol. 559, *Lecture Notes in Networks and Systems* (2023).

[3] Katharine Jenkins, "Amelioration and Inclusion: Gender Identity and the Concept of Woman," *Ethics* 126, no. 2 (2016): 394-421. Also, Georgios Tsitas and Athanasios Verdis, "Proposing a Frame of Ethical Principles for Educational Evaluation in Modern Greece," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 135.

Ethics as a set of unwritten rules is designed to resolve conflicts and problems of various levels that arise permanently in society. This is the main significance of the ethical system. Its flexibility is the main advantage in social relations. Some of its rules may eventually be fixed in legal codes. Here they become clearer in definition and execution, but at the same time they lose the flexibility that is necessary in the culture of social communication.

The noted features of the ethical system are very similar to the natural language of communication. The same ontological information and communication status, the same flexibility in everyday communication, a similar fixation in grammar.

These systems, like many others, are transpersonal structures (in relation to science as a transpersonal information structure, this term is disclosed in)[4] although its direct carriers are specific members of society, or rather, society as a system integrity.

The first interesting results of the formalization of the ethical system were obtained by V. A. Lefebvre. In particular, this allowed him to identify two different ethical systems that different peoples adhere to. In one system, the union of good and evil was evaluated as evil; in another system, the union of good and evil was evaluated as good.[5]

At present, tools for the formal description of the ethical system are proposed, allowing for a subsequent transition to quantitative research. These tools are based, in particular, on the consideration of society as a neural network.[6]

II. Ethics as a transpersonal information and communication subsystem of the noosphere

Any communication between individuals physically resolves into the exchange of signals between the neurons of their brain, connected with sensory organs. As follows from the modern theory of neural networks,[7] the memory of neural networks depends non-linearly on the number of elements included in it. In the process of communication between individuals, a com-

---

[4] Ibragim E. Suleimenov et al., "Dialectics of Scientific Revolutions from the Point of View of Innovations Theory," *WISDOM* 24, no. 4 (2022): 25-35.

[5] Vladimir A. Lefebvre, *Algebra of Conscience: Second, Revised Edition* (Reidel, 1982; expanded 2nd ed., Kluwer, 2001).

[6] Ibragim E. Suleimenov et al., "Artificial Intelligence: What Is It?" in *Proceedings of the 2020 6th International Conference on Computer and Technology Applications (ICCTA 2020)* (Association for Computing Machinery, 2020), 22-25; Ibragim E. Suleimenov et al., "Neural Networks and the Philosophy of Dialectical Positivism," *MATEC Web of Conferences* 214, no. 02002 (2018).

[7] Ibragim E. Suleimenov et al., "Distributed Memory of Neural Networks and the Problem of the Intelligence's Essence," *Bulletin of Electrical Engineering and Informatics* 11, no. 1 (2022): 510-520.

mon neural network arises, the memory of which exceeds the total memory of these individuals taken separately. Consequently, an additional segment of the information space arises, which is only indirectly connected with the memory of individuals. Moreover, the exchange of signals between neurons in a common neural network leads to the appearance of non-rivial information objects – transpersonal information structures. The mechanism of their appearance is completely analogous to the appearance of human consciousness, for instance, it is also associated with the exchange of signals between neurons, but since such an exchange takes place within a common network, the resulting information structures relate precisely to the transpersonal level of information processing. It is appropriate to emphasize that the conclusions of the theory of transpersonal information objects can be made very clear. For example, such a system as a university can also be considered on the basis of a direct analogy with a neural network,[8] in it is shown that any voting Council with a sufficient density of horizontal connections is converted into a direct analogue of Hopfield neuroprocessor, etc. It should also be noted that the approach correlating with the one proposed in the works cited above is also reflected in the current literature.[9]

The conclusion about the existence of a transpersonal level of information processing also clearly correlates with modern trends in the application of neural network theory in the natural sciences at the level where they are adjacent to philosophy. Thus, in it was demonstrated that the Universe as a whole can also be considered as a neural network, which is completely consistent with the understanding of the "complex" proposed in.[10]

The main idea of our statement is not only to fix the transpersonal nature of the ethical system, but also to prove the possibility of its scientific research on an interdisciplinary basis, for instance, in parallel by means of philosophy and information theory.

The ethical system, the essence of which is revealed through the theory of transpersonal information objects, is an important (framework) component

---

[8] Ibragim E. Suleimenov et al., "University as an Analogue of the Neural Network," *E3S Web of Conferences* 258, no. 07056 (2021): 07056; Ibragim E. Suleimenov et al., "Voting Procedures from the Perspective of Theory of Neural Networks," *Open Engineering* 6, no. 1 (2016): 318-321.

[9] George A. Mashour et al., "Conscious Processing and the Global Neuronal Workspace Hypothesis," *Neuron* 105, no. 5 (2020): 776-798; Panagiotis Kormas et al., "Implications of Neuroplasticity to the Philosophical Debate of Free Will and Determinism," in *Handbook of Computational Neurodegeneration*, eds. P. Vlamos, I. S. Kotsireas, and I. Tarnanas, 1-19 (Springer International Publishing, 2022).

[10] Yelizaveta S. Vitulyova et al., "Interpretation of the Category of 'Complex' in Terms of Dialectical Positivism," in *IOP Conference Series: Materials Science and Engineering* 946, no. 1 (2020): 012004; Vitaly Vanchurin, "The World as a Neural Network," *Entropy* 22, no. 11 (2020): 1210.

of the emerging noospheric reality, which can be characterized as an information and communication environment. Otherwise, it can be argued that at the present stage the noosphere has already been converted into a human-machine system, an important component of which is telecommunications networks.

The new quality of the noosphere leads to the accelerated formation of new transpersonal information objects[11] and today it is more than important to make this process manageable, which returns to the basic thesis of this work – the consideration of ethics as one of the transpersonal information structures that perform the functions of a regulator of social behavior.[12]

Remembering the hopes of V. I. Vernadsky, which he pinned on the mind of mankind, today we can talk about these ideas of an outstanding naturalist in a different perspective. As an educator and humanist V. I. Vernadsky believed in the rational nature of humans, which should change the ethics of their behavior and, as a consequence, the social structure of the world.

> The idea of a single state unification of all mankind is becoming a reality only in our time, and that, obviously, is becoming only a real ideal, the possibility of which cannot be doubted. It is clear that the creation of such unity is a necessary condition for the organization of the noosphere, and humanity will inevitably come to it.[13]

This, in fact, belief in the idea of progress of great Enlightenment figures, unfortunately, did not justify itself. It burned down in two world wars, when the scientific and technological achievements of mankind were used to destroy millions of people, as well as in numerous conflicts of various kinds, which continue at the present time.

However, qualitative changes in the noosphere inspire certain hopes – if, of course, it adequately disposes of the resource that a consistent theory of transpersonal information structures creates.

III. Ethics in the context of the coevolution of human and nature

There are objective reasons for this. Namely, there is every reason to believe that the ethical evolution of humans is not the result of random processes.

---

[11] Almaz S. Bakirov et al., "Internet Users' Behavior from the Standpoint of the Neural Network Theory of Society: Prerequisites for the Meta-Education Concept Formation," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVI-4/W5-2021 (2021): 83-90.

[12] See Julian Savulescu and Evangelos D. Protopapadakis, "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu," *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

[13] Vladimir I. Vernadsky, *The Biosphere and the Noosphere* (Airis-Press, 2004), 324.

This is a natural consequence of objective laws reflecting the peculiarities of the coevolution of man and nature, in particular, the shells of the Earth. The first attempts to establish the laws of coevolution of living and inert matter were made by J. Lovelock, who put forward the concept of Gaia as a kind of "superorganism" in which biota plays the role of one of the regulators. Discussions around Lovelock's concept continue to this day[14] and the question of whether Gaia can possess certain forms of consciousness is also debated.[15]

The utilization of the theory of transpersonal informational structures and neural network theory to analyze the interaction between society and dormant matter has enabled us to understand the regulatory functions of Gaia's rational component at a higher level.

It is crucial to note that both Vernadsky and Lovelock assert that human civilization, as the intelligent component of Gaia, is a transformative force for the Earth spheres. Vernadsky likened the outcomes of human activity to geological disturbances on a smaller scale. Consequently, the concept of co-evolution implies that humans exercise regulatory functions on a global scale, which necessitates a different perspective on the transpersonal information structures that regulate human social behavior.

As more and more powerful and effective technical means appear at a person's disposal, ethics becomes more and more important, and already as a factor of coevolution of living and inert matter. Continuing Vernadsky's logic, it can be argued that the transformations of ethics become events of a planetary scale.

In order to reveal this thesis consistently, the stages of human evolution in the context of transformations of its interdependence with nature are briefly considered.

According to this criterion, human evolution has undergone (or is undergoing to one degree or another) five "optimizations" ("gastronomic," "mechanical," "demographic," "mnemotic," "gerontological,") each of which can be considered from the standpoint of coevolution, understood by Lovelock, and with positions of transformations of transpersonal information structures.

The first ("gastronomic") stage was associated with the development of fire. The possibility of cooking food subjected to heat treatment can be interpreted as a physiological optimization, in which part of the functions performed by the gastrointestinal tract was "outsourced" which led, among other things, to the evolution of the jaw and intestinal tract).

---

[14] Jaime A. T. Da Silva and Panagiotis Tsigaris, "The Relevance of James Lovelock's Research and Philosophy to Environmental Science and Academia," *Frontiers of Environmental Science & Engineering* 17, no. 3 (2023): 1-2; Tim Radford, "James Lovelock at 100: The Gaia Saga Continues," *Nature* 570, no. 7762 (2019): 441-443.

[15] Dorion Sagan, "James Lovelock and Consciousness: An Obituary," *Journal of Consciousness Studies* 29, nos. 11-12 (2022): 226-231.

From the point of view of the theory of transpersonal information structures, this stage is associated with the formation of ancient mythology, which surprisingly reproduces the same plots (despite the fact that the peoples who created this mythology were often significantly geographically dispersed).[16] A classic in this respect is the myth of Prometheus, who gave fire to people. Transpersonal information structures, which are "read" according to ancient myths reflecting this historical period, act as a kind of active principle with which a person – or his "representatives", such as Prometheus – can enter into real interaction. This translates into the emergence of innovations that initiated the further development of civilization (potter's wheel, sail, sustainable fire production systems, etc.).

The second ("mechanical") stage is associated with the optimization of muscle work. The domestication of animals, and most importantly, the creation of mechanisms for the use of external natural forces, can be interpreted as the transfer of a significant part of heavy physical labor "to outsourcing." In this historical period, a new specific form of transpersonal information structures associated with abstract thinking appears. Without moving to such a level of understanding of physical reality, it is impossible to formulate the laws of mechanics even at the most primitive level. It is not by chance that mechanics as a science was born much later than geometry. Humanity first had to generate and perceive the corresponding transpersonal information objects.

We emphasize that abstract thinking is obviously not inherent in cultures that are even at the highest stages of barbarism, this is an achievement of civilized peoples. Even the simplest forms of logical thinking are not characteristic of primitive cultures.[17] At the same time, abstract thinking cannot appear – much less develop – as something individual. An appropriate environment is needed. Hence the thesis that abstract thinking, even of an individual, can exist only as a transpersonal information structure and is realized through its projection onto the consciousness of an individual.[18]

The third ("demographic") stage is expressed in the optimization of population reproduction. Unlike the first two, it cannot be considered already completed, however, a demographic transition has obviously occurred in developed countries, reflected by official statistics: low birth rate with low mortality, including infant mortality.

This transition is also obviously associated with transformations of transpersonal information structures. With high infant mortality, the only way to preserve the population is to ensure a high birth rate.

---

[16] David J. Wong-MingJi et al., "Cross-Cultural Comparison of Cultural Mythologies and Leadership Patterns," *South Asian Journal of Global Business Research* 3, no. 1 (2014): 79-101.

[17] Lucien Lévy-Bruhl, *Le surnaturel et la nature dans la mentalité primitive* (Presses Universitaires de France, 1963).

[18] Bakirov et al., "Internet Users' Behavior," 83-90.

The desire to ensure a high birth rate in one way or another permeates any religion or ideology formed before the 20th century. The ban on abortions, the subordinate position of a woman, whose main function in the eyes of society was childbearing, categorical rejection of any other forms of marital relations – all this is a reflection of the fact that the transpersonal information structures formed before the 20th century were focused precisely on the quantitative reproduction of the population.

The roots of such transpersonal information structures go back to the times when they were formed on the basis of individual clans or tribes. For that historical epoch, the value of an individual's life was negligible compared to procreation.

Currently, the situation is changing dramatically. Modern medical achievements allow us to interpret placental pregnancy, breastfeeding, etc. as a kind of biological prejudice, or rather, a social atavism. Reproduction of the population is moving from quantitative to qualitative, which cannot but cause the transformation of those transpersonal structures that are responsible for the reproduction of the population.

The following two stages are considered ("mnemotic" and "gerontological") partly implemented, partly manifest themselves in the form of emerging trends.

The prerequisites for the implementation of the first of them are obvious. The simplest forms of intellectual (more precisely, not related to physical labor) activity can already be transferred to digital technologies, which is partly implemented in practice through the creation of appropriate artificial intelligence systems. There is no longer any doubt that the very existence of "office plankton," for instance, small clerks engaged in obviously useless work are determined only by the imperfection of legislation and problems easily solved by information logistics. Nevertheless, it is the "office plankton" that largely determines the structure and content of the modern information space. People who really have nothing to do at work spend time on online social networks, forming relevant requests and an information agenda.

Any significant economic crisis, however, will certainly lead to the fact that society will easily sacrifice this ballast. Accordingly, a change in the social fabric is predicted, and, consequently, the accompanying transpersonal information structures.

The least obvious is the content of the "gerontological" stage of evolution. On the one hand, the thesis of ensuring "digital immortality" sounds more and more clearly, and, as shown, for example, in this issue is partly solved already at the existing level of information technology development.[19]

However, even if we exclude the question of a drastic increase in life expectancy (which is quite expected, even if we do not take into account

[19] Bakirov et al., "Practical Implementation," 25.

the achievements of nanotechnology), the "gerontological" revolution in the formation of transpersonal information structures has already largely occurred. Here is the evidence.

IV. Modern consequences of coevolution and possible prospects

Transpersonal information structures cannot but be visualized (or personified) to one degree or another, albeit with noticeable distortions. Otherwise, they will remain not perceived by the ordinary consciousness. Such structures in the era of the Neolithic revolution were perceived through the images of Ancient Gods, totems of the genus, etc.

In Modern Times, such structures could not but transform, as well as their personification. The "avatar" of atavistic tribal structures became the head of the family, personifying and defending the corresponding set of values, and, consequently, the corresponding set of informal ethical norms.

Visualizing (or personifying) transpersonal information structures to some extent is essential, as without it, these structures would be imperceptible to ordinary consciousness. In the Neolithic era, these structures were perceived through the images of Ancient Gods, clan totems, and similar visual representations.

In modern times, the personification of transpersonal information structures has undergone significant changes. In the past, the head of the family represented the avatar of tribal structures and upheld a specific set of values and ethical norms.

Until the mid-20th century, the corresponding transpersonal information structures were relatively stable, which was determined by the nature of the age pyramid reflected in numerous studies on demography.[20]

So, at the beginning of the 20th century, the age pyramid of Belgium had the appearance characteristic of a patriarchal society. The population belonging to the older age groups was significantly lower than the corresponding indicator for groups aged about 30 years.

By the end of the 20th century, the situation in the EU countries had changed drastically. The number of different age groups in the age range from 0 to 80 years remains approximately the same.

In such conditions, there is simply no place for the "heads of the family." Indeed, as observations of the customs of societies in which patriarchal traditions are strong show, the "aqsaqal" was valuable for society, first of all, as a source of information (knowledge, life experience, etc.). This attitude towards the older generation also supported the existence of appropriate transpersonal information structures.

---

[20] John R. Gillis, *Youth and History: Tradition and Change in European Age Relations, 1770 – Present* (Elsevier, 2013).

This role of "aqsaqal" is now lost. Moreover, it is becoming increasingly obvious that the classical monogamous family no longer meets the needs of society, if only due to factors related to the transformations of the return pyramid and an increase in life expectancy.

Indeed, there are a very significant number of formats of activity (and critically important for the existence of civilization) that require the acquisition of high qualifications, and, consequently, a very long period of study. So, in such an area as medicine or science, a person becomes wealthy on average by the age of 35. Biologically, this corresponds to the age of a grandfather or grandmother: individuals of our species become reproductive by the age of 17.

In fact, this means that the classical monogamous family has already been significantly transformed. A young couple, for example, focused on an academic career, will not survive without the support of the older generation. If we take into account that even in those countries where traditions are strong, tribal relations are becoming more and more weakened, this means that we are dealing with a three-age family. Its peculiarity, unlike classical patriarchal families, is that not one, but two generations are in the position of children requiring care and guardianship.

The considerations expressed, however, are mainly illustrative. They are designed to emphasize that ethics is a flexible transpersonal structure that arises and develops not by itself, but as a coevolutionary component of the biosphere.

It is not difficult to prove this statement in modern conditions. Ethical norms that correspond to the era of maximum stimulation of childbearing are now a thing of the past. It is not the quantity that becomes important, but the quality of the human resource. Accordingly, ethical norms are also changing dramatically, which is clearly visible on the homosexual agenda.

However, such transformations – in accordance with the ideas of coevolution – cannot but correlate with the evolution of the Earth's spheres. Humans have become a factor of a planetary scale, accordingly, the norms (both formal and informal) that set the vector of development of society also take a planetary scale.

Summing up, even a brief overview of the stages of evolution, highlighted in accordance with the criterion formulated above, shows that following (more precisely coherently) objective changes in social life, that is, the change of Homo ontology and information and communication interaction, the ethical system also changed. This is especially evident in the example of family and marriage relations.

Currently, humanity is at the point of bifurcation. Its principal feature is as follows. In terms of information and communication, humanity has previ-

ously evolved through a qualitative transition from the first signal system to the second. Here is the reminder of the essence of this evolution.

Signaling systems are systems of conditional connections that combine the first (sensory) and second (conceptual) signal systems in the brain, providing adequate adaptation to the environment. Both systems work in interaction, perceiving signals from the outside world, and the first signaling system is in humans and animals, while the second signaling system is only in humans. The concept of a "signal system" was introduced by I. P. Pavlov. In 1932, he defined the concept of a "signal system" as central to his teaching about the laws of the brain. The principles of "signaling" act starting from the simplest organisms and then become more and more complicated in the process of evolution. The ability to respond in a timely and adequate manner to the "signal" of the environment is a matter of survival. To meet the vital needs of the body, almost any natural agents (sound, smell, visual image) can serve as signals, so the first signal system common to humans and all living organisms is formed. The qualitative difference between humans and animals lies in the appearance in the process of evolution of the second signaling system, generalized signaling-speech. The first signal system is described by I. V. Pavlov as follows:

> This is what we have in ourselves as impressions, sensations and representations from the surrounding external environment, both common and from our social, excluding the word, audible and visible. This is the first signal system of reality that we have in common with animals.[21]

For him, the second signaling system is a system of conditioned reflex connections in the human brain, where the conditional stimulus is a word, speech. It arises on the basis of the first signal system in the process of communication between people. It is the second signal system that is the regulator of higher nervous activity, the basis of written and oral speech, abstract-logical thinking.

During the evolution of the animal world at the stage of formation and initial development of the Homo sapiens species, a qualitative modification of the alarm system occurred, providing active and collective adaptive behavior of a person to the surrounding world, both natural and social. Using the second signaling system, people have learned not only to communicate and transmit information to each other, but also to accumulate it, process it, store it and transmit it from generation to generation, first orally, and then in writing.

---

[21] Ivan V. Pavlov, *Complete Works*, vol. 3, bk. 2 (1951), 335-336.

In his work *On the Beginning of Human History* the famous paleopsychologist B. F. Porshnev carried out a deep analysis of the process of formation of a person, his speech and consciousness.[22] The biopsychological and social conditions of the formation of brain structures (neocortex) and the formation of languages have been subjected to in-depth analysis by anthropologists and linguists since the end of the 19th century in connection with the study of Sanskrit by European science and with the advent of comparative linguistics.

Human speech communication is not just the perception of signals (words), it is the understanding of their meaning and meaning. Speech as a second signaling system acts as a semiotic system of meanings. The second signal system and the memory are one. Thanks to speech, the human world "doubles," the word allows you to mentally operate with objects even in their absence. Human consciousness carries out a holistic perception of the surrounding world in terms of concepts.

The above brief excursus into the history of the evolution of the signal system was necessary to draw attention to its fundamental importance. Developing this evolutionary line, we assert that a third signaling system is being formed – ethics as a transpersonal structure. Just that transforms humanity into a fundamentally new format – Homo Deus.[23] Otherwise, it will disappear, self-destruct, and natural evolution will look for new ways of development.

Homo Erectus (labor) – Homo Sapiens (labor, speech, sociality) – Homo Deus (continuation of evolution – a new materiality is added – information and communication environment, transpersonal structures – outsourcing of intelligence, regulator – ethics as a transpersonal structure). A qualitative transition will take place in this direction, and a new ontology of humanity will be based on this.

There is nothing to derive the social from, except from the biological, but it does not come down to it. This is the first antinomy. From the social, a third signaling system will be born – the ethics of human behavior. This is the second antinomy. Their solution lies in a neural network model that explains the formation of transpersonal structures.

Other regulators of socio-natural human behavior are not visible. In this direction, humanity will need to reach the level of macrocivilizational subjects and later to the level of self-identification as a human civilization with a single ethical system or coordinated at the level of intercivilizational integration.

---

[22] Boris F. Porshnev, *On the Beginning of Human History: Problems of Paleopsychology* (Aleteya, 2007), 207.

[23] Yuval Noah Harari, *Homo Deus: A Brief History of Tomorrow* (Vintage, 2017).

Author contribution statement

Both authors have contributed equally to the conception and design of the work, the drafting and revising of the manuscript, and the final approval of the version to be published.

Funding statement

The research was carried out at the expense of a grant from the Russian Science Foundation (project No. 24-28-00413; scientific supervisor, Doctor of Philosophy, Professor O.A. Gabrielyan).

References

Bakirov, Akhat, Ibragim E. Suleimenov, and Yelizaveta Vitulyova. "To the Question of the Practical Implementation of 'Digital Immortality' Technologies: New Approaches to the Creation of AI." In *Proceedings of the Future Technologies Conference (FTC) 2022, Volume 1*, edited by Kohei Arai. Springer, 2023.

Bakirov, Akhat, Yelizaveta Vitulyova, A. A. Zotkin, and Ibragim E. Suleimenov. "Internet Users' Behavior from the Standpoint of the Neural Network Theory of Society: Prerequisites for the Meta-Education Concept Formation." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVI-4/W5-2021 (2021): 83-90.

Da Silva, Jaime A. Teixeira, and Panagiotis Tsigaris. "The Relevance of James Lovelock's Research and Philosophy to Environmental Science and Academia." Frontiers of Environmental Science & Engineering 17, no. 3 (2023): 1-2.

Gillis, John R. *Youth and History: Tradition and Change in European Age Relations, 1770–Present*. Elsevier, 2013.

Harari, Yuval Noah. *Homo Deus: A Brief History of Tomorrow*. Vintage, 2017.

Jenkins, Katharine. "Amelioration and Inclusion: Gender Identity and the Concept of Woman." *Ethics* 126, no. 2 (2016): 394-421.

Juozelis, Egidijus. "Religious Dimensions in Transhumanist and Posthumanist Philosophies of Science." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 125-133.

Kormas, Panagiotis, Antonia Moutzouri, and Evangelos D. Protopapadakis. "Implications of Neuroplasticity to the Philosophical Debate of Free Will and Determinism." In *Handbook of Computational Neurodegeneration*, edited by Panayiotis Vlamos, Ilias S. Kotsireas, and Ioannis Tarnanas, 1-19. Springer International Publishing, 2022.

Lefebvre, Vladimir A. *Algebra of Conscience*. Reidel, 1982; expanded 2nd ed. Kluwer, 2001.

Lévy-Bruhl, Lucien. *Le Surnaturel et la Nature dans la Mentalité Primitive*. Presses Universitaires de France, 1963.

Mashour, George A., Pieter Roelfsema, Jean Pierre Changeux, and Stanislas Dehaene. "Conscious Processing and the Global Neuronal Workspace Hypothesis." *Neuron* 105, no. 5 (2020): 776-798.

Pavlov, Ivan V. *Complete Works*, vol. 3, book 2, 1951.

Porshnev, Boris F. *On the Beginning of Human History: Problems of Paleopsychology*. Aleteya, 2007.

Radford, Tim. "James Lovelock at 100: The Gaia Saga Continues." *Nature* 570, no. 7762 (2019): 441-443.

Sagan, Dorion. "James Lovelock and Consciousness: An Obituary." *Journal of Consciousness Studies* 29, nos. 11-12 (2022): 226-231.

Savulescu, Julian, and Evangelos D. Protopapadakis. "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu." *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

Suleimenov, Ibragim E., Akhat Bakirov, Guliyash Niyazova, and Dina Shaltykova. "University as an Analogue of the Neural Network." *E3S Web of Conferences* 258 (2021): 07056.

Suleimenov, Ibragim E., Aliya Massalimova, Akhat Bakirov, and Oleg Gabrielyan. "Neural Networks and the Philosophy of Dialectical Positivism." In *MATEC Web of Conferences*, vol. 214, 02002. EDP Sciences, 2018.

Suleimenov, Ibragim E., Dinara Matrassulova, Inabat Moldakhan, Yelizaveta Vitulyova, Sherniyaz Kabdushev, and Akhat Bakirov. "Distributed Memory of Neural Networks and the Problem of the Intelligence's Essence." *Bulletin of Electrical Engineering and Informatics* 11, no. 1 (2022): 510-520.

Suleimenov, Ibragim E., Oleg Gabrielyan, and Yelizaveta Vitulyova. "Dialectics of Scientific Revolutions from the Point of View of Innovations Theory." *WISDOM* 24, no. 4 (2022): 25-35.

Suleimenov, Ibragim E., Sergey Panchenko, Oleg Gabrielyan, and Ivan Pak. "Voting Procedures from the Perspective of Theory of Neural Networks." *Open Engineering* 6, no. 1 (2016): 318-321.

Suleimenov, Ibragim E., Yelizaveta Vitulyova, Akhat Bakirov, and Oleg Gabrielyan. "Artificial Intelligence: What Is It?" In *Proceedings of the 2020 6th International Conference on Computer and Technology Applications (ICCTA '20)*, 22-25. Association for Computing Machinery, 2020.

Tsitas, Georgios, and Athanasios Verdis. "Proposing a Frame of Ethical Principles for Educational Evaluation in Modern Greece." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 135-158.

Van der Haak, Donovan. "Death Anxiety, Immortality Projects and Happiness: A Utilitarian Argument Against the Legalization of Euthanasia." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 159-174.

Vanchurin, Vitaly. "The World as a Neural Network." *Entropy* 22, no. 11 (2020): 1210.

Vernadsky, Vladimir I. *The Biosphere and the Noosphere*. Airis-press, 2004.

Vitulyova, Yelizateva, Akhat Bakirov, Saltanat Baipakbayeva, and Ibragim E. Suleimenov. "Interpretation of the Category of 'Complex' in Terms of Dialectical Positivism." In *IOP Conference Series: Materials Science and Engineering* 946, no. 1 (2020): 012004.

Wong-MingJi, Diana J., Eric H. Kessler, Shaista E. Khilji, and Shanthi Gopalakrishnan. "Cross-Cultural Comparison of Cultural Mythologies and Leadership Patterns." *South Asian Journal of Global Business Research* 3, no. 1 (2014): 79-101.

A. Gounaris, G. Kosteletos, & M. A. Kolliniati · *Conatus* 10, no. 1 (2025): 127-152
doi: https://doi.org/10.12681/cjp.40628

Virtue in the Machine: Beyond a One-size-fits-all Approach and Aristotelian Ethics for Artificial Intelligence

Alkis Gounaris
National and Kapodistrian University of Athens, Greece
E-mail address: alkisg@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0002-0494-6413

George Kosteletos
National and Kapodistrian University of Athens, Greece
E-mail address: gkosteletos@philosophy.uoa.gr
ORCID iD: https://orcid.org/0000-0001-6797-8415

Maria-Artemis Kolliniati
Ruprecht Karls Universität Heidelberg, Germany
E-mail address: martkolliniati@gmail.com
ORCID iD: https://orcid.org/0000-0003-1553-7014

Abstract

This paper explores the application of Aristotelian virtue (arête), as quality of excellence and as a key notion of ethics, to AI systems as classified in the EU Artificial Intelligence Act. It argues that while the Act's approach based on 'ethical data' and 'prima facie values' aligns with the Rossian paradigm, such principles may not be suitable for all AI systems, particularly those in 'limited' or 'minimal risk' zones. The paper suggests that the Aristotelian concept of virtue can be effectively applied to designing, training, operating and using no-risk or low-risk AI systems. However, its application to the design and training of high-risk areas such as migration, asylum, border control, and justice, where clearly defined objectives are essential, requires ongoing consideration. The paper concludes that by distinguishing between (a) design, development, training, deployment, operation and use, (b) by stage evaluation of systems, and c) virtuous use of the systems, Aristotelian ethics can serve as a post ex evaluating method for all-risk AI systems, while further research and the potential use of regulatory sandboxes are needed to explore the integration of Aristotelian virtues into the design, development and training of such applications. Finally, we propose a virtuous-based 'AI Seal of Excellence' certification process, which empowers the virtuous use of AI systems.

Keywords: *AI ethics; AI virtues; virtuous agents; EU AI Act; arête; Aristotelian ethics for AI; seal of excellence for AI; virtuous use of AI; liberalism; borders and AI*

I. Introduction

The EU Artificial Intelligence Act classifies AI systems into four distinct risk zones, aiming to protect "fundamental rights, democracy, and the rule of law" (EU AI Act). This paper sets out to achieve three primary objectives. First, it asserts that the EU AI Act aligns with an approach based on 'ethical data' and 'prima facie values or duties,' resembling the Rossian paradigm.[1] This alignment is attributed to the Act's objective of ensuring the integrity of Artificial Intelligence (AI) systems, which are considered 'trustworthy AI' and must comply with eight core criteria, including transparency, non-discrimination, and fairness.[2] Second, it explores the potential for applying the "aretological" concept of 'virtue in the machine' to 'limited' and 'minimal and no risk' AI systems. Third, the paper aims to demonstrate that while Aristotelian ethics-based criteria may effectively *evaluate* 'high-risk' AI systems, there are challenges in applying them to the design, training, and operation of such systems. Through hypothetical scenarios, we argue that Aristotelian ethics may not be well-suited for guiding the development and deployment of AI systems in high-risk domains such as migration, asylum, border control, justice, etc. However, it can still serve as a valuable framework for *evaluating* these systems and as a method for guiding users on their virtuous use.

In addressing the first objective, the paper clarifies the EU AI Act's approach, illustrating that a uniform treatment across risk zones is impractical. Provisions on transparency and non-discrimination apply in particular to high-risk AI systems, which must address and overcome the 'value-loading problem.' The development of ethical AI agents requires overcoming this challenge by aligning AI with human values, often through 'prima facie' moral models. However, challenges persist, including the inability to predict all potential scenarios[3] and the necessity for external assessments of machine moral agency.[4]

For the second objective, the paper argues that 'ethical data' and 'prima facie values' may not apply or may not be checkable or even necessary to 'limited' or

---

[1] William D. Ross, *The Right and the Good* (Oxford University Press, 2002).

[2] European Commission, *Ethics Guidelines for Trustworthy AI* (Office for Official Publications of the European Communities, 2019).

[3] Eliezer Yudkowsky, "Complex Value Systems in Friendly AI," in *Artificial General Intelligence*, eds. J. Schmidhuber, K. Thč risson, and M. Looks, 388-393 (Springer, 2011); see also Eliezer Yudkowsky, "The Value Loading Problem," *EDGE*, July 12, 2021, https://www.edge.org/response-detail/26198.

[4] Michael Anderson and Susan Leigh Anderson, "Machine Ethics: Creating an Ethical Intelligent Agent," *AI Magazine* 28, no. 4 (2007): 15; also, Michael Anderson, Susan Leigh Anderson, Alkis Gounaris, and George Kosteletos, "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 177-202.

'minimal and no risk' AI systems. Drawing on examples such as the AI tutor,[5] we anchor our proposal on the concept of Aristotelian virtue (*arête*) by emphasising the cultivation of character and excellence in achieving goals throughout the AI system's life cycle. By doing so, we present a behavioural framework that focuses on the development of virtues rather than a purely functional framework, allowing for an external approach to evaluating the performance of AI systems in low-risk scenarios. This approach suggests using external criteria to address the 'value loading problem,' emphasising fostering individual virtues within a social context. The paper proposes defining a 'virtuous agent' guided by both social and individual ends[6] in alignment with the intrinsic nature of machines.

This "aretological" framework opens the possibility of considering machines as virtuous agents[7] recognised for their contributions to social and individual goals,[8] with their performance evaluated according to their achievements over time[9] rather than procedural aspects.[10]

For the third objective, the paper argues that Aristotelian ethics is ill-suited for guiding the design, development, training and deployment of AI systems engaged in high-risk activities, particularly those involved in areas such as migration, asylum and border control, justice,[11] education, and healthcare. In such contexts, it becomes essential to define the ultimate goal, or '*telos,*'[12] in advance, as AI systems lack the capacity for human-like deliberation concerning their ultimate goals.[13] The paper questions whether this goal can be adequately addressed through an Aristotelian approach alone. It contends that Aristotelian ethics is not a universal theory and that, by the same token, the EU AI

---

[5] John Tasioulas, "First Steps Towards an Ethics of Robots and Artificial Intelligence," *Journal of Practical Ethics* 7, no. 1 (2019): 61-95.

[6] Michael Sandel, *Justice: What's the Right Thing to Do?* (Farrar, Straus and Giroux, 2010).

[7] Martin Gibert, M. "The Case for Virtuous Robots," *AI and Ethics* 3 (2022): 135-144; see also Massimiliano Cappuccio, Eduardo Sandoval, Omar Mubin, Mohammad Obaid, and Mari Velonaki, "Can Robots Make us Better Humans? Virtuous Robotics and the Good Life with Artificial Agents," *International Journal of Social Robotics* 13 (2021): 7-22.

[8] Silviya Serafimova, "Whose Morality? Which Rationality? Challenging Artificial Intelligence as a Remedy for the Lack of Moral Enhancement," *Humanities and Social Sciences Communications* 7 (2020): 1-10.

[9] Sandel.

[10] John Tasioulas, "The Rule of Algorithm and the Rule of Law," Lecture at the University of Vienna, October 15, 2021.

[11] Alkis Gounaris and George Kosteletos, "Writing the Algorithm of Good: Artificial Intelligence as a Machine of Justice," *Ithiki* 19 (2024): 6-27 [in Greek].

[12] Aristotle, *The Nicomachean Ethics*, ed. L. Brown, trans. D. Ross (Oxford University Press, 2009).

[13] Tasioulas, "First Steps." For a Heidegger-inspired analysis on the lack of agency on behalf of AI systems, see also Ashley Roden-Bow, "Killer Robots and Inauthenticity: A Heideggerian Response to the Ethical Challenge Posed by Lethal Autonomous Weapons Systems," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 477-486.

Act cannot adopt a one-size-fits-all approach. While Aristotelian ethics may be appropriate for minimal-risk and no-risk AI systems and for the sole evaluation of high-risk AI systems, it may be inadequate to guide the design, training and operation of systems deployed in high-risk activities such as migration, asylum, warfare,[14] and border control. Accordingly, the paper calls for further philosophical discussion and empirical research, suggesting that the potential use of regulatory sandboxes to explore behavioural evaluation criteria based on Aristotelian ethics may lead us to the safe and virtuous use of high-risk AI applications. Furthermore, we suggest that the virtuous use of AI systems can be realised through the introduction and application of the 'AI Seal of Excellence' certification process, which will be based on virtuous principles.

II. The EU AI Act, trustworthy AI and the 'value loading problem'

The EU Artificial Intelligence Act (EU AI Act) classifies AI systems into four distinct risk categories. First, unacceptable risk, which includes systems like AI used for biased criminal justice decision-making or social scoring. Second, high risk, which encompasses applications in areas such as healthcare, justice, border control, education, hiring, and autonomous vehicles. Third, limited risk, covering systems like chatbots and online shopping recommendation algorithms; and fourth, minimal/no risk, which includes systems such as weather forecasting or spam filters. This categorisation acknowledges that AI systems pose varying degrees of risk to fundamental rights, safety, and societal values, aiming to foster innovation while safeguarding the rule of law.

The EU AI Act adopts a uniform framework to ensure the development of 'trustworthy AI,' which aligns with 'ethical data' and 'prima facie values or duties,' reflecting the Rossian paradigm.[15] The Act aims to ensure that AI systems fulfil eight core criteria, including accountability, sustainability, privacy, and fairness, promoting technical robustness and societal well-being.[16] These ethical principles, such as transparency, non-discrimination, and fairness, ensure that AI systems comply with both functional and moral standards. However, applying these 'prima facie' ethical principles to lower-risk AI systems – such as those in the 'limited' or 'minimal/no risk' zones – raises questions. For such systems, an Aristotelian virtue-based approach may offer a more appropriate ethical framework, particularly as any uniform treatment across all risk categories would not be practical.

---

[14] See Ioanna Lekea, George Lekeas, and Pavlos Topalnakos, "Exploring Enhanced Military Ethics and Legal Compliance through Automated Insights: An Experiment on Military Decision-making in Extremis," *Conatus – Journal of Philosophy* 8, no. 2 (2023), 345–372.

[15] Ross.

[16] European Commission, *Ethics Guidelines for Trustworthy AI*.

While provisions on transparency and non-discrimination are crucial for high-risk AI systems, they are part of the broader 'value loading problem.' This challenge involves harmonising AI systems with human values, which often requires the application of Rossian 'prima facie duties' to a priori moral models. However, significant challenges persist, including the difficulty of foreseeing all possible scenarios, the need for external assessments to evaluate machine moral agency,[17] and the lack of freedom within AI systems to navigate conflicting moral principles and norms. In line with Aristotelian ethics, a 'virtuous' AI system would ideally have the freedom to choose the best course of action based on its deliberative faculties. Yet, this autonomy is absent in most AI systems, particularly in high-risk environments like border control, justice, healthcare or education, where conflicting human interests may arise.[18] For example, AI in the justice system could influence decisions that affect fundamental rights, with the risk of increasing bias or error in legal decisions, especially in 'hard cases' where reasonable lawyers and judges have to discover what the rights of the parties involved are in these contestable cases.[19] In healthcare, errors or biases in diagnosis or treatment could have serious, life-threatening consequences.[20] In education, students' future opportunities and, thus, life plans can be affected by biased AI applied to assessing students. Similarly, the management of border controls, on which we will focus, may involve conflicting interests – such as national security versus humanitarian concerns – and the choice between conflicting interests may have implications for individual freedoms, deprivation of individual rights, abuse and discrimination. If we assume an AI system has Aristotelian ethics, it would still struggle to resolve such conflicts without the capacity for moral deliberation or freedom of choice to make value-based judgments. For example, a border control AI system may face a conflict between national security concerns (e.g., controlling migration) and the humanitarian duty to protect asylum seekers. Since AI systems in these contexts lack the freedom to navigate such moral dilemmas, Aristotelian ethics would be inadequate for resolving these tensions.

Previous research into the 'value loading problem' has proposed the use of various ethical models, such as utilitarianism, deontology, prima facie val-

---

[17] Anderson and Anderson, 15; also, Anderson et al.

[18] In the light of this, especially when it comes to high-risk systems as war drones, some advocate "an international treaty banning all weaponized UAVs." See Joshua M. Hall, "Just War contra Drone Warfare," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 217-239.

[19] Ronald Dworkin, *Taking Rights Seriously* (Harvard University Press, 1978).

[20] This would also undermine the doctor-patient relationship, since patients seem to be quite sensitive on the introduction of AI tools; see Georgia Livieri, Eleni Mangina, Evangelos D. Protopapadakis, and Andrie G. Panayiotou, "The Gaps and Challenges in Digital Health Technology Use as Perceived by Patients: A Scoping Review and Narrative Meta-synthesis," *Frontiers in Digital Health* 7 (2025): 1474956.

ues, and virtues.[21] However, these approaches face difficulties with conflicting moral principles and the challenge of predicting all possible future scenarios. Behaviour-oriented research[22] may potentially offer the only viable solution to this issue. In light of these obstacles and the structure of the "risk zones," we argue that while the EU AI law aims to ensure the ethical development of AI across different risk categories, in practice, it seeks to adopt a generic solution that may have limitations when applied to lower-risk AI systems. A one-size-fits-all solution is not appropriate in such cases. A virtue-based approach grounded in Aristotelian ethics may provide more suitable guidance for these systems, emphasising moral development within specific contexts rather than rigid adherence to pre-loaded duties or Rossian "prima facie duties." In the next chapter, the characteristics of an evaluation framework built on Aristotelian virtues are presented as a solution to tackle the value-loading problem.

III. Towards virtuous AI: An Aristotelian approach to overcoming the 'value loading problem'

In addressing the challenges of behaviour-oriented approaches to the development, evaluation, and use of AI systems, we propose an *evaluation* framework based on Aristotelian virtues.

As Aristotle explains in *Nicomachean Ethics*,[23] virtues (*arêtes*) must be understood in light of our characteristic function (*ergon*). For humans, this function is the activity of the rational part of the soul conducted well or in accordance with excellence. Hence, the cultivation of virtue is inseparable from our ultimate purpose (telos), since it enables us to perform our function in a fully realised manner. In so doing, we attain our proper end (*entelecheia*) and achieve genuine human flourishing (*eudaimonia*).[24]

Similarly, in other texts, such as *On the Soul*,[25] Aristotle distinguishes between the ultimate purpose and the characteristic function (ergon) of tools, exemplified by the axe, whose function is to cut well, thereby conferring upon it functional value.

Furthermore, in *Nicomachean Ethics*, Aristotle, identifies distinct types of virtues. Organic or functional virtues, as the virtue of the eye lies in its capacity

---

[21] Gounaris and Kosteletos, "Writing the Algorithm of Good."

[22] Nathan I. N. Henry, Mangor Pedersen, Matt Williams, Jamin L. B. Martin, and Liesje Donkin, "A Hormetic Approach to the Value-Loading Problem: Preventing the Paperclip Apocalypse," *arXivLabs* (2024), https://arxiv.org/abs/2402.07462.

[23] Aristotle, *The Nicomachean Ethics,* A, 7, 1097b22 – 1098a20.

[24] For an innovative account of Aristotle's views on flourishing or eudaimonia, see Pia Valenzuela, "Fredrickson on Flourishing through Positive Emotions and Aristotle's Eudaimonia," *Conatus – Journal of Philosophy* 7, no. 2 (2022): 37-61.

[25] Aristotle, *On the Soul*, trans. J. A. K. Thomson (Harvard University Press, 1959), II 1, 412b10-15.

of the eye to see clearly (B, 6, 1106a15-20), intellectual virtues (B, 1, 1103a 14-25) are related to logic, computation and learning, and moral virtues (*ibid*), such as practical wisdom (*phronesis*),[26] related with *ethos*, which is regarded as a dispositional choice (*hexis*).[27]

Although most of the above virtues can be attributed to AI systems, the attribution of the moral virtue remains debatable. This is primarily because moral virtue is inherently tied to freedom of choice, raising doubts about its applicability to machines whose operations are bounded by predefined purposes or ends (*telos*). Additionally, the ontological dimension of virtue takes precedence over its moral dimension, as ethics cannot exist without an underlying ontology. In this framework, moral virtue – as a choice of appropriate means – serves as the pathway through which an individual attains their ontological end.

In the same vein, an Aristotelian-based evaluation framework shifts the focus from solely examining the pre-loaded values or internal function of AI systems or their decision-making processes to considering the broader societal context in which these systems operate. By evaluating AI from an external – behaviour-based perspective, we can better address the challenges of creating genuinely virtuous agents. This includes examining the factors that contribute to their instrumental, functional, computational and/ or moral qualities, as well as their potential impact and benefits to society. Such an approach ensures that AI systems are technically effective and align with societal values, fostering their integration and positive contribution to human well-being.

Drawing on examples such as the AI-equipped tutor robot,[28] our approach emphasises the cultivation of character and virtue throughout the lifecycle of an AI system, presenting a behavioural rather than a functionalist theory. This externalist perspective proposes criteria rooted in virtue ethics to address the 'value loading problem,' highlighting the importance of cultivating virtues within a social context. In this context, we do not aim to create 'moral agents,' but instead, we argue that minimal or no risk AI systems can play the role of a 'low-risk agent' that can be considered virtuous because it meets all instrumental, functional, and intellectual criteria, without concern for whether it is a 'literally moral' agent. In fact, questioning its moral status would amount to an anthropomorphic projection onto the agent. However, when it comes

[26] Although Aristotle, in his *Nicomachean Ethics* (A, 13, 1103a 4-5), classifies *phronesis* among the intellectual virtues (alongside *sophia* and *synesis*), he nonetheless deems it indispensable for the realisation of ethical conduct. In this work, we treat phronesis as a foundational element inextricably tied to moral behaviour rather than a value that can be fully captured through computational representation.

[27] Richard Kraut, "Aristotle's Ethics," *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), eds. Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

[28] Tasioulas, "First Steps."

to human decisions of purely moral weight, such as those examined in section 3, the 'quasi-moral' behaviour of the agent[29] becomes relevant in determining whether it can be entrusted with high-risk responsibilities.

At this point, it should be stressed that a virtue-ethical approach to using AI in high-risk contexts not only requires AI agents to be virtuous themselves. It also requires virtuous human users of AI systems. This is relevant, for example, in the context of the virtuous use of AI weapons.[30] [31] The cultivation of a virtuous use could be respectively achieved by the cultivation of virtue in the human users themselves through their interaction with virtuous AI agents. In this case, the latter will perform a significant social task, becoming the means for an exercise of the human users' character (we return to a further analysis of this idea in section 5 of the present text).

This process transcends mere computational abilities, incorporating the expression of virtues within the broader social context. The proposed framework suggests that an ethical agent cannot be understood without its societal role, highlighting its organic function within a symbiotic system.[32] In such a system, practical machine learning occurs, knowledge is continuously acquired, and a feedback loop of virtuous behaviour is established – without relying on pre-loaded values. This "aretological" approach solves ontological, epistemological and other 'value loading' challenges by prioritising behavioural outcomes over internal functional metrics. It thereby opens the theoretical possibility of considering machines as virtuous agents whose contributions to both social and individual goals are evaluated based on their cumulative achievements throughout their existence.

A virtuous agent, in the context of AI ethics, embodies several key characteristics that align with Aristotelian virtues and emphasise the system's integration into social contexts. One fundamental attribute is the ability to learn by doing. This involves learning through practice, where AI systems acquire knowledge and refine their behaviour through iterative processes. Such a dynamic approach enables them to adapt to changing contexts and improve their functionality over time. Another essential characteristic is the capacity for meaningful social interaction. Virtuous AI systems must actively participate in societal networks, drawing on these interactions to align their behaviour with the values and expectations of the communities they serve. This social embeddedness is crucial for fostering trust and ensuring that their actions resonate with the needs of society. Furthermore, virtuous agents must

---

[29] Alkis Gounaris, "Can We Literally Talk About Artificial Moral Agents?" 2020.

[30] Henrik Syse and Martin Cook, "Robotic Virtue, Military Ethics Education, and the Need for Proper Storytellers," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 667-680.

[31] Nigel Biggar, "An Ethic of Military Uses of Artificial Intelligence: Sustaining Virtue, Granting Autonomy, and Calibrating Risk," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 67-76.

[32] Joseph C. R. Licklider, "Man-Computer Symbiosis," *IRE Transactions on Human Factors in Electronics* HFE-1, no. 1 (1960): 4-11.

maintain neutrality concerning predefined moral principles, commonly referred to as 'prima facie values and duties' as described above. Unlike traditional approaches that impose a priori moral obligations and are ultimately linked to the 'value loading problem,' virtuous agents should prioritise fulfilling their specific individual and societal purposes instead. This neutrality allows them to develop moral qualities organically through their actions and integration into their social environments rather than being constrained by rigid ethical models.

To evaluate the success of AI systems as virtuous agents, it is imperative to address two fundamental questions: What is their purpose, and what role do they play in society? Defining their purpose involves identifying their intended contributions, whether in solving problems, enhancing efficiency, or fostering innovation. Meanwhile, understanding their societal role requires assessing how they integrate into existing frameworks and address broader needs, such as social cohesion, fairness, and individual well-being. AI systems can evolve into virtuous agents by embodying these characteristics and aligning them with their defined roles and purposes. This evolution focuses on interaction with behaviour rather than static ethical models, emphasising adaptability, learning, and meaningful contributions to society.

The measure of their success lies not in compliance with fixed principles but in their ability to accomplish their work and, at the same time, to foster both human fulfilment and community flourishing (*eudaimonia*) throughout their lifecycle.[33] This externalist, virtue-based perspective offers a robust framework for addressing the ethical challenges posed by AI development. However, while this externalist virtue-based model is suitable for the evaluation of limited or minimal-risk AI systems, a what-if scenario can exemplify why its application to high-risk AI systems should be avoided. In these cases, a priori moral values should be taken seriously and applied accordingly in these systems' design, operation and use.

IV. What-if scenario: Applying Aristotelian ethics in 'high-risk' AI design, operations, and use

We argue that high-risk AI systems rely on complex factors and cannot be addressed "horizontally" through a general regulatory framework. Instead, special criteria need to be established for the various phases of their design, operation and use.

An illustrative example can be found in AI systems acting as tutors or university faculty members. Assigning such roles to AI entities meets the two fun-

[33] See Sandel, Chapter 8, where he discusses Aristotle's example of the flute. Sandel argues that granting the best flute to the best flutist realises three convergent ends: the instrument's telos (producing excellent music), the musician's personal fulfillment, and the community's overarching good. According to Sandel's interpretation of Aristotle, it is precisely this alignment of individual and collective purposes that constitutes eudaimonia. In such a case, the instrument itself can be seen as participating in virtue, insofar as its proper use fulfills its own function while simultaneously contributing to the flourishing of both the individual user and the wider community.

damental criteria of a virtuous agent: achieving a defined individual goal and contributing to a broader social purpose. The individual goal of the AI system could involve conducting original research in its field or mentoring students. Its social purpose would focus on promoting learning and research or improving the efficiency and optimisation of time dedicated to both teaching and research activities. However, despite being designed with virtue-based principles or performing well in evaluations of its operation, the design, operation, and even the use of such a system might ultimately fail to be truly 'virtuous.' For instance, such a system could be misused to disseminate propaganda, facilitate academic dishonesty – such as enabling students to cheat – or advance the interests of specific social groups within the educational sector.

Since the ethical issues associated with the use of high-risk agents cannot be exhaustively anticipated and the application of the concept of 'moral virtue' fails to adequately extend to such agents, designing these systems based on aretological criteria becomes particularly problematic. The core challenges that arise are typically related to issues of autonomy, freedom of choice, and cognitive limitations, which render these systems incapable of fully understanding their functional roles. The absence of autonomy and freedom of choice in AI systems – especially in high-risk areas such as border control – shows that Aristotelian ethics alone may not be sufficient to address the complex moral dilemmas that arise in these contexts.

In high-risk systems, such as those related to migration, asylum, and border control, it is crucial to determine the system's ultimate purpose, or *telos*,[34] in advance. AI lacks the capacity for human-like deliberation regarding its ultimate objectives,[35] which brings the 'value loading' problem back into focus.

When applying the virtuous agent model to migration policies, such as asylum procedures or border control, it is critical to assess whether it can promote just systems. These systems involve significant ethical and legal complexities, impacting fundamental rights and freedoms. This raises the question: Can a virtuous agent model produce fair outcomes, or does it face limitations in high-risk operations?

As we have emphasised, a virtuous agent, according to Aristotelian ethics, should promote both individual and societal flourishing (*eudaimonia*). However, in high-risk systems, operational effectiveness alone is insufficient. Specific criteria must also be ensured, such as the protection of vulnerable individuals, legal compliance, and the safeguarding of human rights. In the context of migration policies, AI systems can ideally expedite administrative procedures, such as asylum applications, or detect fraud and identify vulnerable individuals. At the same time, however, concerns arise regarding privacy protection, discrimination, and injustices that may have profound impacts on human lives.

---

[34] Aristotle, *The Nicomachean Ethics*.

[35] Tasioulas, "First Steps."

The objectives of such systems may include legal compliance, border security, or facilitating access to asylum. However, these individualized goals are shaped by broader social and political directives and decisions, which precondition both the effectiveness and ethical behavior of these systems.

In Aristotelian ethics, as mentioned above, virtues are cultivated through action, with an emphasis on the development of moral character and practical wisdom (*phronesis*). However, AI systems lack the capacity for moral reasoning and the ability to navigate complex, context-dependent ethical dilemmas, such as those encountered in migration management. Unlike human agents, AI systems do not possess the cognitive and emotional capabilities necessary for ethical deliberation concerning the consequences of their decisions on vulnerable populations. Thus, the application of virtues such as justice, courage, and temperance[36] in these systems becomes problematic. The "virtue" of artificial intelligence is ultimately based on pre-defined criteria rather than genuine moral reasoning.

Moreover, applying the virtuous agent model to high-risk AI systems in the areas of migration and asylum control overlooks the political and value-laden nature of these fields. The use of AI in this context is inherently tied to societal debates on issues such as open versus closed borders, migration, human rights, and the diverse values associated with different approaches to political theory.[37] These issues are subject to shifting political agendas and public opinion.

In practice, such systems may prove inadequate in addressing these contentious issues, as their social responsibilities – such as fraud detection or enhancing security – may conflict with goals related to justice and human rights. For example, even within liberal theory, there are divergent approaches, with some liberals arguing for open borders and freedom of movement as a central element of human life planning,[38] while other liberal approaches may suggest that in an idealised 'realistic utopia,' forced migration, in particular, would be eliminated,[39] or they may raise concerns about the divergent political principles of different communities,[40] echoing issues raised by communitarians.[41]

---

[36] Andrew P. J. Mullins, "What Does Self-control Look Like? Considerations About the Neurobiology of Temperance and Fortitude," *Conatus – Journal of Philosophy* 10, no. 1 (2025): forthcoming.

[37] Maria-Artemis Kolliniati, *Interpreting Human Rights: Narratives from Asylum Centers in Greece and Philosophical Values* (Routledge, 2024).

[38] Joseph Carens, "Migration and Morality: A Liberal Egalitarian Perspective," in Free Movement: Ethical Issues in the Transnational Migration of People and Money, eds. B. Barry and R. Goodin, 25-47 (Harvester Wheatsheaf, 1992).

[39] John Rawls, *The Law of Peoples* (Harvard University Press, 2002).

[40] John Rawls, *A Theory of Justice* (Harvard University Press, 1999).

[41] Joseph Carens, *The Ethics of Immigration* (Oxford University Press, 2013), as cited in Kolliniati, *Interpreting Human Rights*.

Given these factors, applying the virtuous agent model to such systems requires modifications that incorporate pre-defined values. However, the regulatory and politically charged environment in which they operate complicates the development of ethically sustainable systems. Without human moral judgment, the risk of discrimination against vulnerable groups, such as asylum seekers, is heightened. Consequently, Aristotelian virtues are insufficient for the design, training, and operation of these systems, as they are likely to encounter conflicting dilemmas.

On the one hand, the goal of maintaining national sovereignty and protecting the local community calls for closed borders for asylum seekers. Virtues such as prudence, responsibility, and justice prioritise the well-being and security of the state and its citizens.

On the other hand, international legal obligations, such as human rights treaties, advocate for open borders to ensure the fair assessment of asylum claims. Virtues such as compassion and respect for human dignity support the protection of vulnerable individuals fleeing persecution.

Aristotelian ethics, emphasising achieving a final purpose (*telos*) through virtue, struggles to resolve such dilemmas. The two objectives – protecting the local community and upholding international obligations – can both be considered virtuous but often come into conflict. For instance, the virtue of prudence might favour closed borders to safeguard security, while the virtue of compassion demands open borders for humanitarian reasons. Aristotle's concept of practical wisdom (*phronesis*) suggests that virtuous actions should be context-sensitive and aim for balance. However, the competing virtues in this case offer no clear solution. The tension persists, as reconciling national sovereignty with the fulfilment of international law obligations proves difficult through Aristotelian ethical balancing.

At this point, a broader problem inherent in Virtue Ethics emerges. In order to explain why a particular trait qualifies as a virtue or to prioritize one virtue over another in situations of moral conflict, Virtue Ethics must often appeal to concepts and criteria from other ethical frameworks, such as ethical egoism or social contract theory.[42]

The dilemma presented here illustrates the conflict between protecting the well-being of the local community – defensible under a version of social contract theory – and the right to human dignity, even for asylum seekers or those crossing borders illegally – defensible through rights-based or Kantian approaches.

This incompatibility highlights the incompleteness of Virtue Ethics, which compels us – in the context of high-risk AI systems – to adopt prima facie moral values. This, in effect, undermines the value of aretology and simultaneously leads us back to the value loading problem that we sought to avoid with our virtue-based

[42] James Rachels, *The Elements of Moral Philosophy* (McGraw-Hill, 2015), 172.

approach in this paper. By contrast, in low-risk and minimal-risk contexts, moral dilemmas requiring strict prioritisation of ethical principles do not typically arise.

The inability to apply this model to high-risk AI systems, such as those used in migration and border control, reveals inherent issues with these systems, which can be summarised in six additional points. First, they do not understand – or it remains uncertain whether they understand – the concept of morality.[43] Second, under current conditions, they cannot be regarded as 'moral persons,' and as such, they cannot bear responsibilities or be held accountable.[44] Third, they are incapable of voluntarily cultivating virtues, as they are constrained by their objectives and, therefore, cannot freely err. Fourth, they cannot demonstrate equity. This is due to the 'frame problem' of AI, namely the fact that systems are programmed by finite programs, or trained by a finite number of examples (e.g. in the case of artificial neural networks), and therefore there are cases for which they do not have all the critical information, and thus their behaviour in these cases is 'rigid.' Nevertheless, this problem could potentially be overcome if AI systems were given the capacity for true understanding. However, as mentioned above, it is highly doubtful that they have the capacity for true understanding due to the 'other minds problem.'[45]

Fifth, learning by example in reinforcement learning systems could, under certain conditions, be considered analogous to the continuous life experience that leads to the development of *phronesis* (practical wisdom). However, a reasonable question is, "How long should this process of experience and refinement take place before an AI entity can reach the level of *phronesis*?" For humans, this is often a lifelong process. However, when it comes to AI systems deployed in high-risk contexts, a continuous self-improvement process without clear time boundaries would be difficult to accept. Therefore, even if we consider learning by example in deep learning systems as a sufficient analogy to the lifelong experience that leads to *phronesis*, the temporal indeterminacy of achieving *phronesis* is something we allow for humans – since life experience and the accumulation of knowledge are continuous – but not for AI systems, particularly when these are intended to operate in high-risk contexts. Moreover, the absence – or our inability to verify – of key cognitive characteristics in AI entities that are prerequisites for *phronesis*, such as moral sensitivity and moral attentiveness, further heightens our reluctance to adopt a virtue ethics model for AI agents operating in high-risk environments.

---

[43] Gounaris and Kosteletos, "Writing the Algorithm of Good."

[44] Alkis Gounaris and George Kosteletos, "Licensed to Kill: Autonomous Weapons as Persons and Moral Agents," in *Personhood*, eds. D. Prole and G. Rujević, 137-189 (The NKUA Applied Philosophy Research Lab Press, 2020).

[45] Anita Avramides, "Other Minds," *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), eds. Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/win2023/entries/other-minds/.

Additionally, in the same context, even meeting operational criteria does not necessarily imply the 'virtuous use' of high-risk AI systems. A virtue, on its own, is insufficient to produce ethical behaviour. For example, HAL 2000's commitment to ensuring the mission's safety led to catastrophic decisions for the spacecraft crew. A virtue detached from the concept of a 'moral person' can also enable unethical or unlawful behaviours. For instance, the virtue of courage can embolden a criminal or a terrorist. Aristotle argues that the misuse of virtue for harmful purposes is prevented by its combination with *phronesis* (practical wisdom). He distinguishes between perfect virtue and natural virtue, the latter being a primitive version of perfect virtue – an early form shaped merely by predisposition and emotion (as often seen in children) rather than by rational choice and *phronesis*. Children, as well as adults who, despite good intentions, fail to help others unintentionally, lack *phronesis* or practical wisdom. They fail either because they do not know what is necessary to implement their good intentions or because they do not correctly recognise what is beneficial or harmful to others. Thus, *phronesis* requires specific cognitive skills, such as the ability to evaluate which features of a particular situation are most significant from a moral standpoint. In this sense, *phronesis* presupposes the presence of intellectual capacities such as moral sensitivity, moral attentiveness, and moral imagination.[46] AI systems do not possess – or, due to the 'other minds problem,' we cannot ascertain whether they possess – such cognitive traits.

Sixth, in such contexts, it is crucial to define the system's ultimate purpose, or '*telos*'[47] in advance, given that AI lacks the capacity for human-like deliberation on its end goals.[48]

However, despite the fact that Aristotelian ethics can lead to several pitfalls in the design of high-risk AI systems, the conclusion is different when it comes to the *evaluation* of such systems. In the next chapter, we examine the idea of adopting a virtue ethics approach to the evaluation of high-risk AI systems.

V. Towards an Aristotelian evaluation method: Adopting virtue-based criteria for the assessment of low, medium and high-risk AI systems

Given the classification framework of the EU AI Act, which categorises AI systems into various risk zones, our analysis aims to demonstrate through hypothetical scenarios that while characterising an AI system as 'virtuous' is feasible in low- or medium-risk environments, Aristotelian ethics is not an appropriate

---

[46] Rosalind Hursthouse and Glen Pettigrove, "Virtue Ethics," *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), eds. by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

[47] Aristotle, *The Nicomachean Ethics*.

[48] Tasioulas, "First Steps."

framework for guiding the design, construction, and training of AI systems involved in high-risk activities. However, Aristotelian ethics is perfectly applicable to the evaluation of high-risk systems. In particular, it can serve as a valuable reference for establishing evaluation criteria for the operation and use of such systems. By distinguishing between the phases of design, operation, and use, it is argued that virtue-based criteria can indeed be applied, according to the following Table 1.

Application of Virtue-Based Criteria	Can virtue-based criteria be introduced during the Design, Development, and Training Phase (virtuous by design – a priori assessment)	Can virtue-based criteria be introduced for evaluating the Deployment and Operation Phase (a posteriori virtuous assessment)	Can virtue-based criteria be introduced to evaluate the use of the systems?
High-Risk Zone	NO	PROBABLY YES	YES
Medium -Risk Zone	YES	YES	YES
Low -Risk Zone	YES	YES	YES

Table 1.

In other words, we argue that in the hypothetical scenario where such high-risk systems are developed and deployed, they can still be evaluated in use (post ex or a posteriori virtuous assessment) using Aristotelian criteria based on their behaviour. Specifically, they can be assessed through their outcomes and behaviour within a regulatory sandbox to determine whether they meet the conditions to be characterized as 'virtuous' agents or systems. As controlled environments, regulatory sandboxes offer flexibility for experimentation under real-world conditions, allowing regulatory bodies to evaluate system behaviour before full implementation while maintaining strict oversight and control.[49] This process reduces the risk of unintended consequences, promotes compliance with human rights standards, and addresses issues

[49] Dirk A. Zetzsche, Ross P. Buckley, Janos N. Barberis, and Douglas W. Arner, "Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation," *Journal of Corporate and Financial Law* 23, no. 1 (2017): 31-103.

such as algorithmic bias. Within this context, the systems can be examined to determine whether they meet the criteria of virtue and justice prior to their broader deployment.

In such environments, evaluation criteria based on virtuous behaviour and respect for fundamental values can be introduced. During the operation of high-risk AI systems, such as those in migration and asylum applications, evaluators can monitor the system's interactions with vulnerable populations using indicators such as fairness in decision-making and the avoidance of biases. For example, the system can be assessed on its ability to recognize cases that require exceptions to the rule, thereby demonstrating equity, a key virtue in Aristotelian ethics. Additionally, *phronesis* (practical wisdom) can serve as a criterion for the system's adaptability to different ethical and legal frameworks. The 'a posteriori' assessment of such systems can function as a feedback-loop refinement,[50] enabling continuous improvement to better align with ethical principles and regulatory requirements.

An AI system or agent involved in decision-making for asylum applications should, for example, demonstrate transparency by providing justifications for each rejection and allowing for appeals and revisions. The virtue of responsibility demands accountability, which can be achieved through regular performance evaluations and the implementation of reporting mechanisms for errors or discriminatory practices. Thus, the sandbox can serve as a feedback and evaluation environment for high-risk systems using criteria grounded in virtue ethics.

Table 2 below describes the relationship between various aspects of moral virtues and AI systems. It compares moral virtues in terms of their connection to reason, freedom of choice, and their role in achieving *eudaimonia* – that is, human flourishing. The table also explores how these virtues can manifest in AI, particularly in high-risk applications, such as border control systems for asylum evaluation.

It is emphasized that while AI itself may not possess moral virtues, it can exhibit moral behaviour if correctly designed, programmed, and trained. The table further highlights the importance of the EU AI Act in regulating these systems, with a specific focus on justice, transparency, and human rights in high-risk zones.

[50] For the reinforcement learning problem, see Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, 2018).

Virtue category	Relationship with Reason	Relationship with Freedom of Choice	Prerequisite for Eudaimonia	AI Example	Connection to EU AI Act and Risk Zones
Ethical Virtues	Guided by practical wisdom (*phronesis*)	Require freedom of choice	Yes, for humans	AI may not possess ethical virtue due to ontological, epistemological and efinitional limitations (as described above)	**High-Risk Zone:** AI systems for border control (e.g., asylum claim assessment). Must ensure fairness, and human rights, balancing conflicting societal interest
Quasi-Moral Virtues	Requires complex reasoning, depends on predetermined purpose and design based on value loading	No freedom or deliberation	No, but meets instrumental and functional excellence	AI that complies with moral principles and executes its designed role excellently.	**High-Risk Zone:** AI used in automated border control, ensuring compliance with international law and preventing bias in decision-making. Must adhere to rustworthiness criteria.
Intellectual Virtues	Requires complex reasoning	Not dependent on freedom	Yes, for humans. No, for AI Agents but meets instrumental and functional excellence	AI can solve problems, explain reasoning, analyse data, simulating science or prudence.	**Medium-RiskZone:** AI for detecting fraud in migration data or/ and providing recommendations - suggestions for asylum eligibility. Aligns with legal frameworks.
Functional Virtues	Requires high intelligence, Influenced by reason but not fully	Not dependent on freedom	No, but meets instrumental and functional excellence	AI coordinating processes within a society or organisation based on moral principles and logic.	**Medium-Risk Zone:** AI managing asylum processes, ensuring fair treatment and legal compliance with transparency and non-discrimination

Instrumental Virtues	Do not involve reason	Do not require freedom	No, but meets nstrumental excellence	Reactors in AI that function correctly to collect data. Systems working optimally for its purpose, e.g., a chatbot answering questions correctly.	**Low-Risk Zone:** AI systems in administrative tasks (e.g., virtual assistants for immigration queries). AI systems for biometric data collection at border points.

Table 2.

Therefore, based on the above, even if the application of virtue-based criteria is insufficient for the creation and design of high-risk agents, once such agents are deployed – regardless of how they are constructed – the establishment of virtue-based criteria for evaluating the operation and use of high-risk AI systems becomes imperative. Such an application could have immediate practical value. The evaluation of machines through an Aristotelian virtue assessment system could lead to a process of certifying AI systems based on the virtues they exhibit. In this context, we propose the development of a 'Seal of Excellence' based on Aristotelian virtues. This is described in the next section of this paper.

VI. A Seal of Excellence for AI: From virtuous systems to virtuous users

A method of evaluation checks, such as the one presented in the previous section, could also serve as a general guide for system developers during the design, development, and training phases for medium- or low-risk systems or agents, as well as during deployment, operation, and use across all risk categories. We propose that this assessment method would also be highly beneficial for users and policymakers.

Inspired by the successful European Commission's model, we propose a Virtue-Based Seal of Excellence Certificate based on Aristotelian criteria tailored to virtue categories, use, and risk zones. The existing EU Seal of Excellence is awarded to project proposals that meet the high-quality standards of EU funding programmes, certifying their excellence and enabling access to alternative funding sources.[51] It enhances project credibility, attracts investment at multiple levels, and ensures security through digital sealing.[52]

[51] European Commission, *Seal of Excellence*, https://commission.europa.eu/funding-tenders/find-funding/seal-excellence_en.

[52] European Commission, *How Can Seal Holders Use the Seal of Excellence?* European Com-

Our approach extends this concept to AI by embedding ethical evaluation into system design, deployment, and governance. We propose a certification process whereby a "Seal of Excellence" for virtuous AI systems or agents serves as a mark of distinction, signifying that a system demonstrates virtuous behaviour. In practical terms, this entails not only fulfilling its specific design objectives but also adhering to ethical standards that contribute positively to broader societal aims. In this sense, the certification of virtuous agents could function as a bridge between philosophy, social imperatives, and system design – integrating these domains in a concerted effort to ensure that AI development remains fully aligned with human values and goals.

The Seal of Excellence for AI could be considered best practice, as, from a virtue ethics perspective, an AI system or agent could even be awarded or decorated in a form of commendation, analogous to the decoration of animals for their service. Notably, the PDSA Dickin Medal has been awarded to animals, particularly dogs, in recognition of their bravery and contributions in military conflicts. A recent example is the commendation of Diesel, a police dog honoured posthumously after being killed in action during an anti-terror operation in Paris in 2015.[53]

Importantly, this would also serve an instructive and exemplary function for human users of AI. Discussions on AI ethics often emphasise the goal of responsible use. However, from a virtue ethics perspective, the notion of responsible use -and consequently, its objective- could be reframed as virtuous use. It is essential to cultivate the character of users so that they engage with AI in a virtuous manner. A society that achieves *eudaimonia* through AI does not merely require virtuous AI systems; it also demands virtuous users capable of interacting with them in ethically sound ways. More broadly, insofar as human life is increasingly intertwined with AI, a virtuous life – one that ensures *eudaimonia* – must encompass virtuous interaction with AI and, consequently, its virtuous use. Furthermore, if the virtuous life necessitates *phronesis* (practical wisdom), and *phronesis* itself is cultivated through lived experience and learning by example, then AI systems with which humans interact function both as exemplars and as integral components of human experiential learning. Considering the limitations of Tables 1 and 2 (above), by designing AI systems as virtuous agents – and even more so by conferring distinctions upon them – we effectively establish them as paradigms of virtue from which human users can derive practical wisdom. As AI increasingly permeates daily life, the ethical exemplars surrounding us will no longer pertain

mission, 2021, https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/seal-excellence/how-can-seal-holders-use-seal-excellence_en.

[53] PDSA, *PDSA Dickin Medal*, https://www.pdsa.org.uk/what-we-do/animal-awards-programme/pdsa-dickin-medal.

solely to traditional aspects of human existence but will extend to our interaction with AI. Consequently, our engagement with AI systems will shape a significant part of our moral development and real-world conditioning. If this interaction involves engagement with virtuous AI systems or agents, then we will have reinforced the presence of virtuous exemplars in our environment. In fact, the cultivation of virtuous use can take place at two different levels of user consciousness: a) a fully conscious level at which users perceive and accept AI systems as role models because of the AI Seal of Excellence that these systems carry, and b) a less conscious – perhaps even unconscious – level at which users' daily interaction with virtuous AI systems – even with virtuous AI systems that have not yet received the AI Seal of Excellence – inevitably shapes users' characters in a virtuous way. In this case, it is the AI systems that set the tone for their interaction with humans. This interaction inevitably takes place in virtuous contexts because of the virtuous nature of the AI systems themselves. Thus, everyday interaction with AI becomes a process of habituation (i.e. a less conscious process) through which users acquire virtue. In this case, it's not the conscious process of modelling, but the less conscious – or even automatic – process of being molded by everyday practice. Through AI, we will have created an ecosystem that, at least in its technological dimension, cultivates virtue in human users, guiding them towards moral excellence, shaping their character, and fostering the development of *phronesis*. This perspective could serve as a response to the legitimate concerns that human interaction with AI might erode their virtues.[54] On the contrary, engagement with virtuous AI has the potential to strengthen them. Conclusively, a virtuous AI system or agent could explicitly motivate, educate and/or even demand its optimal use by the user or implicitly lead the end user in such use. By analogy with Aristotle's flute's instrumental virtue (see footnote 34 above), virtuous AI could be a flute, making the flutist a better musician.

The AI Seal of Excellence embodies a transformative approach to ethical AI, one that transcends mere harm mitigation and aspires towards moral cultivation. By recognising AI systems that exemplify virtue – promoting *phronesis*, justice, beneficence, honesty, and social virtues – the certification does more than validate ethical compliance; it establishes AI as an active agent in shaping human morality. This perspective challenges the prevailing concerns that AI may erode human virtues, suggesting instead that well-designed AI can reinforce them. Just as virtuous AI inspires virtuous users, a society shaped by AI engagement must prioritize both ethical system design and ethical user development. By institutionalizing this vision through thorough evaluation, public engagement, and adaptive governance, the Seal of Excellence for AI

[54] Nir Eisikovits and Dan Feldman, "AI and Phronesis," *Moral Philosophy and Politics* 9, no. 2 (2022): 181-199.

represents a significant shift. It moves from AI systems that simply follow ethical guidelines to those that actively promote moral engagement, creating a beneficial cycle between technology and ethics.

VII. Conclusion

In this paper, we have argued for an innovative application of Aristotelian virtues (*arêtes*) as qualities of excellence and a key notion to Artificial Intelligence ethics, mindful of the diverse risk categories delineated in the EU AI Act. At the heart of this discussion lies the realisation that the one-size-fits-all approach of the AI Act – rooted in 'ethical data' and 'prima facie' or 'a priori' values – has intrinsic limitations when extended to all-risk systems and in particular to those that do not typically face irreconcilable moral dilemmas. As we have demonstrated, an AI system can indeed exhibit forms of virtue, particularly in low- or medium-risk settings, by achieving functional excellence and serving a defined purpose (*telos*). Nonetheless, it lacks the freedom of choice that, in Aristotelian thought, is a prerequisite for genuine moral virtue. Even if an AI system is 'loaded' with moral values for decision-making, it does not develop moral will per se; it relies on preexisting values set by its designers. The question of authentic consciousness or internal volition, moreover, foregrounds the 'other minds problem.'[55] We have no definitive way of confirming whether an AI truly 'feels' or 'thinks' in a human-like manner. In this light, although a virtue-based perspective offers a useful methodology for evaluating system behaviour, the strictly moral dimension of virtue (which depends on freedom and practical wisdom) is challenging to replicate in high-risk applications. That is why, as we have emphasised, virtue ethics is most appropriate mainly as an external, ex post (or a posteriori) evaluation criterion, rather than as a foundation for the initial design of AI systems in critical domains such as border management or judicial procedures.

Through an external behaviour-based evaluation, both a human cognitive system and a high-risk AI system may be assessed as capable of achieving their final purpose. However, simply arriving at the desired outcome does not demonstrate the presence of *phronesis* (practical wisdom). Indeed, a system can display computational virtue and operational and/or instrumental excellence without possessing the moral imagination or sensitivity that *phronesis* presupposes. Computational virtue, which is adequate for complex reasoning and problem-solving procedures, is insufficient for genuine Aristotelian *phronesis*, as features such as moral imagination, moral sensitivity, and other cognitive capacities remain unverified in current AI systems. Additionally, while humans often require a lifetime to develop *phronesis*, AI systems are 'born'

[55] Avramides.

fully developed;[56] there is no clear developmental arc in which they gradually cultivate moral discernment. As a result, even if they fulfill key goals effectively, we cannot justify calling them quasi-moral agents.

At last, the pre-loading of values and the predefinition of final purposes by system designers – since, as Tasioulas notes, AI entities lack the capacity for human-like deliberation – leads us to refer to this as quasi-*phronesis*. These systems may mimic some elements of virtuous conduct but cannot autonomously choose, revise, or weigh moral ends in the robust sense implied by Aristotelian virtue.

In conclusion, as we have demonstrated, virtue ethics encounters formidable challenges when one attempts to build or train AI systems in high-risk sectors – including those that affect migration, asylum, border control, healthcare, education, and justice. In these domains, conflicting (normative) objectives arise frequently, often hinging on complex legal, societal, or humanitarian considerations that demand immediate and pre-specified moral imperatives. While virtues encourage context-sensitive discernment, contemporary AI systems cannot replicate the kind of *phronesis* (practical wisdom) that is central to Aristotelian thought. As we emphasised, lacking genuine autonomy and freedom of choice, such systems are ill-equipped to engage in genuine moral deliberation. For this reason, we proposed that an Aristotelian model is unsuitable for *designing* and *training* these high-risk systems. However, Aristotelian ethics retains value in *assessing* how high-risk systems perform and are used post-deployment. Adopting a virtue-based *a posteriori evaluation* method, if needed within regulatory sandboxes, enables policymakers and researchers to observe whether AI systems or agents uphold fairness, mitigate bias, and promote collective *eudaimonia* in the sense of fulfilling social objectives. This form of dynamic, behaviour-oriented oversight aligns with the notion that AI's real-world performance should be judged not only by technical metrics but also by the social outcomes it produces. Even in domains where moral dilemmas are acute, tracking whether a system's operation demonstrates virtuous behaviour helps identify potential improvements and fosters user trust.

In this direction, we have proposed an Aristotelian *evaluation method* that adopts virtue-based criteria for assessing low, medium, and high-risk AI systems. This method leads to the Virtuous AI system or agent's acknowledgement, certification and 'decoration.' As a central contribution, the virtue-based "AI Seal of Excellence" underscores how a series of criteria can

[56] For a defence of the opposing view, i.e., the position that *phronesis* can develop in AI systems, see John P. Sullins, "Automated Ethical Practical Reasoning: The Problem of Artificial Phronesis," in *Robophilosophy: Philosophy of, for, and by Social Robotics*, eds. J. Seibt, R. Hakli, and M. Nørskov (MIT Press, 2025), forthcoming.

serve as a constructive framework for both developers and users across diverse risk levels. Inspired by existing European certification models, we propose an 'AI Seal of Excellence' which awards AI systems that exhibit virtuous behaviour over time, as measured against clearly defined operational excellence and social benefit thresholds. Recognising such systems publicly would not only motivate industry-wide adherence to higher ethical standards but also encourage a reciprocal dynamic in which virtuous AI acknowledgement fosters virtuous user practices. From a broader philosophical view, involving both designers and users in an ongoing effort around virtue cultivation holds promise for aligning AI's expanding role in society with human flourishing. Insofar as future work can integrate philosophical depth with technological sophistication, the Aristotelian paradigm may help create AI that does more than merely minimize harm, instead contributing positively to the shared pursuit of *eudaimonia*.

Author contribution statement

All authors have contributed equally to the conception and design of the work, the drafting and revising of the manuscript, and the final approval of the version to be published.

References

Anderson, Michael, and Susan Leigh Anderson. "Machine Ethics: Creating an Ethical Intelligent Agent." *AI Magazine* 28, no. 4 (2007): 15.

Anderson, Michael, Susan Leigh Anderson, Alkis Gounaris, and George Kosteletos. "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 177-202.

Aristotle. *On the Soul*. Translated by J. A. K. Thomson. Harvard University Press, 1959.

Aristotle. *The Nicomachean Ethics*. Edited by L. Brown. Translated by D. Ross. Oxford University Press, 2009.

Avramides, Anita. "Other Minds." *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), edited by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/win2023/entries/other-minds/.

Biggar, Nigel. "An Ethic of Military Uses of Artificial Intelligence: Sustaining Virtue, Granting Autonomy, and Calibrating Risk." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 67-76.

Cappuccio, Massimiliano, Eduardo Sandoval, Omar Mubin, Mohammad Obaid, and Mari Velonaki. "Can Robots Make Us Better Humans? Virtuous Robotics and the Good Life with Artificial Agents." *International Journal of Social Robotics* 13 (2021): 7-22.

Carens, Joseph. "Migration and Morality: A Liberal Egalitarian Perspective." In *Free Movement: Ethical Issues in the Transnational Migration of People and Money*, edited by B. Barry and R. Goodin, 25-47. Harvester Wheatsheaf, 1992.

Carens, Joseph. *The Ethics of Immigration*. Oxford University Press, 2013.

Dworkin, Ronald. *Taking Rights Seriously*. Harvard University Press, 1978.

Eisikovits, Nir, and Dan Feldman. "AI and Phronesis." *Moral Philosophy and Politics* 9, no. 2 (2022): 181-199.

European Commission. *Ethics Guidelines for Trustworthy AI*. Office for Official Publications of the European Communities, 2019.

European Commission. *How Can Seal Holders Use the Seal of Excellence?* European Commission, 2021. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/seal-excellence/how-can-seal-holders-use-seal-excellence_en.

Gibert, Martin. "The Case for Virtuous Robots." *AI and Ethics* 3 (2022): 135-144.

Gounaris, Alkis, and George Kosteletos. "Writing the Algorithm of Good: Artificial Intelligence as a Machine of Justice." *Ithiki* 19 (2024): 6-27 [in Greek].

Gounaris, Alkis. "Can We Literally Talk About Artificial Moral Agents?" 2020.

Gounaris, Alkis, and George Kosteletos. "Licensed to Kill: Autonomous Weapons as Persons and Moral Agents." In Personhood, edited by Dragan Prole and Goran Rujević, 137-189. The NKUA Applied Philosophy Research Lab Press, 2020.

Henry, Nathan I. N., Mangor Pedersen, Matt Williams, Jamin L. B. Martin, and Liesje Donkin. "A Hormetic Approach to the Value-Loading Problem: Preventing the Paperclip Apocalypse." *arXivLabs* (2024), https://arxiv.org/abs/2402.07462.

Hall, Joshua M. "Just War contra Drone Warfare." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 217-239.

Hursthouse, Rosalind, and Glen Pettigrove. "Virtue Ethics." *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), edited by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

Kolliniati, Maria-Artemis. *Interpreting Human Rights: Narratives from Asylum Centers in Greece and Philosophical Values*. Routledge, 2024.

Kraut, Richard. "Aristotle's Ethics." *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), edited by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

Lekea, Ioanna, George Lekeas, and Pavlos Topalnakos. "Exploring Enhanced Military Ethics and Legal Compliance through Automated Insights: An Experiment on Military Decision-making in Extremis." *Conatus – Journal of Philosophy* 8, no. 2 (2023), 345-372.

Licklider, Joseph C. R. "Man-Computer Symbiosis." *IRE Transactions on Human Factors in Electronics* HFE-1, no. 1 (1960): 4-11.

Livieri, Georgia, Eleni Mangina, Evangelos D. Protopapadakis, and Andrie G. Panayiotou. "The Gaps and Challenges in Digital Health Technology Use as Perceived by Patients: A Scoping Review and Narrative Meta-synthesis." *Frontiers in Digital Health* 7 (2025): 1474956.

Long, David. *The Animals' VC: For Gallantry and Devotion: The PDSA Dickin Medal-inspiring Stories of Bravery and Courage*. Random House, 2012.

Mullins, Andrew P. J. "What Does Self-control Look Like? Considerations About the Neurobiology of Temperance and Fortitude." *Conatus – Journal of Philosophy* 10, no. 1 (2025): forthcoming.

PDSA. *PDSA Dickin Medal*. https://www.pdsa.org.uk/what-we-do/animal-awards-programme/pdsa-dickin-medal.

Rachels, James. *The Elements of Moral Philosophy*. McGraw-Hill, 2015.

Rawls, John. *A Theory of Justice*. Harvard University Press, 1999.

Rawls, John. *The Law of Peoples*. Harvard University Press, 2002.

Roden-Bow, Ashley. "Killer Robots and Inauthenticity: A Heideggerian Response to the Ethical Challenge Posed by Lethal Autonomous Weapons Systems." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 477-486.

Ross, William David. *The Right and the Good*. Oxford. Oxford University Press, 2002.

Sandel, Michael. *Justice: What's the Right Thing to Do?* Farrar, Straus and Giroux, 2010.

Serafimova, Silviya. "Whose Morality? Which Rationality? Challenging Artificial Intelligence as a Remedy for the Lack of Moral Enhancement." *Humanities and Social Sciences Communications* 7, no. 119 (2020): 1-10.

Sullins, John P. "Automated Ethical Practical Reasoning: The Problem of Artificial Phronesis." In *Robophilosophy: Philosophy of, for, and by Social Robotics*, edited by J. Seibt, R. Hakli, and M. Nørskov. MIT Press, 2025, forthcoming.

Sutton, Richard S., and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 2018.

Syse, Henrik, and Martin Cook. "Robotic Virtue, Military Ethics Education, and the Need for Proper Storytellers." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 667-680.

Tasioulas, John. "First Steps Towards an Ethics of Robots and Artificial Intelligence." *Journal of Practical Ethics* 7, no. 1 (2019): 61-95.

Tasioulas, John. "The Rule of Algorithm and the Rule of Law." Lecture at the University of Vienna, October 15, 2021.

Valenzuela, Pia. "Fredrickson on Flourishing through Positive Emotions and Aristotle's Eudaimonia." *Conatus – Journal of Philosophy* 7, no. 2 (2022): 37-61.

Yudkowsky, Eliezer. "Complex Value Systems in Friendly AI." In *Artificial General Intelligence*, edited by J. Schmidhuber, K. Thč risson, and M. Looks, 388-393. Springer, 2011.

Yudkowsky, Eliezer. "The Value Loading Problem." *EDGE*, July 12, 2021. https://www.edge.org/response-detail/26198.

Zetzsche, Dirk A., Ross P. Buckley, Janos N. Barberis, and Douglas W. Arner. "Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation." *Journal of Corporate and Financial Law* 23, no. 1 (2017): 31-103.

A. Harillo Pla · *Conatus* 10, no. 1 (2025): 153-163
doi: https://doi.org/10.12681/cjp.37227

Robo-Eroticism: Designing Desire via Creativity in Sexual Robots

Adrià Harillo Pla
Independent scholar, Spain
E-mail address: adria.harillo@gmail.com
ORCID iD: https://orcid.org/0000-0002-4005-9643

Abstract

In this article, I argue that the integration of creativity into the development of sexual robots is a vital and indispensable factor for their widespread acceptance and functionality. I contend that creativity serves as a linchpin in the realization of fully functional sexual robots. To substantiate this assertion, I present four key stages. In the first stage, I offer an in-depth analysis of the contemporary state-of-the-art in sexual robots, encompassing both their current practical implementations and the theoretical foundations that underpin them. The second stage delves into the intricate connection between sexuality and creativity, drawing on diverse case scenarios that exemplify the interdependence of these domains as a condition of possibility. The third stage underscores the paramount importance of creativity in the design and development of sexual robots, elucidating its strategic role in securing broad societal acceptance. In the fourth and final stage, I discuss the imperative need to establish a new interdisciplinary field of study, which I propose to term "Robo-Eroticism." This domain promises to comprehensively explore the fusion of technology and human sexuality from a creative perspective. By traversing these four stages, this article unveils the pivotal role of creativity in shaping the future evolution of sexual robots.

Keywords: *sex robots; human-robot interaction; sensuality; ambiguity; human-centered creativity*

I. From imagination to reality: Sexual robots nowadays

Regardless of opinions, the topic of sexual robots is present in our society. In the creative field of science fiction, humanoid robots which interacted with humans existed for decades. A scary popular example is The Terminator. Johnny Five is another example, friendly in this case. However, nowadays, robots are much more pres-

ent in our lives than ever before, leaving the exclusive realm of imagination.[1]

The cheaper production, the improvement of hardware and software, a society which every time is more used to human-robot interactions, and the Internet of Things played a significant role in it.[2] As Hans Peter Moravec puts it, a robot is "any automatically operated machine that replaces human effort, though it may not resemble human beings in appearance or perform functions in a humanlike manner."[3] This is a general definition of a robot. Nevertheless, I am focusing here exclusively on those robots which have anthropomorphic characteristics and perform functions in a human-like manner. That is, it: humanoid robots.

I admit, however, two factors. The first one, is that there are robots which could satisfy sexual desires without the need of resembling humans, as in the case of certain paraphilias, zoophilia or objectophilia being just two examples.[4] I also recognize that the word "effort" from the definition of "robot" can be contradictory to those who consider a sexual activity to take place if, and only if, all the parts involved in the sexual activity are receiving the same sexual gratification.[5] However, sexual abuse, oral sex, or masturbation, can still be considered sexual practices, although the intention behind them is not to have reciprocal sexual pleasure.[6] Therefore, sex can sometimes be considered an effort motivated by different reasons for at least one of the parts involved. As a consequence, I consider sexual practices in which all the agents involved receive the same sexual gratification to be a possibility, but not a necessity.

Sexual humanoid robots represent, therefore, one of the many possible kinds of human-robot interaction.[7] Precisely because of this, coun-

---

[1] International Federation of Robotics, "Robots in Daily Life: The Positive Impact of Robots on Wellbeing," 2021, https://ifr.org/papers/robots-in-daily-life-information-paper.

[2] Sachin Kumar et al., "Internet of Things Is a Revolutionary Approach for Future Technology Enhancement: A Review," *Journal of Big Data* 6 (2019): 1.

[3] Hans Peter Moravec, "Robot," *Britannica*, January 9, 2025, https://www.britannica.com/technology/robot-technology.

[4] Andrea M. Beetz and Anthony L. Podberscek, eds., *Bestiality and Zoophilia: Sexual Relations with Animals* (Purdue University Press, 2005); Melanie Weixler and Herwig Oberlerchner, "Objektophilie – Die Liebe zu Dingen," *psychopraxis. neuropraxis* 21, no. 5 (2018): 210-213.

[5] Chantelle Ivanski and Taylor Kohut, "Exploring Definitions of Sex Positivity Through Thematic Analysis," *The Canadian Journal of Human Sexuality* 26, no. 3 (2017): 216-225; Zoë D. Peterson and Charlene L. Muehlenhard, "What Is Sex and Why Does It Matter? A Motivational Approach to Exploring Individuals' Definitions of Sex," *The Journal of Sex Research* 44, no. 3 (2007): 256-268.

[6] Alan H. Goldman, "Plain Sex," *Philosophy & Public Affairs* 6, no. 3 (1977): 270.

[7] Robbie Arrell, "Sex and Emergent Technologies," in *The Routledge Handbook of Philosophy of Sex and Sexuality*, eds. Brian D. Earp, Claire Chambers, and Lori Watson (Routledge, 2022), 586.

tries such as Spain and Russia have begun to open prototype brothels with robots.[8] A TV series like *Better Than Us* presented the potential impact of this kind of robot in society. From a production side, Abyss Creations created the RealDoll.[9] Another company providing this kind of humanoid robot is Sex Doll Genie, which offers them in different looks, genders, ages, and sizes.[10] In Spain, the robots Samantha or Ava are alike.[11] This new scenario, did not interest only companies who started obtaining money from it, but also the academicians. A proof of this is the organization of the *Congress on Love and Sex with Robots*, from which more than nine editions were held up to date. Books like *Love and Sex with Robots*, or *Turned On: Science, Sex and Robots* are just another example.[12] Nevertheless, like many other innovations, this one also finds its resistance.[13] An example of organized resistance is the campaign launched in 2015, led by Kathleen Richardson and Erik Billing, called *The Campaign Against Sex Robots*.[14]

Often, the opinion regarding this type of robot depends on the perceived and imagined consequences, both in the robot-human interaction and in its externalities for human-human interaction.[15] Different expectations lead to different conclusions regarding whether this type of robot is possible, salutary, and desirable.[16]

---

[8] Alfonso L. Congostrina, "Trouble in Spain's Uncanny Valley of the Sex Dolls?" *El País*, March 15, 2017, https://english.elpais.com/elpais/2017/03/15/inenglish/1489581889_495823.html; Will Stewart, "Russia's First Sex Robot Brothel Opens Ahead of World Cup," *Daily Mail Online*, May 10, 2018, http://www.dailymail.co.uk/news/article-5713369/Russias-sex-robot-brothel-opens-ahead-World-Cup-bid-cash-fans-players.html.

[9] Joan Alvado, "Sex and Love with Robots: No Longer Science Fiction," *Equal Times*, July 14, 2017, https://www.equaltimes.org/sex-and-love-with-robots-no-longer; Ry Crist, "Behind the Scenes of a Sexbot Factory," *CNET*, August 10, 2017, https://www.cnet.com/pictures/sex-robots-sexbots-abyss-creations-factory-realdoll-harmony/.

[10] "About Sex Dolls by SDG," Sex Doll Genie, effective 2022, https://sexdollgenie.com/pages/about-us.

[11] Alvado, "Sex and Love with Robots."

[12] David Levy, *Love and Sex with Robots: The Evolution of Human-Robot Relationships* (HarperCollins, 2007); Kate Devlin, *Turned On: Science, Sex and Robots* (Bloomsbury Sigma, 2018).

[13] Sven Heidenreich and Patrick Spieth, "Why Innovations Fail – the Case of Passive and Active Innovation Resistance," *International Journal of Innovation Management* 17, no. 5 (2013): 1; Thomas S. Kuhn, *The Structure of Scientific Revolutions*, ed. Ian Hacking (University of Chicago Press, 2012).

[14] Andrea Morris, "Meet The Activist Fighting Sex Robots," *Forbes*, September 27, 2018, https://www.forbes.com/sites/andreamorris/2018/09/26/meet-the-activist-fighting-sex-robots/.

[15] Noel Sharkey, Aimee van Wynsberghe, Scott Robbins, and Eleanor Hancock, "Our Sexual Future with Robots: A Foundation for Responsible Robotics Consultation Report," *Foundation for Responsible Robotics*, July 1, 2017.

[16] Georgios Arabatzis, "Pornography and Stress," *Conatus – Journal of Philosophy* 7, no. 2 (2022): 145-146; Anco Peeters and Pim Haselager, "Designing Virtuous Sex Robots," *Interna-*

Whether these robots can be socially salutary is a topic which must be answered from an ethical perspective.[17] However, human-robot interaction is often characterized by safety, usability, and functionality. My statement is that for a sexual robot to be desirable, and fully functional, only achieving creative behavior is a condition of necessity, although not of sufficiency.

II. Methodology

In this article, I employ a qualitative research approach centered on the analysis of secondary data to investigate the intricate interplay between sexual robots and creativity. The rationale for choosing this method is grounded on the nature of the research question, the accessibility of existing literature, and the potential to synthesize and reinterpret established knowledge within the topic.

My decision to utilize a qualitative analysis of secondary data was driven by several factors. First, the research question is exploratory and conceptually focused, seeking to delve into the complexities of a relatively novel area of study. Second, there is a significant amount of scholarly work on social robotics, human-robot interactions, and human-centered creativity, making secondary data readily accessible. Lastly, this method allows for the integration of diverse perspectives from multidisciplinary sources, enabling a comprehensive exploration of the research question.

To undertake the aforementioned qualitative analysis, I performed a systematic literature review and data compilation. I conducted a comprehensive search across reputable academic databases, encompassing peer-reviewed articles, books, conference proceedings, and relevant reports. I meticulously selected relevant materials based on their alignment with the research topic and the conceptual framework.

The analysis entailed a multi-step process. Initial data collection involved identifying and categorizing sources that provided insights into social robots in general, and sexual robots in particular. As well, in human sexuality, and creativity. Subsequently, I employed a thematic analysis approach, focusing on identifying recurring themes, concepts, and patterns across the selected sources.

---

tional Journal of Social Robotics 13 (2021): 64; Harro Van Lente, "Imaginaries of Innovation," in *Handbook on Alternative Theories of Innovation*, eds. Benoît Godin, Gérald Gaglio, and Dominique Vinck, 23-36 (Edward Elgar Publishing, 2021).

[17] Arrell, "Sex and Emergent Technologies," 586.

The methodology that I have chosen aligns coherently with the research objectives and the nature of the research question. By leveraging secondary data, this methodology provides a cost-effective and efficient means of addressing the research question, particularly considering the exploratory nature of the investigation. At the same time, it reduces the ethical concerns derived from human experimentation. This methodology enables us to derive meaningful conclusions, generate new theoretical insights, and provide a foundation for future empirical research endeavors in this emergent and thought-provoking domain.[18]

III. Sex and creativity

Human sexuality can be a complex activity. We could define a sexual activity as an action that tends to satisfy sexual desire via sexual pleasure for at least one of the members involved in the activity. This does not imply excessive creativity by default. However, some sexual practices do. Sexual role-plays are one example.

Defining creativity is not a simple task, and differences between scholars and disciplines often show this. To avoid semantic disputes, it is important to clarify what I mean by "creativity." By creativity, I refer to the "tendency to generate or recognize ideas, alternatives, or possibilities that may be useful in solving problems, communicating with others, and entertaining ourselves and others."[19] In fact, in the following chapters, I will present this tendency, and particularly, its linked skill, as a key. Key to generating useful alternatives for solving the communication and entertainment problems that sometimes arise within sexual interactions. Sexual human interactions which require "for novel, varied, and complex stimulation."[20]

While I acknowledge that the philosophical notion of creativity is often related to that of personal identity, I do not consider this as strictly necessary.[21] In the field of artificial intelligence, the concept of generative intelligence, and especially, the one of combinational creativity, are questioning that notion of the self as a condition of necessity.[22] Applied to robotics, I consider that the upcoming book of

---

[18] Kerstin Dautenhahn, "Methodology & Themes of Human-Robot Interaction: A Growing Research Field," *International Journal of Advanced Robotic Systems* 4, no. 1 (2007): 15.

[19] Robert E. Franken, *Human Motivation* (Brooks/Cole Publishing Company, 1998), 396.

[20] Ibid.

[21] Nikos Erinakis, "What Makes Free Will Free: The Impossibility of Predicting Genuine Creativity," *Conatus – Journal of Philosophy* 5, no. 1 (2020): 55-69.

[22] Giancarlo Frosio, "The Artificial Creatives: The Rise of Combinatorial Creativity from Dall-E to

Iva Apostolova approaching the importance of touch in the process of having mental representations and how this could impact robots, will contribute positively to this meaningful topic.

Having clarified this fact, it is time to state that individuals can engage sexually in scenarios where they take on different identities creating intricate narratives. These imaginative activities add novelty to intimate encounters and allow individuals to explore their desires and fantasies in a safe and consensual setting. The understanding of this situation is often ambiguous and non-literal, which seems to be a potential conflict for the safety, usability, and functionality of a sexual robot nowadays.

An aesthetic and creative understanding of the world at the same time transforms engagement into sexual activities. One illustrative example is eroticism and sensuality linked to the concept of nude, in comparison with the naked. While the naked body is a state of undress, vulnerability, and even embarrassment, the concept of the nude suggests other things. Things such as beauty, sensuality, touch, warmth, and acceptance of the body. The nude, static or in motion, is creative since it generates a pleasant engagement.[23]

In addition, creative communication within intimate relationships has the potential to generate sensuality and eroticism.[24] When partners engage in imaginative and artful conversations, they create a unique mental and emotional connection that is deeply stimulating. The exchange of seductive words, fantasies, and desires, through creative dialogues, allows individuals to tap into sensuality. By weaving stories and using language to evoke vivid mental images, they tease each other and craft a shared narrative of desire. This creative communication serves as a powerful aphrodisiac, setting the stage for a more profound and electrifying physical connection.

Humans seeking to engage in sexual experiences often engage in creative practices. Whether it is the introduction of new scenarios, experimentation with verbal communication, or immersive role-playing, these endeavors reflect a commitment to novelty, personal connection, and the exercise of the imagination to foster deeper intimacy.[25]

ChatGPT," in *Handbook of Artificial Intelligence at Work: Interconnections and Policy Implications*, eds. Martha Garcia-Murillo, Ian MacInnes, and Andrea Renda (Edward Elgar Publishing, 2024), 240-245.

[23] Georges Bataille, *Eroticism*, trans. Mary Dalwood (John Calder, 1962); Kenneth Clark, *The Nude: A Study of Ideal Art* (John Murray, 1956); John Money, "The Development of Sexuality and Eroticism in Humankind," *The Quarterly Review of Biology* 56, no. 4 (1981): 379-404.

[24] Maggie Geuens and Patrick De Pelsmacker, "Affect Intensity Revisited: Individual Differences and the Communication Effects of Emotional Stimuli," *Psychology & Marketing* 16, no. 3 (1999): 195.

[25] Tabea M. Zorn, Lena Feilhauer, Elina Juhola, Kaja Gottwald, and Charmaine Borg, "Does Sexual Creativity Enhance Sexual Satisfaction? Examining the Effect of Weekly Creative Sexual

As a result, I support that a relationship between sexual practices and creativity exists. I state that this relationship is based on a condition of possibility, and not of necessity nor sufficiency. However, when applied in a consensual activity, creativity seems to have a positive effect on the satisfaction of sexual desire via sexual pleasure, thanks to its generation of eroticism and sensuality.

IV. Sexual robots and creativity

My conclusion is that the concept of functionality in sexual robots is undeniably important, as it forms the very core of their purpose. However, I recognize the critical importance of other factors, such as safety and usability. The research suggests that creativity serves as a bridge with the potential to connect functionality to a wider audience, making the interaction with sexual robots resemble a qualitatively expected human-robot interaction with sexual purposes.

Creativity has the power to enrich the user experience by enhancing the aesthetic and psychological engagement between humans and these robots. It has the potential of enabling sexual robots to adapt, surprise, and evolve, ensuring that the experience remains fresh, engaging, and satisfying. This adaptability is crucial for the long-term success of sexual robots, as it prevents users from growing bored or disinterested over time, and therefore, prevents its potential decline.[26] Without creativity, sexual robots risk becoming mere mechanical devices, unable to provide the nuanced responses and interactions that their users desire.

Furthermore, creativity opens doors to diverse and inclusive experiences, catering to a broader range of user preferences. It allows for the customization of personalities, appearances, and communication styles, which can help users feel more connected and engaged with their robotic partners. In doing so, creativity is pivotal in addressing various human needs and desires, making sexual robots an inclusive and versatile option for users of different backgrounds, cultures, and orientations.

Overall, the integration of creativity into the design and functionality of sexual robots should not be understood as a luxury. It has

---

Tasks in Monogamous Long-Term Heterosexual Couples," *The Journal of Sexual Medicine* 19, suppl. 4 (2022): S114.

[26] Zahar Koretsky, Harro Van Lente, Bruno Turnheim, and Peter Stegmaier, "Introduction: The Relevance of Technologies in Decline," in *Technologies in Decline: Socio-Technical Approaches to Discontinuation and Destabilisation*, eds. Zahar Koretsky, Peter Stegmaier, Bruno Turnheim, and Harro Van Lente (Routledge, 2022), 7.

the potential to be the key to expanding their reach and ensuring that they remain relevant and desirable companions for humans. By offering unique, evolving, and emotionally enriching experiences, sexually creative robots can genuinely engage and connect with a broader audience, fulfilling their promised functionality.

V. In praise of a "Robo-Eroticism"

In light of the importance of creativity and its significant role in enhancing the functionality of sexual robots, I believe there is a compelling case for creating a dedicated subfield within the broader realm of human-robot interaction. This subfield, which I propose to be termed "Robo-Eroticism." This would focus on the interdisciplinary study and development of sexual robots with an emphasis on creative, psychological, and esthetic engagement.

The emergence of "Robo-Eroticism" as a distinct subfield is a response to the growing interest and investment in the development of sexual robots, and the recognition that these artificial companions are more than mere tools for sexual gratification, or at least, they have the potential to become more than that. Rather, they represent a potential paradigm shift in how humans experience sexual practices with non-biological entities. Therefore, I consider it is imperative that we approach this field with a more comprehensive perspective. A perspective that goes beyond the technical and ethical aspects, while expressing all the recognition, respect, and interest for the technical and ethical contributions to this topic.

"Robo-Eroticism" would involve experts in fields such as artificial intelligence, robotics, psychology, human-computer interaction, ethics, aesthetics, and sexuality studies. By fostering collaboration among experts from these diverse backgrounds, this subfield can address the multifaceted challenges and opportunities presented by sexual robots.

Within this subfield, researchers and developers can explore how creativity can be used to improve the user experience. In doing so, they can pave the way for a future where sexual robots serve as not just tools for pleasure, but as sensual and erotic entities for human engagement.

Furthermore, the establishment of "Robo-Eroticism" could also serve to legitimize and regulate the industry by setting standards for manufacturers and designers ensuring that sexual robots are developed with a creativity which is human-centered, since it has the goal to appeal to the human sensitivity. This could help mitigate potential risks and controversies surrounding the technology and promote its acceptance in society.

In conclusion, the integration of creativity and the importance of user engagement in the context of sexual robots are compelling reasons to consider the development of a dedicated subfield, "Robo-Eroticism." By creating this subfield, we acknowledge the multifaceted nature of sexual robotics and work toward harnessing technology to create meaningful, emotionally fulfilling human-robot relationships while addressing the ethical and practical challenges they present. "Robo-Eroticism" could pave the way for a creative, responsible, and ethical future for this emerging field.

Acknowledgements

I would like to express my sincere gratitude to the scholars who have contributed to the body of knowledge surrounding this topic, both directly and indirectly. Their collective efforts have enriched my understanding and paved the way for meaningful discussions and advancements in this field. I hereby declare that I have no conflicts of interest or financial disclosures to report in relation to the subject discussed in this article. There are no affiliations or financial involvements that might be perceived as affecting the objectivity or integrity of the information presented.

References

Alvado, Joan. "Sex and Love with Robots: No Longer Science Fiction." *Equal Times*, July 14, 2017. https://www.equaltimes.org/sex-and-love-with-robots-no-longer.

Arabatzis, Georgios. "Pornography and Stress." *Conatus – Journal of Philosophy* 7, no. 2 (2022): 143-156.

Arrell, Robbie. "Sex and Emergent Technologies." In *The Routledge Handbook of Philosophy of Sex and Sexuality*, edited by Brian D. Earp, Claire Chambers and Lori Watson. Routledge, 2022.

Bataille, Georges. *Eroticism*. Translated by Mary Dalwood. John Calder, 1962.

Beetz, Andrea M., and Anthony L. Podberscek, eds. *Bestiality and Zoophilia: Sexual Relations with Animals*. Purdue University Press, 2005.

Clark, Kenneth. *The Nude: A Study of Ideal Art*. John Murray, 1956.

Congostrina, Alfonso L. "Trouble in Spain's Uncanny Valley of the Sex Dolls?" *El País*, March 15, 2017. https://english.elpais.com/elpais/2017/03/15/inenglish/1489581889_495823.html.

Crist, Ry. "Behind the Scenes of a Sexbot Factory." *CNET*, August 10, 2017. https://www.cnet.com/pictures/sex-robots-sexbots-abyss-creations-factory-realdoll-harmony/.

Dautenhahn, Kerstin. "Methodology & Themes of Human-Robot Interaction: A Growing Research Field." *International Journal of Advanced Robotic Systems* 4, no. 1 (2007): 103-108.

Devlin, Kate. *Turned On: Science, Sex and Robots*. Bloomsbury Sigma, 2018.

Erinakis, Nikos. "What Makes Free Will Free: The Impossibility of Predicting Genuine Creativity." *Conatus – Journal of Philosophy* 5, no. 1 (2020): 55-69.

Franken, Robert E. *Human Motivation*. Brooks/Cole Publishing Company, 1998.

Frosio, Giancarlo. "The Artificial Creatives: The Rise of Combinatorial Creativity from Dall-E to ChatGPT." In *Handbook of Artificial Intelligence at Work: Interconnections and Policy Implications*, edited by Martha Garcia-Murillo, Ian MacInnes, and Andrea Renda. Edward Elgar Publishing, 2024.

Geuens, Maggie, and Patrick De Pelsmacker. "Affect Intensity Revisited: Individual Differences and the Communication Effects of Emotional Stimuli." *Psychology & Marketing* 16, no. 3 (1999): 195-209.

Goldman, Alan H. "Plain Sex." *Philosophy & Public Affairs* 6, no. 3 (1977): 267-287.

Heidenreich, Sven, and Patrick Spieth. "Why Innovations Fail – the Case of Passive and Active Innovation Resistance." *International Journal of Innovation Management* 17, no. 5 (2013): 1-42.

International Federation of Robotics. "Robots in Daily Life: The Positive Impact of Robots on Wellbeing." 2021. https://ifr.org/papers/robots-in-daily-life-information-paper.

Ivanski, Chantelle, and Taylor Kohut. "Exploring Definitions of Sex Positivity Through Thematic Analysis." *The Canadian Journal of Human Sexuality* 26, no. 3 (2017): 216-225.

Koretsky, Zahar, Harro Van Lente, Bruno Turnheim, and Peter Stegmaier. "Introduction: The Relevance of Technologies in Decline." In *Technologies in Decline: Socio-Technical Approaches to Discontinuation and Destabilisation*, edited by Zahar Koretsky, Peter Stegmaier, Bruno Turnheim, and Harro Van Lente. Routledge, 2022.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*, edited by Ian Hacking. University of Chicago Press, 2012.

Kumar, Sachin, Prayag Tiwari, and Mikhail Zymbler. "Internet of Things Is a Revolutionary Approach for Future Technology Enhancement: A Review." *Journal of Big Data* 6 (2019): 1-21.

Levy, David. *Love and Sex with Robots: The Evolution of Human-Robot Relationships*. HarperCollins, 2007.

Money, John. "The Development of Sexuality and Eroticism in Humankind." *The Quarterly Review of Biology* 56, no. 4 (1981): 379-404.

Moravec, Hans Peter. "Robot." *Britannica*, January 9, 2025. https://www.britannica.com/technology/robot-technology.

Morris, Andrea. "Meet The Activist Fighting Sex Robots." *Forbes*, September 27, 2018. https://www.forbes.com/sites/andreamorris/2018/09/26/meet-the-activist-fighting-sex-robots/.

Peeters, Anco, and Pim Haselager. "Designing Virtuous Sex Robots." *International Journal of Social Robotics* 13 (2021): 55-66.

Peterson, Zoë D., and Charlene L. Muehlenhard. "What Is Sex and Why Does It Matter? A Motivational Approach to Exploring Individuals' Definitions of Sex." *The Journal of Sex Research* 44, no. 3 (2007): 256-268.

Sex Doll Genie. "About Sex Dolls by SDG." Effective 2022. https://sexdollgenie.com/pages/about-us.

Sharkey, Noel, Aimee van Wynsberghe, Scott Robbins, and Eleanor Hancock. "Our Sexual Future with Robots: A Foundation for Responsible Robotics Consultation Report." *Foundation for Responsible Robotics*, July 1, 2017.

Stewart, Will. "Russia's First Sex Robot Brothel Opens Ahead of World Cup." *Daily Mail Online*, May 10, 2018. http://www.dailymail.co.uk/news/article-5713369/Russias-sex-robot-brothel-opens-ahead-World-Cup-bid-cash-fans-players.html.

Van Lente, Harro. "Imaginaries of Innovation." In *Handbook on Alternative Theories of Innovation*, edited by Benoît Godin, Gérald Gaglio, and Dominique Vinck. Edward Elgar Publishing, 2021.

Weixler, Melanie, and Herwig Oberlerchner. "Objektophilie – Die Liebe zu Dingen." *psychopraxis. neuropraxis* 21, no. 5 (2018): 210-213.

Zorn, Tabea M., Lena Feilhauer, Elina Juhola, Kaja Gottwald, and Charmaine Borg. "Does Sexual Creativity Enhance Sexual Satisfaction? Examining the Effect of Weekly Creative Sexual Tasks in Monogamous Long-Term Heterosexual Couples." *The Journal of Sexual Medicine* 19, suppl. 4 (2022): S114.

A. Mullins · *Conatus* 10, no. 1 (2025): 165-191
doi: https://doi.org/10.12681/cjp.36665

What Does Self-control Look Like? Considerations about the Neurobiology of Temperance and Fortitude

Andy Mullins
University of Notre Dame Australia, Australia
E-mail address: apjm.vic@gmail.com
ORCID iD: https://orcid.org/0000-0003-1540-3796

Abstract

Our subject is the neurobiological characteristics of virtuous emotional responses and their integration into character. Drawing raw material from the self-reported thoughts and actions of Dr Takashi Nagai, present in Nagasaki at the time of the atomic bomb, our methodology is to conduct a highly granular examination of one specific moment where his self-control is greatly tested. This permits us then to offer an analysis of the neurobiological processes, pathways, and systems that underpin the management of emotional reactions, and in effect, draw insight into the neurobiology of virtues of temperance and fortitude, understood from an Aristotelian-Thomistic perspective. The neurobiological considerations are preceded by discussion of philosophical prerequisites founded on a Thomistic metaphysics of participation. In conclusion we offer some thoughts about the benefits of neurobiological investigations in relation to character and the aptness of the development of virtue for human beings.

Keywords: *Aristotelian-Thomistic virtue; emotion; neurobiology; self-control; atomic bomb; temperance; fortitude*

I. Introduction

Emotions are neither good nor bad in themselves, they are our responses to sense apprehensions. We are able to draw great good from them; they mediate relationships and our understanding of the world, and they are our primary motivators, but they require integration into our rationality. This integration is facilitated by the virtues, good habits of

responding to pain, pleasure, and the rights of others. Aristotle held that it was by these habits, the moral virtues, that "We are directed well or ill in reference to the passions."[1] The Aristotelian-Thomistic doctrine of the virtues offers a model for the successful integration of emotions, which it does not distinguish from passions, into character.[2]

The classical pre-eminence of the four cardinal virtues undergoes remarkable development in Aquinas's repackaging of Aristotle's vision of virtue within a psychology recognising essential powers each needing their own perfection by habit if we are to flourish: "Every power which may be variously directed to act, needs a habit whereby it is well disposed to its act."[3] These powers consist of the concupiscible and irascible sense appetites and the intellectual powers of knowing and choosing, and they each require their specific perfecting habit.[4] These habits are the cardinal virtues. When the sense appetites are activated, an emotion is manifested; emotions are a response to sense apprehensions of whatever we find pleasant or arduous. The habits perfecting the sense appetites are temperance and fortitude, and those perfecting intellectual operations are prudence, and justice which Aquinas identifies as the virtue of the will, wherein every choice must take into account the rights of others.

When our emotional responses have become habitually malleable to reason, we are said to have developed the virtues of temperance and fortitude. Temperance is the habitual pursuit of pleasures reasonably assessed as good for us, and fortitude, the habitual readiness to overcome fear of difficulty for a good reason. These virtues are seen as the engine room of self-control.

For this discussion of the neurobiology of virtue I adopt this Aristotelian-Thomistic understanding of virtue. In distinguishing between sense appetites relating to pleasure and pain, and intellectual powers, possibilities open for a neurobiological discussion of the neural reformation of those appetites.

This doctrine has proven remarkably resilient as a model for human personality. Not only does it appear to accord with human experience, and the wisdom of religious and secular literature, but it is evident across all cultures. Classical virtue is now incarnated in contemporary virtue ethics and contemporary positive psychology, whereby the neurobiological systems and

---

[1] Aristotle, *Nicomachean Ethics*, II.5,1105b29.

[2] As nuances in the nature of emotion are not the focus of this paper, I will adopt an Aristotelian focus on emotion, without entering into contemporary debates about the nature of emotion.

[3] Aquinas, *Summa Theologica* (*S.T.*), I II 50.5.

[4] Ibid., I II 61.2. Previously he has explained: "Every power which may be variously directed to act, needs a habit whereby it is well disposed to its act." (*S.T.*, I II 50.5). The four powers of concupiscible and irascible appetites, rational will, and intellect, all require their specific perfecting habit. These habits are the cardinal virtues.

processes of virtuous behaviours are being identified. This will constitute the focus for this paper. However, my starting point will be close observation of a one precise moment of heightened emotion and subsequent self-control. By this approach, I hope to reinforce the conviction that virtues are not theoretical, albeit coherent, constructs for understanding behaviour, but manifest objective attributes of the person that facilitate behaviours which are good for us as human beings.

II.

a. Takashi Nagai (1908-1951)

We anchor our discussion of self-control in the real world of sensation, passion, apprehension, deliberation, choice and action of Dr Takashi Nagai, in the early days of August 1945 in Nagasaki. We recall Elizabeth Anscombe's advice that, before we lose ourselves in ethical theory, we must account for the richness and complexity of real people who manifest the tension between rationality and emotion, free and impulsive behaviours, intrinsic and extrinsic motivation, and fulfilled and unfulfilled lives.[5]

Nagai is an exemplary figure; a medical doctor and nuclear physicist who, in the years after WWII, through his writings under the most difficult circumstances, was a great force for spiritual healing in a country gutted by the horrors of war.[6] Nagai's response to the cataclysmic event of August 9, 1945, led to his subsequent fame. His personal account of the immediate aftermath of the blast in Nagasaki became a best seller in Japan, and the subject of movie and popular song. He wrote some twenty books in the six years after the war.

Nagai demonstrated that he was a man of deep convictions and intelligence, concern for his fellow man, austerity of life, and courage. Greatly influenced by Pascal's *Pensées*, by his wife and perhaps also by his experiences with the Imperial Army during its invasion of China, Nagai had become a Christian before the war. He was a devoted father and compassionate doctor; zealous for the advancement of medicine, he carefully documented the effects and effective treatments of radiation illness. Through all his suffering he manifested no bitterness towards the Americans for the bombing that destroyed millions of lives, and in some estimates, half the urban areas of Japan. In post war Japan during the six years as he lay dying of leukemia, Nagai

---

[5] Elizabeth Anscombe, "Modern Moral Philosophy," in *Virtue Ethics*, eds. Roger Crisp and Michael Slote (Oxford University Press, 1997), 26.

[6] Introductory texts for an understanding of Takashi Nagai's life: Takashi Nagai, *The Bells of Nagasaki,* trans. William Johnston (Kodansha America Inc, 1994); original title: *Nagasaki no kane* (1949). Paul Glynn, *A Song for Nagasaki* (Marist Fathers Books, 1988).

rose to prominence as a national spiritual leader as much for his extraordinary writings encouraging reconciliation and peace, as for his remarkable humility and inner peace.

> In the postwar years Nagai became a symbol of strength and optimism, a central figure in the spiritual and moral reconstruction of Japan. [...] His influence on the collective unconscious of Japan was very great.[7]

He was praised in the contemporary press as the "Gandhi of Japan." The nobility of Nagai's mature, considered actions and the richness of his emotional life typify what most would agree to be a state of virtue.

b. 11.02 am August 9, 1945

On the morning the bomb fell, Nagai was at work in his bunkered laboratory, 500m from the epicentre, choosing X-ray films to teach students the art of diagnosis. Without warning there was a flash of blinding light. With observational skills honed by scientific training, Nagai described his own experience:

> I immediately tried to throw myself to the ground, but before I could do so, the glass of the windows smashed in and a frightening blast of wind swept me off my feet into the air – my eyes wide open. Pieces of broken glass came in like leaves blown off a tree in a whirlwind. I felt that the end had come. [...] It was as though a huge invisible fist had gone wild and smashed everything in the room. The bed, the chairs, the bookcases, my steel helmet, my shoes, my clothes were thrown into the air, hurled around the room with a wild clattering noise, and all piled on top of me as I lay helpless on the floor. Then the blast of dusty dirty wind rushed in and filled my nostrils so I could scarcely breathe. I kept my eyes open, looking always at the window. And as I looked everything outside grew dark. There as a noise like a stormy sea, and the air everywhere swirled round and round. My clothes, the zinc roof, pieces of wood, and all kinds of other objects were performing a macabre dance in that dark sky. Then it gradually became cold, as at the end of autumn, and a strange and silent emptiness ensued. Clearly this was no ordinary event.[8]

---

[7] William Johnston in the preface of Nagai, *Bells,* xx and xxii.

[8] Nagai, *Bells,* 11.

Everywhere there were the dead and dying, convulsing, strangely swollen, skin peeling. Soon fires were raging. Weakened by previously contracted leukemia, and despite his own grave injuries, Nagai worked to utter exhaustion caring for the dying and injured through the day of the blast and the days that followed. He had to improvise everything. Only after three days did he return to his family home finding the charred bones and melted rosary of his beloved wife, Midori.

c. Tested self-control

Nagai's capacity to describe with precision internal states and subjective responses, as well as external events, make his writing ideal for the purposes of this study. On the day following the blast, he describes one moment in which, despite exhaustion and discouragement, he demonstrates the process of mastering impulsive emotion. In this excerpt he captures the moment when the reality dawns upon him that the destruction was wrought by a nuclear device and consequently Japan must be defeated:

> The chief nurse came running up and handed me a sheet of paper. It was one of the leaflets dropped by enemy planes the previous night. As I glanced at it I shouted out spontaneously: "The atomic bomb!" In the depth of my being I felt a tremendous shock. The atom bomb has been perfected! Japan is defeated! [...] Conflicting emotions churned in my mind and heart as I surveyed the appalling atomic wasteland around me. [...] A bamboo spear lay on the ground. I kicked it fiercely and it made a dull, hollow sound. Grasping it in my hand, I raised it to the sky, as tears rolled down my cheeks. The bamboo spear against the atomic bomb! What a tragic comedy this war was! This was no longer a war. Would we Japanese be forced to stand on our shores and be annihilated without a word of protest? These are the words written on the leaflet:
>
> *To the People of Japan*
> *Read carefully what is written in this leaflet. The United States has succeeded in inventing an explosive more powerful than anything that has existed until now. The atomic bomb now invented has a power equal to the bomb capacity of two thousand huge B-29s. You must reflect seriously on this terrible fact. We swear that what we say here is the solemn truth. [...] The President of the United States has already given you an outline of thirteen conditions for an honourable surrender. We advise you to accept these conditions and to being rebuilding a new and better peace-*

> *loving Japan. [...] If you do not do this, we are determined to use this bomb and other excellent weapons to bring this war to a swift, irresistible conclusion.*
>
> I read the leaflet once and was stunned. I read it a second time and felt they were making fools of us. I read it a third time and was enraged at their impudence. But when I read it a fourth time I changed my mind and began to think it was reasonable. After reading it a fifth time I knew that this was not a propaganda stunt but the sober truth.[9]

This moment brings the thunderclap insight that the war is now inevitably lost. It is interwoven with a sense of the burning shame, associated with any defeat, inculcated by his culture since childhood. Nagai's description of how his passionate reaction subsides on successive readings of the leaflet gives us a remarkable insight into how deliberation can enable mastery of passion. What is initially less obvious is the interior battle that Nagai has to fight in order to respond rationally to the news. Had he torn up the leaflet after the first readings, he would not have come to the same conclusion. Unsaid too is any reference to his education and upbringing that empowered him to exhibit the self-control required to allow the news to sink in.

His acceptance of the truth about the bomb is followed by very swift reasoning and insight that the Japanese defenders would be powerless on beaches against landings supported by atomic weapons. He then deliberates over whether he should accept the demand of the leaflet. Only in his fourth reading he "changes his mind" and now finds the words "reasonable." Grasp of one truth leads to deliberation and reasoning that leads to the grasp of another truth. During the period of deliberation he keeps his emotions under control sufficiently to continue his deliberation.

Complexity is further added by the fact that, although it is not explicitly stated in this passage, he is aware he has an audience. By his leadership and decision making throughout the previous day, he had already demonstrated an appreciation of his responsibility to those suffering around him. It is reasonable to surmise that this sense of responsibility as well as his military and scientific training assisted him in applying sufficient deliberation before committing himself to judgement.

d. Distilling temperance

In the example above, Nagai strives to harmonise his emotional responses with his reason; he chooses to react to his emotion in a way that permits

[9] Ibid., 52-53.

him to grasp the objective truth of the situation in which he finds himself. Only on the fourth and fifth readings did he accept the truth of what he was reading. This effort to confront something deeply unpalatable bespeaks qualities of character and a habitual restraint that withholds final judgement and continues deliberation without committing to an unbridled emotional response until a matter has been thoroughly considered. It illustrates well the capacity for passionate reactions schooled (or "conditioned," to use a term applicable to neural responses) to accept the guidance of reason, albeit with initial reluctance. We are witnessing the effects of the already acquired habitual disposition in Nagai's sensitive appetites to respond to his direction. Also, we see in his rational appetite, his capacity to make well deliberated choices responsive to the rights of others. This is the very stuff of virtue. Aristotle and Aquinas both hold that our appetites, both sensible and rational, need to have been trained if a person is to possess such restraint and self-control as we see in Nagai. Yet, every subsequent virtuous action, further reinforces the earlier disposition.

We witness the evident interplay of the cardinal virtues in this internal dialogue. Nagai draws on habitual management of impulse (temperance), as well as a habitual readiness to apply himself in a difficult situation (fortitude), along with a habitual sense of duty to countrymen and country (justice), and an openness to truth and readiness to reflect (prudence). All is at the service of actions informed by reality. In such a moral dissection, the unity of the virtues appears virtually self-evident.

By mastering his humiliation and anger to consider the implications of the Allied leaflet, Nagai demonstrates how fundamental to human existence are the dual challenges of curbing wayward internal passions and overcoming external difficulties in the pursuit of difficult goals.

Aristotle places the cultivation of man's passions at the very heart of human appetition: to seek pleasure and avoid pain. At this most elemental level, temperance in response to hedonism, and fortitude in response to fear and pain, are dispositions to self-mastery. He explains that pain and pleasure enter into *both* temperance and fortitude.

> The self-indulgent man craves for all pleasant things or those that are most pleasant, and is led by his appetite to choose these at the cost of everything else; hence he is pained when he fails to get them and when he is merely craving for them (for appetite involves pain).[10]

[10] Aristotle, *Nicomachean Ethics*, III.1, 1119a1-5.

And direction of our sensitive appetites is the very matter of daily life.[11] Aquinas stressed that these emotional responses are positively good but only if they are managed by reason.

> Emotion leads away from moral behaviour in so far as it is uncontrolled by reason; but in so far as it is rationally directed, it is part of the virtuous life.[12]

III. Metaphysical pre-requisites

A discussion of the biophysical bases of self-control should be conducted within a sufficient philosophical anthropology. Purely physicalist notions of rationality seem inadequate. Roger Scruton has noted,

> [T]here is a problem about accounting for rationality and the general difference between man and the other animals [...] in the end we need some kind of teleological metaphysics to make sense of our condition.[13]

Human beings have the capacity to act for ends that transcend sense experience. If this view is accepted, rational processes may not be reduced to electrical impulses, chemical processes, neural systems and pathways and brain regions. Human rationality may not be reduced to processes of reasoning, or to subjective inner life, qualia, consciousness, or other subjective manifestations. Indeed, John Haldane questions the possibility of consciousness studies offering a way forward in philosophy of mind.[14] Ultimately human maturity seems inseparable from the capacity to know reality and make choices, and to find fulfilment in interpersonal loving relationships, life self-directed intentionally for others. "Human beings have the capacity to understand their world, and at the same time, to stand above it, to seek fulfilment in elective, loving relationships with other rational beings."[15] This understanding of rationality seems essential, of the essence, to what it means to be human.

a. A Thomistic metaphysics of participation

It is beyond the scope of this paper to enter in any depth into the great variety of views on rationality. A teleological metaphysics of participation within

[11] Ibid., III.2, 1111b13-16.

[12] Aquinas, *S.T.*, I II 24.2.

[13] Andy Mullins, "Can Neuroscientific Studies Be of Personal Value?" *International Philosophical Quarterly* 57, no. 4 (2017): 444.

[14] John Haldane, "A Return to Form in the Philosophy of Mind," *Ratio* 11, no. 3 (1998): 253-277.

[15] Andy Mullins, "Philosophical Prerequisites for a Discussion of the Neurobiology of Virtue," *Ethical Perspectives* 23, no. 4 (2016): 689-708.

hylomorphic personalism offering an adequate philosophical underpinning for a discussion respecting both evident freewill and the evidence of neurobiology in supporting human activity is a Thomistic metaphysics of participation (TMP). Within TMP,

> Rationality should not be seen as a ghostly process exclusive of the world of matter, but rather as a transcendent process within matter itself by virtue of a participated power.[16]

I will offer an overview of TMP, drawing some contrasts with non-reductive physicalism and with hylomorphic approaches that manifest dualism.

Such a metaphysics of participation *in esse subsistens* has been absent from philosophy of mind and is found only, more broadly in Anglo-American Thomism, in the work of philosophers such as Norris Clarke, Koterski,[17] Wippel,[18] Hankey,[19] and Cullen.[20] The Thomistic argument for rationality commences with a grasp of the contingency of living things, and of rational subjects in particular. It is dependent upon the close attention to reality, the "complex materiality of things" alluded to by Martha Nussbaum.[21] Aquinas' argument for the inadequacy of the physical to account for intellectual life, "no corporeal power can produce the intellective soul"[22] is founded on the principle: "the greater is not brought about by the lesser, for nothing acts outside its species."[23] Utilising a simile, he emphasised the metaphysical and existential dependence on a first cause: "the form of fire emerges when the fire itself is produced."[24] His argument for the necessity of participation in being is drawn from contingency. First, he notes that existence and essence are distinct notions:

---

[16] Andy Mullins, "A Thomistic Metaphysics of Participation Accounts for Embodied Rationality," *International Philosophical Quarterly* 62, no. 1 (2022): 83.

[17] Joseph W. Koterski, "The Doctrine of Participation in Thomistic Metaphysics," in *The Future of Thomism*, eds. D. Hudson and D. Moran (University of Notre Dame, 1992), 185-196.

[18] John F. Wippel, *The Metaphysical Thought of Thomas Aquinas: From Finite Being to Uncreated Being* (The Catholic University of America Press, 2000).

[19] Wayne J. Hankey, "Placing the Human: Establishing Reason by Its Participation in Divine Intellect for Boethius and Aquinas," *Res Philosophica* 95, no. 4 (2018): 583-615.

[20] Christopher M. Cullen and Franklin T. Harkins, eds., *The Discovery of Being & Thomas Aquinas: Philosophical and Theological Perspectives* (Catholic University of America Press, 2019).

[21] Martha C. Nussbaum and Hilary Putnam, "Changing Aristotle's Mind," in *Essays on Aristotle's De Anima*, eds. Martha C. Nussbaum and Amélie Oksenberg Rorty (Clarendon Press, 1995), 56.

[22] Aquinas, *The Summa Contra Gentiles (S.C.G.),* 2. 86.7.

[23] Aquinas, *S.T.,* III 79.2 ad3.

[24] Aquinas, *S.C.G.*, 2. 87.3.

> Every essence or quiddity can be understood without understanding anything about its existence: I can understand what a man is or what a phoenix is and nevertheless not know whether either has existence in reality. Therefore, it is clear that existence is something other than the essence or quiddity.[25]

These contingent living things require a principle of being, unity, and operations; this is the soul. This is evident because at death there is not only dissolution of material unity, but also of the subject, who once present has departed, despite the fact that the component material elements remain. At death this principle of unity is lost. Aristotle regarded the soul as the principle of activities following on the nature of the living substance, but Aquinas argued that the soul must be principle of existence.[26]

Note that Roger Scruton above referred specifically to the need to account for the difference between man and other animals. There is no objection that animal souls emerge from matter. But because rational subjects act with a certain immateriality of thought and with a certain freedom of will, which are operations that transcend matter, Aquinas argued that a rational soul cannot have emerged from matter. Therefore, the rational subject must receive being and its operative powers from beyond, or receive a sharing in such powers. By reflection on the contingency of human intellectual subjects, we deduce the necessity of a principle of participation in being, through which rationality being an essential property of human nature, is shared, or bestowed also in some way from another source.

Existence is through the *actus essendi* of the human soul, the principle of contingent existence, unity, and function, which participates in Being, *in esse subsistens*, that Norris Clarke calls "the Ultimate Source."[27] Fabro regarded participation *in esse subsistens* as the key to Aquinas' metaphysics: "It is from the concept of *esse* as ground-laying first act that Thomas develops his own notion of participation and his entire metaphysic."[28]

Moving beyond Aristotle's reservations,[29] Aquinas recast the Neoplatonic notion of participation into a highly original synthesis

---

[25] Aquinas, *De ente et essentia*, IV.

[26] Aquinas, *S.T.*, I 3.4.

[27] William N. Clarke, S.J. *The One and the Many: A Contemporary Thomistic Metaphysics* (University of Notre Dame Press, 2001), 87.

[28] Cornelio Fabro and B. M. Bonansea, "The Intensive Hermeneutics of Thomistic Philosophy: The Notion of Participation," *Review of Metaphysics* 27, no. 3 (1974): 463.

[29] Aristotle, *Metaphysics,* X.5-6, 1056b4-1057a15.

proposing essence and existence to be really distinct.[30] To it he applied the notion of act and potency to being, giving primacy to the act of being. "All other beings that are not their own being but have being by participation must proceed from that one thing,"[31] and elsewhere, "That which has existence but is not existence, is a being by participation."[32] He presented *esse* as *actus essendi*, in contrast to *existentia* of Augustinianism and of rationalism, and presented form not only as a principle of function or unity, but as a participation *in esse subsistens*.

> In the Platonic tradition, the term "participation" signifies the fundamental relationship of both structure and dependence in the dialectic of the many in relation to the One and of the different in relation to the Identical, whereas in Christian philosophy it signifies the total dependence of the creature on its Creator.[33]

The consistently underlying presence of participation in the thought of Aquinas,[34] came to light in the mid-twentieth century through the work of Fabro, Gieger and others. In line with the sixteenth century commentator Dominic Banez, Fabro drew attention to Aquinas' primacy of subsistent being.

> It is not man who determines Being and imposes it on its varied forms, since it is through Being and in view of, that is, because of, Being that man works in the world.[35]

He regarded the dialectic of participation as "the hermeneutic key of the originality of Thomism."[36]

---

[30] Cornelio Fabro, "Platonism, Neo-Platonism and Thomism, Convergencies and Divergencies," *The New Scholasticism* 44, no. 1 (1970): 69-100.

[31] Aquinas, *De potencia* Dei, 3.5.

[32] Aquinas, *S.T.,* 1.3.4.

[33] Fabro and Bonansea, "The Intensive Hermeneutics of Thomistic Philosophy," 449.

[34] Thomas A. Fay, "Participation: The Transformation of Platonic and Neoplatonic Thought in the Metaphysics of Thomas Aquinas," *Divus Thomas* 76 (1973): 50-64.

[35] Cornelio Fabro, "The Problem of Being and the Destiny of Man," *International Philosophical Quarterly* 1 (1961): 407.

[36] Cornelio Fabro, *Esegesi Tomistica* (Libreria Editrice della Pontificia Università Lateranense, 1969), xxxiii.

b. Features of such a view of rationality[37]

i. Rationality and intellect are by participation

Within TMP, rationality is mediated by, but not reducible to, physical processes. Phantasms are made actually intelligible by the active intellect. This active intellect consists of immaterial and intelligible species and is a participation in "Divine light."[38] This point is crucial: the intellect is a participated power, a power not properly its own, but one which "belongs to another fully."[39]

Aquinas argued that goodness and truth participate in Being. Human subjects possess being, and rationality, both *by participation*. "It is necessary that that which is greatest in being and truth be the cause of being and truth in all other beings."[40] It can only be through participation in being, that non-material perfections of the Ultimate Source – the capacity to know (to grasp reality) and to will (to love on the basis of intellectual choice) – are present in the particular as essential properties.[41] The essential operations of knowing and choosing are present also by the same act of being. Aquinas notes in *De Veritate* how being also underpins knowing:

> [...] a thing is apt to be conformed (adaequari) to the intellect in the degree to which a thing has entity (entitas). Consequently, the notion of truth follows upon that of being.[42]

In the hierarchy of being, the intellectual enjoy "intensified substantiality" through richer participation in being. As a consequence they are capable of "immanent and spontaneous activity."[43] Only on a ground of participation in Being from an Ultimate Source, may unity and rationality of an intellectual subject be safeguarded.

From this it follows that rationality is not just a process or one human activity among others, but it is an *essential* mark of human nature, inseparable, not because an individual human being always acts rationally, but because human nature is rational. It is an essential truth about human beings.

---

[37] This view is developed and argued in Andy Mullins, "Rationality and Human Fulfilment Clarified by a Thomistic Metaphysics of Participation," *Scientia et Fides* 10, no. 1 (2022a): 177-195, and Mullins, "A Thomistic Metaphysics of Participation Accounts for Embodied Rationality."

[38] Aquinas, *S.T.*, I 89.1, and I 84.6.

[39] Aquinas, *Expositio* 1.2, cited in Fabro and Bonansea, "The Intensive Hermeneutics of Thomistic Philosophy," 454.

[40] Aquinas, *De substantiis separatis,* 3. 58.

[41] William N. Clarke, S. J. *Explorations in Metaphysics: Being-God-Person* (University of Notre Dame Press, 1994), 65ff and 89ff.

[42] Aquinas, *Questiones disputatae de Veritate*, 1.1 ad 5.

[43] Juan E. Carreño, "From Self Movement to Esse: The Notion of Life and Living Being in Thomas Aquinas," *Angelicum* 92, no. 3 (2015): 347-376.

ii. Mental life all mediated by, but not reducible to, the biophysiological

TMP finds common ground with physicalist non-reductive and emergent accounts. TMP supports the view that in the human subject there is an immaterial life of the mind underpinned in all operations by embodied neural bases. In TMP, biophysiology is the material cause of the operations of rationality: "intellectual knowledge is caused by the senses [...] it is in a way the material cause."[44]

> In the present state of life in which the soul is united to a passible body, it is impossible for our intellect to understand anything actually, except by turning to the phantasms.[45]

iii. An enriched notion of rationality

So, by means of its participation *in esse subsistens,* TMP is able to distinguish the embodied rational person from non-human-animal natures. Rationality is seen as an essential quality of human life. Even if high level cognitive processing should emerge from matter, those behaviours would be one more competing feature or behaviour among others. In this way, the processes, structures and systems of reasoning and mental life are not seen as defining features of rationality, but signs that rationality may or may not be present.

Here too the Thomistic notion of a participative active intellect is of great importance. According to the TMP account, in contrast with essentialist metaphysical systems like that of Aristotle himself, essence and its primary constituent, form, do not play the role of primary repositories of perfection, but take on rather the secondary, derivative role of principles of limitation.[46] Human nature is actualized through participation *in actus essendi*, but limited by the human essence, and so too, the power of intellection, as an essential property of participating being, is bestowed from without.

iv. Teleological implications for the human subject[47]

Within TMP and a Thomistic understanding of the transcendentals, truth is grounded in being, *in esse subsistens*. Rationality is seen as essential property of the soul. The operations of rationality make possible human fulfilment in truth and love. Non-reductive physicalist accounts cannot discuss such

---

[44] Aquinas, *S.T.*, 1.84.6.

[45] Ibid., 1.84.7.

[46] Clarke, *Explorations in Metaphysics*, III.

[47] See extended discussion of these notions in Mullins, "Rationality and Human Fulfilment Clarified by a Thomistic Metaphysics of Participation."

fulfilment as essential, because they offer no basis to prioritise the operations of rationality over other activities.

Thereby TMP provides a coherent account of the capacity for human beings to grasp truth and universal concepts, to make love choices based on these truths. Walker offers philosophical arguments, developed from a communitarian Thomistic perspective, for the view that integral to rationality are loving relationships: that human persons by their very nature are fulfilled in personal loving relationships.[48]

c. Not all hylomorphic anthropology recognises participation in being as the decisive factor. Challenges faced by current Anglo-American philosophy of mind.

By arguing that the immaterial life of the mind correlates with, but is not reducible to, embodied neural bases TMP is free of any taint of dualism which is anathema to contemporary neuroscience, summed up in the words of Nobel laureate Eric Kandel: "Philosophically disposed against dualism, we are obliged to find a solution to the problem in terms of nerve cells and neural circuits."[49] These immaterial operations of rationality are transcendent operations carried out by ensouled matter through a participated power, at all times mediated by the biophysical. Thus unity of the subject in the embodied rational person is defended, and substance dualism, indeed all dualism, is avoided.

On the other hand, when hylomorphic arguments for immaterial properties proper to a human being have been framed on the basis of formal causality, there has been a tendency by the advocates themselves, to frame the operations of the intellect in dualistic terminology. This dualism arises because of a different understanding of "immateriality of the soul." TMP views the active intellect as a participated power and not a formal or constitutive principle, to use the terminology of Fabro.[50] Instead of accounting for immateriality through this participated power, an abstracting "non-physical" faculty,[51] proper to the soul, is proposed. This becomes the "formal or constitutive

[48] Adrian J. Walker, "Personal Singularity and the *Communio Personarum*: A Creative Development of Thomas Aquinas' Doctrine of *Esse Commune*," *Communio* 31, (2004): 457-479.

[49] Eric R. Kandel, James Harris Schwartz, and Thomas Jessell, *Principles of Neural Science* (McGraw-Hill, 2000), 1317.

[50] Cornelio Fabro, *La Nozione Metafisica di Partecipazione Secondo S. Tommaso d'Aquino* (Società editrice internazionale, 1950), 272-273. Reference in Jason A. Mitchell, "Being and Participation: The Method and Structure of Metaphysical Reflection according to Cornelio Fabro" (PhD diss., Pontifical Athenaeum Regina Apostolorum, 2012).

[51] James D. Madden, *Mind, Matter, and Nature: A Thomistic Proposal for the Philosophy of Mind* (The Catholic University of America Press, 2013), 671.

principle" of the intellect,[52] whose operations of "could not be powers actualized in matter."[53] In this way, rather than looking to the infinite Source itself to explain intellectual life, we look to constituent faculties of the human being.

Significantly all the important contemporary Anglo-American texts in *hylomorphic* philosophy of mind that have been published in recent years, including major works by Feser,[54] Madden,[55] and Jaworski,[56] adopt an exclusive focus on the formal causality of the human soul and omit discussion of participation, although Madden makes one reference without elaboration to "ontological causality." Jaworski stands in greater contrast by his primary focus on the structure bestowed by form. Other recent relevant works feature what could be described as an essentialist approach more akin to Aristotle than to Aquinas.[57]

Formal causality without participation is susceptible to the critique of dualism because it downplays full embodiment and suggests immaterial agency. Both Feser and Madden suggest that immateriality of the soul demands a form of property dualism. But as Lycan has argued, how is the property dualist solution not a substance dualist solution.[58] Property dualism risks being an assertion without a coherent explanation. D. M. Armstrong has also picked this up when he noted the "considerable difficulty and confusion which surrounds the philosophical theory of properties," and how this poses a challenge for the "one substance view" which includes hylomorphism.[59]

Without a paradigm of participation, dualism in some form is inevitable if one wishes to preserve the enriched understanding of rationality discussed above. But resort to dualism shuts down the conversation with advocates of non-reductive physicalist and emergent accounts. TMP is able to engage with emergent accounts as both are wholly embodied. In both, the biological mediates all intellectual operations of the embodied subject; but also, as we have seen, TMP offers the advantage of a teleological metaphysics that supports the unique nature of human life.

---

[52] Fabro, *La Nozione Metafisica di Partecipazione*, 272-273.

[53] Madden, *Mind, Matter, and Nature,* 254.

[54] Edward Feser, *The Philosophy of Mind: A Short Introduction* (Oneworld Publications, 2005).

[55] Madden, *Mind, Matter, and Nature*.

[56] William Jaworski, *Structure and the Metaphysics of Mind: How Hylomorphism Solves the Mind-Body Problem* (Oxford University Press, 2016).

[57] Kathrin Koslicki, *Form, Matter, Substance* (Oxford University Press, 2018).

[58] William G. Lycan, "Is Property Dualism Better off than Substance Dualism?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 164, no. 2 (2013): 533-542.

[59] D. M. Armstrong, "Mind-Body Problem: Philosophical Theories," entry in *The Oxford Companion to the Mind,* ed. Richard L. Gregory (Oxford University Press, 2006).

IV. Distilling the neurobiology[60]

Because real-time brain imaging has only been possible since the 1980s, the neurobiological descriptions that follow are hypothetical but thoroughly founded on current research into neurobiological mechanisms, pathways, systems and regions. The originality of this paper is to synthesise this knowledge into an account of the neural bases of a virtuous action. For the sake of clarity I will analyse the sequence of events into eight segments. The principal neurobiological players in the description below are the *prefrontal cortex* (PFC), acknowledged as mediating consciousness. Within the PFC, the *orbitofrontal cortex* (OFC) and the *ventromedial PFC* (VMPFC) are associated with consciousness of emotion, and the *dorsolateral prefrontal cortex* (DLPFC) with conscious regulation of emotion. The *limbic system*, and especially the *amygdala*, play a major role in mediating emotional response and memory. The *basil ganglia* (BG) including the *striatum*, the *nucleus accumbens* (NA), and the *substantia nigra* (SN) is now recognised as playing a major role in linking centres for emotion, reward, and both conscious and unconscious habit formation. *Dopamine* (DA) is a neurotransmitter associated with attention, and reward. Note too that in the course of moral evaluations and judgements there is a complex integration of numerous neural subdivisions which show significantly consistent activation across numerous studies.[61]

a. "The chief nurse came running up and handed me a sheet of paper"[62]

Nagai has already demonstrated that he has finely attuned responsibilities to those he leads, and habitual courtesy. He takes the leaflet from the nurse, conscious of his duties towards her. It is reasonable to surmise that this sense of responsibility as well as his training in scientific objectivity assisted him in applying sufficient deliberation before committing himself to judgement. Pre-established habits of professionalism dispose Nagai to take his duties to the nurse and to the survivors seriously. He will be aware of the importance in this crisis of giving appropriate example and reaching a final deliberation in the best interests of the entire community he now finds himself leading. The principle evident in his actions is that truth, no matter how unpalatable,

---

[60] It is beyond the scope of this paper to justify the implication of the neurobiological systems, pathways, and processes discussed. Data accords with standard reference neurobiological texts such as Larry Squire et al., eds., *Fundamental Neuroscience* (Elsevier, 2008); Kandel et al., *Principles of Neural Science*; Mark F. Bear et al., *Neuroscience. Exploring the Brain* (Lippincott Williams and Wilkins, 2007). For complete text and research papers references see Andrew P. J. Mullins, "An Investigation into the Neural Substrates of Virtue to Determine the Key Place of Virtues in Human Moral Development" (PhD diss., University of Notre Dame Australia, 2012).

[61] Mullins, "Can Neuroscientific Studies Be of Personal Value?" 448-451.

[62] Nagai, *Bells,* 51.

must be faced and for a leader to do so is necessary. Prior learning of duty and courtesy, both disposed by justice, lead Nagai to respond to the nurse's implicit request to read the leaflet.

Nagai is extremely drained and exhausted as it is the day after the bomb has exploded. He has suffered much blood loss and has lost consciousness at least on one occasion. The need to care for the wounded has given him little rest. He is also fatigued from the responsibility of leading the growing group of survivors. Exhaustion will be manifested in lower levels of attention, a heightened capacity for emotionally initiated representations (an overactive imagination), and difficulty in suppressing goal irrelevant sensory inputs. When Nagai is approached by the nurse and becomes aware of the American leaflet, his *fronto-parietal attentional system* is triggered. *Acetylcholine* (Ach) from the *nucleus basalis* is released into the *thalamus* and into areas of the *parietal, frontal and cingulated cortices*, heightening attentiveness. Attentional loops between the BG and the cortex are established. DA, a facilitator of attention, is also activated for emotional and reward responses.[63]

b. "In the depth of my being I felt a tremendous shock. The atom bomb has been perfected! Japan is defeated!"[64]

Nagai's description of how his passionate reaction subsides on successive readings of the leaflet gives us a remarkable insight into how deliberation can enable mastery of passion. What is initially less obvious is the interior battle that Nagai has fights in order to respond rationally to the news. There is an initial movement of complacency towards the truth. He comprehends the American claim and reacts initially dispassionately.

As Nagai glances at the leaflet initial visual input is channelled to the *basolateral complex of the amygdala* (BLA) from the *sensory nuclei of the thalamus* which has filtered out other inputs allowing his full attention on the leaflet. A degree of fear conditioning will be evident in the BLA given the suffering that Nagai has endured. There is an initial numbness, a lack of cortical response associated with a welling sadness (perhaps reflecting a left VMPFC that is underactive in managing right side negative affect), but Nagai falls back, perhaps heightened through his training as a scientist, on an established habit of seeking knowledge. He reads in the text a confirmation of his suspicions of a nuclear blast. His scientific knowledge gives him virtually

[63] Note that it can be only the person as the subject of the action when we are considering a human act. However, when we are not referring to a human act, but to unconscious responses by parts of the body (e.g. ACh heightens attentiveness) or to the action of part of the body (e.g., my eye blinked, my amygdala responded to emotional input), I adopt the convention of specifying a non-personal subject.

[64] Nagai, *Bells,* 51.

an intuitive grasp of the potential of such a bomb. Cortical representations of sadness for his country and people, and disgust at the Americans, flood his cortical memory and a negative emotional state overwhelms reason. Even in a state of physical and psychological exhaustion, Nagai's learned self-mastery is sufficient for him to utilise cognitive pathways to gain some management of passionate predisposition, reacting with deliberation, and consideration of consequences.

c. "[...] Conflicting emotions churned in my mind and heart as I surveyed the appalling atomic wasteland around me. [...] A bamboo spear lay on the ground. I kicked it fiercely and it made a dull, hollow sound. Grasping it in my hand, I raised it to the sky, as tears rolled down my cheeks. The bamboo spear against the atomic bomb! What a tragic comedy this war was! This was no longer a war. Would we Japanese be forced to stand on our shores and be annihilated without a word of protest?"[65]

Conflicting emotions well up within him, of frustration, sadness, anger, patriotism, and shame at defeat. He deliberates briefly about the prospects of victory for Japan in the face of nuclear weapons and concludes there is no hope of victory. He remains resilient to blind passionate assertion. Prudence disposes his readiness to apply himself diligently to assess the truth of the American claims. Justice ensures that does not dismiss the American claims out of hand.

Initially we are witnessing a "low road"[66] emotional response unmediated by the cortex. Although there is consciousness in the PFC of the response, there is little cortical processing. A further surge of emotional memory inputs flood into the BLA via reciprocal connections to the *hippocampal and striatal memory systems*. There are neural activations of fear, sadness, anger and disgust (in the *anterior insula*). Outputs from the BLA trigger a rage response in the *dorsomedial nucleus of the hypothalamus*. This hypothalamic response directs rage related motor patterns. The various limbic aversion centres also activate. Limbic afferents to the PFC via the OFC trigger consciousness of the emotional response and provide an initial justification based hippocampal call up of DA mediated, short term, memories of the suffering witnessed in association with the bomb. Nagai seeks gratification in the pointless action of kicking the spear. As Nagai apprehends the bamboo spear, DA floods the NAc in anticipation. *Globus pallidus* (GP) and the SN select patterns of cortical activity and motor programs for action, drawing on movement sequencers

[65] Ibid.

[66] Terminology popularised by Joseph LeDoux. Cf. Joseph E. LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (Phoenix Press, 1999).

in the *dorsal striatum*. The *hippocampus* is enlisted for goal direction. The BG instruct the motor areas of the PFC via the *BG-thalamo-cortical loop*. PFC and motor cortices deliver executive command. The bamboo is kicked. The BLA is now drawing input from widespread cortical areas, and from, and from cortical, hippocampal and striatal memory systems. Nagai shows a greater awareness of his current state, looking around and reflecting on his situation. Attempts to articulate the situation serve to mitigate his passionate reaction. Nagai indulges in a flight of imagination calling up a succession of cortical representations of future scenarios, and thereby diverts his attention away from sensory input and into deliberation. Effectively he is buying time for cortical processing of the overwhelming emotion. The initial surge of passion, corresponding to a neuromodulating flood of ACh and DA, dissipates somewhat and he is able to adopt a more cognitive response. His training in self-discipline can now take effect. Also he is conscious that he has an audience. His desires to give good example, despite an absence of evident reward, are mediated by overlearned *stimulus response* (S-R) responses in the ventral striatum. His prior training has established this mechanism that is now available to him.

d. "I read the leaflet once and was stunned."[67]

By learned pathways of reflection Nagai sets about the task of reaching the truth of the situation. Quickly Nagai refocuses on reaching the truth or falsehood of the American claim. He seeks to weigh the message of the leaflet with his own assessment. Prior to the commencement of his deliberative reading there is a moment of "election" towards the means to reach the chosen good: by reading and calm consideration he decides he will reach the truth. Despite the almost overwhelming emotion of the moment, he draws on learned neural pathways of determination with the goal of uncovering the truth no matter how unpalatable. He makes a judgement that the truth can be reached by suppressing immoderate emotion and then applies himself rereading the leaflet weighing its assertions against his own knowledge and estimations.

Nagai's self-mastery, established in a habitual way by prior learning, is characterised by obedience of the emotional realm to rational command. Goals and rewards are not needed as incentives. S-R processes give way to intentional *action-outcome* (A-O) learning. Input to the PFC via the *BG-thalamo-cortical loop* reestablishes higher cortical, as opposed to emotion driven, management. Emotional thalamic and limbic centres are now redisposed to cortical inhibition of impulsive response. Prior learning of fortitude and

[67] Nagai, *Bells,* 52.

temperance have established pathways moderating emotional responses, so they are obedient to rational direction; and these neural pathways are activated. The habit of fortitude will be present in preferential pathways in the *BG-thalamo-cortical loop* and strengthened top-down connections between the OFC and the amygdala. The prior habituation of temperance will be present in mediation primarily by the DLPFC and *anterior cingulated cortex* (ACC). In the past it is likely that this habituation was associated with DA rewards that consolidated the pathways of self-control. In this current scenario, with the habit established, it is unnecessary for DA perfusion to take place. Various brain regions now coordinate to regulate emotional reaction: OFC, DLPFC, VMPFC, further areas of the amygdala and of the BG. Emotional regulation leads to cortical direction of reward expectations. Emotional representations are consciously suppressed. Cortical management is consolidated via action plans involving rereading. Even though it is unlikely that this is a conscious strategy, nevertheless it is likely to be a learned strategy to divert oneself into a cognitive task in order to take the heat out of emotion.

e. "I read it a second time and felt they were making fools of us."[68]

Nagai's established dispositions to moderate and divert excessive emotional representations and expressions are the result of prior learning. The actions of these dispositions permit cortical deliberation to occupy his attention. Nagai's response in this situation, and his character in general, is built on previously established behaviours and convictions upon which he can draw.

Nagai's habit of application at the task until the desired outcome is achieved as a consequence of prior training. His election of the goal to read and reread involves deliberative evaluation utilising numerous cortical areas, drawing particularly on episodic memory, self-knowledge, scientific knowledge and skills of critical assessment.

Intrinsic motivations have come to the fore: of commitment to the truth, and that one's duty to others must be fulfilled. These appear to be the result of cortical neuronal pathways established and consolidated by prior experience and DA reinforcement, originating in the *ventral tegmental area* (VTA) and SN and mediated by the *ventral striatum*, by childhood and military training and in happier times. The prior habituation of prudence will be present in this rich and reciprocal connectivity, primarily to and from the DLPFC, with other cortical areas serving memory and somatic and sensory input, with the OFC, DMPFC, BG, and *amygdala* serving emotion regulation, with the *ventral striatum* assisting in goal setting and motivation. Similarly the prior habituation of justice consists of consolidated pathways in the

[68] Nagai, *Bells,* 52.

anterior and medial PFC, the VMPFC, OFC (especially medial OFC), ACC (especially *rostal ACC*), *insula*, *limbic and paralimbic areas*, and the BG. These rich connections serve to give preferential traffic to deliberations about understanding of others, consideration of the impact of one's actions on others, empathy, and considerations of fairness, etc. It is possible to detect in this change of behaviour the classic pattern noted by Graybiel; a passing from reward mediation in the *ventral striatum* (Nagai's kicking fruitlessly against the goad), associated with the emotional gratification, to a dorsal automatisation, carrying out duty as he has trained himself to do (Nagai's determination to grapple with the truth and face it).[69] Such an automatisation is consistent with our knowledge of the character of this wonderful man.

f. "I read it a third time and was enraged at their impudence."[70]

In neural terms Nagai is aware of cortical representations of attractive or aversive sense objects, and that reward systems provide neuromodulatory incentive for preferential attention to, and pursuit of these goals via appropriate action plans. Cognition and the capacity to reach the truth can be overwhelmed by the presentation of attractive or aversive cortical representations. This is the battle that Nagai fights in the first, second and third reading. However, by holding to his action plan of rereading he is able to regulate the emotion sufficiently to allow deliberation and a final judgement as to the truth of the American message. As we are discussing aversional content, direct involvement of reward systems is minimal. However, during prior learning the habits of prudence, justice, fortitude and temperance were established; during this time the reward systems were greatly active, leading to DA mediated reinforcement of regulatory pathways triggered in the OFC and *amygdala* by sense representations. These pathways are available now for Nagai's use in coping with this particularly difficult situation.

g. "But when I read it a fourth time I changed my mind and began to think it was reasonable."[71]

Nagai has reached the point of making a judgement about the trustworthiness of the American leaflet. Nagai's neural activity is likely to include heightened activation in numerous areas. The DLPFC, ACC, the ventral striatum, and amygdala will reveal principal activation. Above baseline activation will also be evident in medial PFC, VMPFC, OFC, the *posterior cingulate/retrosplenial cortex*, *superior*

[69] Ann M. Graybiel, "Habits, Rituals, and the Evaluative Brain," *Annual Review of Neuroscience* 31 (2008): 359-387.

[70] Nagai, *Bells,* 52.

[71] Ibid.

temporal cortex, STS, the *temporo-parietal junction*, *medial hypothalamus*, and insula. Left PFC and right OFC will be active in the suppression of sadness. Considerations of the social norms of patriotism and of cultural expectations of a leader will involve further integration of areas such as the VMPFC, lateral OFC, and the *anterior temporal lobes*, assisted by storage of social perceptual representations in the *temporal lobes*. In addition, reflecting Nagai's frustration at the shame of defeat and empathy with the pain of others, the *rostral ACC*, and the *anterior insula* will be active. The *posterior cingulate*, and *inferior parietal lobe* will show activity during the Nagai's brief catastrophising.

h. "But when I read it a fourth time I changed my mind and began to think it was reasonable. After reading it a fifth time I knew that this was not a propaganda stunt but the sober truth."[72]

He gives his assent to the assertion of the leaflet. His word "sober" indicates he has achieved mastery of impulsive response, in contrast with his fierce kick to the bamboo spear: destructive, impulsive, futile, and pointless. Acquiescence to the truth implies also a degree of self-mastery. Note the capacity of the brain to operate multiple processes, concurrently and in concert, in support of goals of the person's own choosing.

In Nagai's reading of the flyer, in this single human act seeking the truth of the issue, we have witnessed a highly complex interplay of systems (memory, emotional management, deliberation, goal election, consideration of consequences of action, moral judgement, attention, reward to some extent, and motor execution), brain areas, mechanisms, and pathways. Throughout he has drawn upon learned (prior) responses to emotion, and in this episode, he is able to further reinforce the neural pathways underpinning a virtuous response. The end result is the harmonisation of the emotional life with the rational life. The disorder of actions carried out without cognitive approval and direction, is replaced by a grasping of truth and a quieting of wayward passion. Virtue thus may be understood as supported at the neurobiological level by a harmonised complex of systems.

IV. Conclusion

A close examination of Nagai's response offers an augmented understanding of the neural bases of self-control within an anthropology supported by TMP. Within this view, virtue, as understood by Aristotle and Aquinas, is embodied, mediated by the neural structures, pathways and mechanisms. In this final section, I draw together some insights into human fulfillment and virtue suggested by this neurobiological study.

[72] Ibid.

Appreciation that character has a biological basis offers practical insights into the role of emotional management in development of character: we see the effects of this in Nagai's wrestle with motivation, and also in his final acquiescence. This exploration of Nagai's experience exposes the mechanisms whereby intentional management of emotion not only brings better immediate outcomes for the person, but also that, though use-induced plasticity, behaviours predispose for further emotional management. Nagai's experience serves to demonstrate that personal development lies in choosing the right goals for our actions: that repeated actions build a facility for future action, and that habits of emotional management may be established to support future action.

Emotional regulation is supported by limbic-cortical connectivity permitting bottom-up modification of cortical "decision making," and top-down direction and regulation. Neuroscience and philosophy converge in describing the complementary roles of emotion and reason in a balanced happy life. Given that neural pathways are plastic, strengthening with use, and falling into atrophy in disuse, the very presence of substantial reciprocal neural pathways is firm evidence of both cortical direction and subcortical modification of regulation and decision making. Similarly human motivation and goal election are supported by the reward structures of the brain that are in reciprocal communication with cortical structures.

We see that Nagai's cognitive activity founded in the PFC requires the complement of subcortical structures of the limbic system and the BG for effective emotion regulation and goal election. The involvement of the BG appears crucial, not only by its implication in emotion and reward pathways but also because the BG are the principal seat of automaticity of actions, which has ultimately assisted Nagai in bringing the tension to a conclusion. Such automaticity is not necessarily opposed to conscious, voluntary goal election.

Nagai's account draws attention to the very process of formation of the intentional self-control. Consistent with the weight of current neuroscientific opinion concerning the capacity for economies of interconnection in the brain, we are led to consider the state of a virtue as a complex of systems. Interdependent with processes of emotion regulation we find systems of responses to pleasure, pathways for fear responses, reward evaluation, goal setting, motivation and executive control, all supported by "upstream" systems of plasticity, learning and memory, and capacities for attention, critical learning, imitation, and empathy.

It is apparent, too, that self-management of emotion brings a liberating result evident in the acquiescence of Nagai. Our free and conscious efforts form, or reform, our very biophysical constitution into neural structures,

better capable of supporting effective self-management at the rational level. Hence the process is apparent whereby virtue is acquired, and freedom augmented.

In conclusion, it may be seen that virtue has a material foundation in the neural structures, systems and processes of the brain and that these material aspects can be identified. These neural structures manifest, in their maturity of expression and integration across the entire brain, the role that virtue plays in human fulfilment itself, and that there is a biological aptitude and *predisposition* in human beings for the development of virtue.

That the neurobiology of virtue may be described, and that it is associated with a state of neurobiological perfection, must carry far reaching implications for the study of ethics. Aristotle proclaimed, "Happiness is the reward of virtue," with happiness understood as human fulfilment, flourishing, *eudaimonia*.[73] Happiness, understood as flourishing, seems, to the most considerable extent, a consequence of the neurobiological presence of virtue.

References

Anscombe, Elizabeth. "Modern Moral Philosophy." In *Virtue Ethics*, edited by Roger Crisp and Michael Slote, 26-44. Oxford University Press, 1997.

Aquinas, Thomas. *De ente et essentia*. Edited by Joseph Kenny, O. P. https://isidore.co/aquinas/DeEnte&Essentia.htm.

Aquinas, Thomas. *De potencia Dei*. Edited by Joseph Kenny, O. P. https://isidore.co/aquinas/english/QDdePotentia.htm.

Aquinas, Thomas. *De substantiis separatis*. Edited by Joseph Kenny, O. P. https://isidore.co/aquinas/SubstSepar.htm.

Aquinas, Thomas. *Questiones disputatae de veritate*. Edited by Joseph Kenny, O. P. https://isidore.co/aquinas/QDdeVer.htm.

Aquinas, Thomas. *The Summa Contra Gentiles*. Edited by Joseph Kenny, O. P. Hanover House, 1955-57. https://isidore.co/aquinas/ContraGentiles.htm.

Aquinas, Thomas. *The Summa Theologica*. Translated by Fathers of the English Dominican Province. https://www.newadvent.org/summa/.

Aristotle. *The Complete Works of Aristotle: Volumes 1 and 2*. Edited by Jonathan Barnes. Princeton University Press, 1984.

Armstrong, D. M. "Mind-Body Problem: Philosophical Theories." In *The Oxford Companion to the Mind,* edited by Richard L. Gregory. Oxford University Press, 2006.

[73] Aristotle, *Nicomachean Ethics*, 1099b16.

Barnes, Gordon P. "The Paradoxes of Hylomorphism." *The Review of Metaphysics* 56, no. 3 (2003): 501-523.

Bear, Mark F., Barry W. Connors, and Michael A. Paradiso. *Neuroscience. Exploring the Brain*. Lippincott Williams and Wilkins, 2007.

Carreño, Juan Eduardo. "From Self Movement to *Esse*: The Notion of Life and Living Being in Thomas Aquinas." *Angelicum* 92, no. 3 (2015): 347-376.

Clarke, William N., S. J. *Explorations in Metaphysics: Being-God-Person*. University of Notre Dame Press, 1994.

Clarke, William N., S. J. *The One and the Many: A Contemporary Thomistic Metaphysics*. University of Notre Dame Press, 2001.

Cullen, Christopher M., and Franklin T. Harkins, eds. *The Discovery of Being & Thomas Aquinas: Philosophical and Theological Perspectives*. Catholic University of America Press, 2019.

Fabro, Cornelio, and B. M. Bonansea. "The Intensive Hermeneutics of Thomistic Philosophy: The Notion of Participation." *Review of Metaphysics* 27, no. 3 (1974): 449-491.

Fabro, Cornelio. "Platonism, Neo-Platonism and Thomism: Convergencies and Divergencies." *The New Scholasticism* 44, no. 1 (1970): 69-100.

Fabro, Cornelio. "The Problem of Being and the Destiny of Man." *International Philosophical Quarterly* 1, no. 3 (1961): 407-436.

Fabro, Cornelio. *Esegesi Tomistica*. Libreria Editrice della Pontificia Università Lateranense, 1969.

Fabro, Cornelio. *La Nozione Metafisica di Partecipazione Secondo S. Tommaso d'Aquino*. Società editrice internazionale, 1950.

Fay, Thomas A. "Participation: The Transformation of Platonic and Neoplatonic Thought in the Metaphysics of Thomas Aquinas." *Divus Thomas* 76 (1973): 50-64.

Feser, Edward. *The Philosophy of Mind: A Short Introduction*. Oneworld Publications Limited, 2005.

Glynn, Paul. *A Song for Nagasaki*. Marist Fathers Books, 1988.

Graybiel, Ann M. "Habits, Rituals, and the Evaluative Brain." *Annual Review of Neuroscience* 31 (2008): 359-387.

Haldane, John. "A Return to Form in the Philosophy of Mind." *Ratio* 11, no. 3 (1998): 253-277.

Hankey, Wayne J. "Placing the Human: Establishing Reason by Its Participation in Divine Intellect for Boethius and Aquinas." *Res Philosophica* 95, no. 4 (2018): 583-615.

Jaworski, William. *Structure and the Metaphysics of Mind: How Hylomorphism Solves the Mind Body Problem*. Oxford University Press, 2016.

Kandel, Eric R., James Harris Schwartz, and Thomas Jessell. *Principles of Neural Science*. McGraw-Hill Companies, 2000.

Koslicki, Kathrin. *Form, Matter, Substance*. Oxford University Press, 2018.

Koterski, Joseph W. "The Doctrine of Participation in Thomistic Metaphysics." In *The Future of Thomism*, edited by D. Hudson and D. Moran, 185-196. University of Notre Dame, 1992.

LeDoux, Joseph E. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Phoenix Press, 1999.

Lycan, William G. "Is Property Dualism Better off than Substance Dualism?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 164, no. 2 (2013): 533-542.

Madden, James D. *Mind, Matter, and Nature: A Thomistic Proposal for the Philosophy of Mind.* The Catholic University of America Press, 2013.

Mitchell, Jason A. "Being and Participation: The Method and Structure of Metaphysical Reflection according to Cornelio Fabro." PhD diss., Pontifical Athenaeum Regina Apostolorum, 2012.

Mullins, Andrew P. J. "An Investigation into the Neural Substrates of Virtue to Determine the Key Place of Virtues in Human Moral Development." PhD diss., University of Notre Dame Australia, 2012.

Mullins, Andy. "Can Neuroscientific Studies Be of Personal Value?" *International Philosophy Quarterly* 57, no. 4 (2017): 429-451.

Mullins, Andy. "A Thomistic Metaphysics of Participation Accounts for Embodied Rationality." *International Philosophical Quarterly* 62, no. 1 (2022): 83-98.

Mullins, Andy. "Philosophical Prerequisites for a Discussion of the Neurobiology of Virtue." *Ethical Perspectives* 23, no. 4 (2016): 689-708.

Mullins, Andy. "Rationality and Human Fulfilment Clarified by a Thomistic Metaphysics of Participation." *Scientia et Fides* 10, no. 1 (2022a): 177-195.

Nagai, Takashi. *The Bells of Nagasaki*. Translated by William Johnston. Kodansha America Inc, 1994.

Nussbaum, Martha C., and Hilary Putnam. "Changing Aristotle's Mind." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amélie Oksenberg Rorty, 27-56. Clarendon Press, 1995.

Squire, Larry, Darwin Berg, Floyd E. Bloom, Sascha du Lac, Anirvan Ghosh, and Nicholas C. Spitzer, eds. *Fundamental Neuroscience*. Elsevier, 2008.

Walker, Adrian J. "Personal Singularity and the *Communio Personarum*: A Creative Development of Thomas Aquinas' Doctrine of *Esse Commune*." *Communio* 31 (2004): 457-479.

Wippel, John F. *The Metaphysical Thought of Thomas Aquinas: From Finite Being to Uncreated Being*. The Catholic University of America Press, 2000.

R. T. Oyelakin · *Conatus* 10, no. 1 (2025): 193-206
doi: https://doi.org/10.12681/cjp.30907

Multiple Realizability in the Nature of the Mind and Its Implications for SETI

Richard Taye Oyelakin
Obafemi Awolowo University, Nigeria
E-mail address: richyman2009@yahoo.com
ORCID iD: https://orcid.org/0000-0002-0804-6420

Abstract

Responding to Putnam's computational hypothesis of the mind and the adoption of the Turing machine, it is argued by Churchland and Searle (biological naturalists) that the implementing organic structure is necessary in understanding the nature of mental states. This paper notes that if the term "necessity" is understood in terms of "withoutness," then it is argued, from the idea of multiple realizability, that no particular implementing structure is necessary to the nature of the abstract mental state. Furthermore, drawing implications from the analysis, the paper shows how limited and unjustified human understanding and generalizations about the issue of mental states can be when viewed only from an anthropocentric perspective, and the dire implications this brings on the search for extraterrestrial intelligence (SETI). The paper concludes that there is a need to review our methodology and reorient our technology to make a more promising search. The paper employs philosophical argumentation and analysis as tools of assessment of the metaphysical hypothesis.

Keywords: *phenomenal experience; implementing system; mental state; intelligence; extraterrestrial intelligence*

I. Introduction and discussion of the problem

It is argued in the literature by Ned Block and Patricia Churchland respectively that cooperation and co-evolution between computational hypothesis and neuro-physiological account are expected to provide the desired sufficient account of the nature of mental states.[1] This suggests that both Put-

[1] Patricia Churchland, "The Co-evolutionary Research Ideology," in *Readings in Philosophy and Cognitive Science*, ed. Alvin Goldman, 745-768 (MIT Press, 1993), 745. Ned Block, "The Computer Model of the Mind," in *Readings in Philosophy and Cognitive Science*, ed. Alvin Goldman, 819-832 (MIT Press, 1993), 824.

nam's computational hypothesis and neuro-physiological account are necessary and sufficient for the understanding of the nature of mental states. This view is based on Searle's and Churchland's conviction that mental states are causally produced by the activation of C-fibres in the brain. This is a response to Putnam's abstract computational nature of the mind, which resulted from the adoption of the Turing machine.[2] It means that as the stimulus input strikes the nerve endings, it institutes a neural process which triggers the activation and firings of relevant and appropriate neurons, and this involves some form of energy transfer.[3] Based on the assumption that mental states are "caused by and realized in the neurophysiology,"[4] it appears that it is part of the nature of C-fibre firing in the brain to produce and transfer energy to which Searle was so emphatic.[5] For him, the C-fiber firing occurs at the lower level of the neural process. This lower-level C-fibre firing causes the mental states at the higher level. The neural firings process produces a corresponding mental state.[6] However, Searle has not yet provided a strong explanation on how the brain does it or the sorts of chemical process which combine to produce mental states.[7]

The point that "any causal power the machine might have to cause consciousness, and intentionality would have to be a consequence of the physical nature of the machine,"[8] may seem to imply that the nature of the implementing physical structure which implements the abstract state description is so necessary for an adequate account of the mental states. In fact, this is the view held by Churchland and Searle respectively. However, the term *necessity* of the implementing structure appears ambiguous. What is of interest is the point that as much as multiple realizability is concerned, no particular implementing structure is necessary in re-

---

[2] Alan M. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433-460. The Turing machine has been described by Putnam as "a device with finite number of internal configurations, each of which involves the machine's being in one of a finite number of states, and the machine's scanning a tape on which certain symbols appear." This is a complex organism which implements programs by converting information into symbols and then processing them based on the specified machine table.

[3] John R. Searle, *Philosophy in a New Century: Selected Essays* (Cambridge University Press, 2008), 61; Derk Pereboom and Hilary Kornblith, "The Metaphysics of Irreducibility," *Philosophical Studies* 63, no. 2 (1991): 125-145; Derk Pereboom, "Robust Nonreductive Materialism," *Journal of Philosophy* 99, no. 10 (2002): 499-531.

[4] John R. Searle, "The Critique of Cognitive Reason," in *Readings in Philosophy and Cognitive Science*, ed. Alvin Goldman, 833-847 (MIT Press, 1993), 834. See also Churchland, 745.

[5] Searle, *Philosophy in a New Century*, 70.

[6] Adams and Beighley raise some issues against Searle's perspective and present a different view on the issue. See Fred Adams and Steve Beighley, "The Mark of the Mental," in *The Continuum Companion to Philosophy of Mind*, ed. James Garvey, 64-72 (Continuum International, 2011), 66-67.

[7] Ibid., 72.

[8] Ibid., 62.

lation to the abstract mental states. As a fall-out of this assumption, one of the main questions addressed in the paper is; what implications do these have on the belief in anthropocentricism and the troubling issue of Search for Extraterrestrial Intelligence (SETI). Demonstrating with multiple realizability, the paper notes that, (1) there are multiples of appropriate implementing structures capable of implanting a particular abstract computational state; (2) anthropocentricism is an unassailable limitation to understanding the true nature of mental states and by implication intelligence; (3) it is very plausibly argued that there is a strong possibility of a multiplicity of alien/extraterrestrial intelligence, especially when intelligence is conceived as a function of an abstract mental state. Therefore, the paper argues that "intelligence" should be inclusively and widely defined to provide for a possibility of *coming across* alien intelligence.

II. "Necessity" as withoutness in the structural hypothesis

This section introduces Putnam's machine table. Machine table is what accounts for the functioning of the Turing machine. That is why for Putnam, the machine table describes any Turing machine.[9] This means that machine table is what instructs the machine on what to do when a particular input is received.

> The 'machine table' describes a machine if the machine has internal states corresponding to the columns of the table, and if it 'obeys' the instruction in the table in the following sense: when it is scanning a square on which a symbol s_1 appears and it is in, say, state B, that it carries out the 'instruction' in the appropriate row and column of the table (in this case, column B and row s_1). Any machine that is described by a machine table of sort just exemplified is a Turing machine.[10]

The idea of *Row* and *Column* may be appreciated in a sample machine table below, specifying implementable and computable functions.

		A	B	C	D
(s1)	I	*s1RA*	*s1LB*	*s3LD*	*s1CD*
(s2)	+	*s1LB*	*s2CD*	*s2LD*	*s2CD*
(s3)	blank Space	*s3CD*	*s3RC*	*s3LD*	*s3CD*

[9] Hilary Putnam, "Minds and Machines," in *Mind, Language, and Reality: Philosophical Papers, Volume 2*, ed. Hilary Putnam, 362-385 (Cambridge University Press, 1975), 365.

[10] Ibid.

For instance, if the instruction says, "if you read or scan 1 as input, print 11, proceed to scan the next square to your left, then shift to state B," the machine is constrained by this instruction and cannot perform otherwise. A typical interpretation of machine table instruction is given by Putnam.

> These instructions are read as follows: s_5L A' means 'print the symbol s_5 on the square you are now scanning (after erasing whatever symbol it now contains), and proceed to scan the square immediately to the left of the one you have just been scanning; also, shift into state *A*.[11]

This is an example of an implementable instruction which is contained in the machine table. Possible machine table instructions include: S_1LA, S_2LB, S_3RA, or S_4LD. Each of these programmes is an appropriate instruction. For instance, in the Putnam's computational hypothesis, S_3LB may read as; print S_3 on the square you as now scanning (after erasing whatever symbol it now contains) and proceed to scan the square immediately to the left of the one you have just been scanning, also shift into state B. This is the programme which this machine is to implement if it is appropriate to implement it.

In Putnam's hypothesis, this machine table is abstract and then can be implemented by multiple appropriate physical substrates. The machine programmed to implement an abstract computational instruction cannot do otherwise. It is a deterministic automaton. Notwithstanding, one of the main weaknesses identified against Putnam's computational hypothesis is that it is incapable of accounting for phenomenal experience, which is considered to be a necessary property of the mental states. The biologists argue that computational hypothesis is designed only as a mere symbol manipulation. Consequently, it appears to lack the properties sufficient for actualizing mental states. If this is true, it may plausibly question the possibility of a moral machine.[12] Let us assume that the complete nature of mental states is contained in the abstract machine table. It follows that a particular implementing system, expectedly, implements all that is contained in the instruction and for Putnam, that is all there is to being a mental state. This implies that phenomenal experience is part of mental states, then it is expected to be contained in the instruction and be eventually implemented. Just as phenomenal experience, so is intelligence to be understood as a function of an internal process which is also part of mental states. Being intelligent, therefore, depicts a state which, perhaps, is abstractly and computationally definable relative to

[11] Ibid.

[12] Michael Anderson et al., "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 177-202.

and being implemented by a relevant implementing structure. If this is true then it is also capable of being multiply realized by different appropriate implementing systems. How this is implemented turns out to be a function of each implementing structure.

Hilary Putnam's hypothesis that the same mental state can be realized by different brain states, and/or that the same brain state can realize different mental states, has become orthodoxy in the philosophy of mind.[13] It means that a token abstract mental state is capable of being implemented by more than one implementing structure.[14] That explains the possibility of a multiplicity of the equally possible and appropriate implementing substrates as well as corresponding expected raw experience and intelligent states. The attendant raw experience may range from phenomenal experience to *silicomenal* experience, or *metalomenal* experience, and so on, depending on the nature of a possible and appropriate implementing structure. Intelligent state will also manifest depending on the nature of the respective internal process of the implementing structure. That there is multiplicity of implementing structures suggests that no particular structure is necessary, if "being necessary" is strictly read in the sense that without the implementing structure, there could be no abstract (mental) state.

This implies that if "necessity" is defined in the sense of "withoutness," then no particular implementing structure is necessary in the nature of the abstract instruction. What this suggests is that a particular implementing system is only sufficient but not necessary in relation to the nature of the abstract mental state. If any of these were necessary, then using Kant's understanding of "necessity,"[15] the relationship should have been either that of identity or of containment.

The notion of containment is not relevant to the current discussion. But, dislodging the idea of identity requires some clarification. This is illustrat-

[13] Jerry A. Fodor, "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)," *Synthese* 28 (1974): 97-115; Lawrence Shapiro, "Multiple Realizations," *Journal of Philosophy* 97, no. 12 (2000): 635-654; William Bechtel and Jennifer Mundale, "Multiple Realizability Revisited: Linking Cognitive and Neural States," *Philosophy of Science* 66, no. 2 (1999): 175-207. See also Carl Gillett, "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences," *The Journal of Philosophy* 100, no. 11 (2003): 591-603; and Robert Francescotti, *Physicalism and the Mind* (Springer, 2014), 1-3. Ross, however, approached multiple realizability from the context of causal connection in biology; Lauren N. Ross, "Multiple Realizability from a Causal Perspective," *Philosophy of Science* 87, no. 4 (2020): 640-662.

[14] The notion of multiple realizability has been strongly criticized. See an example of such in Bechtel and Mundale, 1999. Betchel and Mundale suggests *Multiple Consilience* as an alternative which suits neuroscience in guiding our understanding of cognitive systems.

[15] The Kantian sense of "Necessity" is understood in terms of contradiction. For him, a denial of a necessary proposition raises a contradiction. For example, 'The woman is not a female." In the case under discussion, it is understood in the sense that it is impossible to have a particular implementing structure without the abstract mental state. Immanuel Kant, *Critique of Pure Reason*, trans. John M. D. Meiklejohn (J. M. Dent & Sons Ltd., 1978).

ed using these epistemic cases. (1) Only neurophysiological structure implements abstract mental state; (2) only 7+5=12, and (3) only object that extends is matter. A careful inspection of these three statements will reveal some truth. First, statement 3 appears to be necessarily true conceptually, where "necessity" is understood either directly with the relationship of identity or indirectly with that of containment. Even while recognizing Quine's naturalism, it appears impossible to arbitrarily deny the necessary truth of this statement without preparing to step into inconsistency or rupture our linguistic structure. Quine also admitted that statements such as "unmarried men are unmarried" are true come what experience may, only that every epistemic statement is a *de facto* member of a holistic epistemic system.[16] These are examples of statements which, according to Quine, form the nucleus of his system, hence are less susceptible to experience.

Statements 1 and 2 do not possess such sense of necessary truth as 3. For instance, statement 1 cannot be true any more than 2 is true. In the sense of identity, a simple translation of 2 says, for all X, if X is 12, then X is identical to 7+5. Going by the "identity" nomenclature employed here, there is an issue to be explained. This is saying that 12 and 7+5 are identical. That is, (a) 12 is the same thing as 7+5, and (b) nothing else could be 12 apart from 7+5. This breakdown apparently exposes the difficulty in maintaining that statement 2 is necessarily true even when we try to avoid the question raised by "same thing as". First, by mere token representation, it is clear that 12 and 7+5 are not identical. In fact, the correct numerical breakdown of 12 is 1 tens and 2 units, the summation of which is also represented by 10+2. Would that imply that 12 and 10+2 are more identical than 12 and 7+5? The point is that 10+2 is a direct componential breakdown of 12 rather than 7+5; it is not that one is more identical mathematically.

Second, in the strict sense of identity, 12 is neither identical to 7+5 nor to 10+2. This is because there are numerous mathematical relations which can be equal to 12. Examples are 6+6, 9+3, 8+4, 11+1, 13-1, etc. The point which becomes clear is that 12 is multiply realizable by different appropriate mathematical relations within the mathematical system. It follows that 7+5 and 10+2 are not necessary to realizing 12; they are only sufficient in the sense of necessity of an implementing system in relation to abstract mental state. The same thing applies to statement 1. Therefore, no one particular implementing structure is identical to and therefore necessary to the nature of the abstract mental state. Without electrochemical organic structure, abstract mental state can be implemented by other possible sufficient implementing systems. Therefore, if "necessity" is understood in the sense of "identity" or

[16] Willard Van Orman Quine, *From a Logical Point of View: Logico-Philosophical Essays* (Harvard University Press, 1961), 42-43.

"withoutness," then electrochemical-based structure is not *necessary* in the nature of the abstract mental state. However, when considering the internal process peculiar to a particular implementing structure, the issue of necessity may be raised in relation to what experience is produced no doubt. It is rarely in this sense that neurophysiological implementing structure might be deemed necessary in producing phenomenal experience which is peculiar to organic structure alone.

III. Effects of anthropocentricism on the hypothesis

The focus of this section is to show that anthropocentricism is a questionable inhibition against the adequacy of the required knowledge concerning the nature of mental state and by implication, intelligence. Anthropocentricism is a belief that human beings occupy the central determinant position in the universe. That is, and this has played out in all research attempts and inquiries; all phenomena, ontology, definitions and description are human dependent and determined. This belief is accentuated by the biblical depiction in Genesis chapter 2, verse 19.

> And out of the ground the Lord God formed every beast of the field, and every fowl of the air; and brought them unto Adam to see what he would call them: and whatsoever Adam called every living creature, that was the name thereof.

However, much as this appears to be soothing, it poses a danger in the sense that it constitutes a seemingly unassailable hindrance in the understanding of the real nature of mental states. By 'real' I mean the original human independent nature. This anthropocentric belief looms large over judgments, assertions, points of view, beliefs, etc., and it manifests in every attempt to pursue an inquiry into the nature of reality in every aspect of human inquiry. A thing is seriously and unapologetically assumed to be whatever human beings can prove or define them to be! Presumably, the issue of the nature of mental states will become pretty convenient to deal with, once we can conceive that strictly speaking, human being is only *a* componential part, and not the sole determinant of nature.

Building upon the thesis of multiple realizability, and an insight into anthropocentricism, there are some observable points which are congruent to the issue of discussion. First, a fallacy ensues if one were to assert that the capability of internal states leading to the possibility of mental states is only restricted to human electrochemical property. On the contrary, it is strongly assumable that there are multiples of implementing system in the world, with an appropriate implementing structure, which are capable of implementing

the abstract machine table leading to the possibility of internal states. Arguing otherwise may be running into *ad ignorantiam*.

Second, electrochemical implementing structure should caution against over-assumption that only this category, capable of realizing mental states, is able to possess phenomenal experience and thereby, for instance, feel pain or be intelligent. Plausibly, electro-metallic, or silico-metallic, etc., which are other possible systems might possess advanced processing structure just as, or more than, human's electrochemical system. Then, the question now is no longer; what category of the animate is capable of possessing mental states, but, rather; which category of systems, in nature, is incapable of implementing a particular abstract state *depending* on its own implementing structure? This is because the probability of the assumption that a silico-metallic system is able to realize an experience through its internal process, similar to human feeling of pain is very strong.

Third, as "we theorize that our universe may be rich with planets populated by intelligent beings who, like us, can search for evidence of other technological civilizations,"[17] it follows that being *intelligent*, which is considered a property of mental state, may be multiply realized by several appropriate implementing systems, again, relative to the nature of their implementing structure. This is because being intelligent, as already demonstrated, depicts a possible state which is abstractly and computationally definable relative to a relevant implementing structure. This state can also be multiply realized by different appropriate implementing systems.

The difficulty here is not in the assumption of multiple realizability, but in the perceived and troubling effects of anthropocentricism. This is because it is assumable that biological function of the electrochemical system might restrict it from ascertaining the possibility of other implementing structures capable of realizing intelligence beyond the level of mere assumptions. The level of mere assumption is the hypothetical level of attributing intelligence capacity to other systems from or by human judgment alone. To justify beyond the level of mere assumptions, there must be the ideal- implementing-structure whose system superintends overall. Nonetheless, this is what Putnam[18] argued that only God (whatever this may mean or refer to)[19] could

[17] Bernard M. Oliver, "The Windows of SETI – Frequency and Time in the Search for Extraterrestrial Intelligence," *The Planetary Report* 7, no. 6 (1987): 23-25.

[18] Hilary Putnam, *Representation and Reality* (MIT Press, 1988), 89.

[19] For the view that metaphysical anthropocentrism requires and implies monotheism and thus we are ontologically committed to God, see Åke Gafvelin, "No God, No God's Eye: A Quasi-Putnamian Argument for Monotheism," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 83-100.

have. Nagel's[20] question; "what is it like to be a bat?" is a mortal question indeed serving to justify this restriction. Therefore, when 'being intelligent' is seen as an abstract state of the mental, then there is nothing that says that other natural systems are incapable of realizing intelligence.

Just as Jackson[21] argues, the electrochemical system should therefore be wary of imposing human judgment on other objects or systems in the cosmos. The fact, if it is a fact, that electrochemical may not completely fathom the true nature of an internal process and experience of other systems does not warrant that they (other systems) should be conceived as incapable of realizing some states. This reasoning, correspondingly, challenges Tye's assumption; "Thus, when I feel pain, and I believe that I do, my Zombie replica believes that he feels pain too. It is just that his belief, unlike mine, is false."[22] This assumption is purely based on the privileged information that human beings have regarding the internal make-up and configurations of the zombie, nothing more.

IV. The existence of aliens/extraterrestrial intelligence: Redefining the question

The question which has been seriously troubling the *Homo Sapiens* is whether or not there is extraterrestrial intelligence out there. This question, definitely and quite clearly, has a place in this discussion. This is because intelligence is described as a function of internal state of an implementing system. Consequently, search for signs of life, alien cultures, and intelligence around the universe are part of the main scientific concerns. Existence of aliens is commonly believed to be a hoax and supported by rumors, in some quarters, some of which are through alien sightings with their disc/saucer-like-craft, which is being manned by the little green big-headed beings. However, the search for the possibility of extraterrestrial intelligence or life appears as a worthwhile scientific enquiry in the cosmos. This is what foregrounded the space inquiry and search into the universe which are within the purview of Search for Extraterrestrial Intelligence (SETI). Apparently, human curiosity accounts for more seriousness in the search. Our neural triggers push human curiosity to want to inquire, to want to meet, and, perhaps, interact with our cosmic friends around the universe! The discovery of advanced telescopes for deep penetrations into the hearts of planetary bodies in the universe, also increases this curiosity on the possibility to meet this possibly metallic-made little man.

---

[20] Thomas Nagel, "What Is It Like to Be a Bat?" *The Philosophical Review* 83, no. 4 (1974): 435-450.

[21] Frank Jackson, "Representation and Narrow Belief," *Philosophical Issues* 13 (2003): 99-112.

[22] Michael Tye, *Consciousness Revisited: Materialism without Phenomenal Concepts* (MIT Press, 2009), 191.

However, the point of note is that the methodology and the theoretical frameworks adopted in the search appear to be narrow regarding the conceptual construal of what 'intelligence' is. In line with the view of Slijepcevic and Wickramasinghe,[23] 'intelligence' is largely and restrictively defined from electrochemical point-of-view. This thesis has technically conceived 'intelligence' as a function in the electrochemical abstract mental state. In layman's terms, and simply put, intelligence is observed as a mental capacity which is exhibited by human beings, (electrochemical system), to enable them to appraise and solve problems. This point-of-view, however, explains the quantitative nature of the search. The term "quantitative nature" is employed in the sense of the belief that alien intelligence could be found if and when alien life is found, hence, the point-of-view, methodology and theoretical framework adopted. Whereas to appraise better, we may need to be ready to *sidestep* the electrochemical encumbrances. This is the real issue! This might help to review our methodology and reorient our technology. Reviewing our methodology raises the question of turning a search light towards ourselves with a view to re-examining how human cognitive ingenuity has produced the framework for the present methodology. This may also include evaluating the methodology of its possible limitations. Reorienting our technology is an inevitable and logical result of a sufficient reviewing of our methodology. This may become rewarding eventually as the question may need to be redefined for a more fruitful search.

Side stepping our electrochemical encumbrances may help to distinguish the question of a 'search for life' from a 'search for intelligence'[24]; the two questions which ordinarily appear similar but are sufficiently different. This is so because, according to the view of Hisabayashi,[25] there should be a distinction between a search for extraterrestrial life, and intelligence. However, whereas the definition of a search is expected to be initially and clearly clarified, there is no doubt that the two questions constitute genuine reasons to initiate inquiries into a search. Suppose for instance that human beings have been searching for signs of an intelligent electrochemical organism, on the

---

[23] Predrag Slijepcevic and Chandra Wickramasinghe, "Reconfiguring SETI in the Microbial Context: Panspermia as a Solution to Fermi's Paradox," *Biosystems* 206 (2021): 104441.

[24] See Nathalie A. Cabrol, "Alien Mindscapes – A Perspective on the Search for Extraterrestrial Intelligence," *Astrobiology* 16, no. 9 (2016): 661-676. Some researchers, however, actually defined their search towards extraterrestrial life in the universe. For example, see much more defined enquiry which is about the possibility of extraterrestrial life in Steven J. Dick, "NASA and the Search for Life in the Universe," *Endeavour* 30, no. 2 (2006): 71-75; Baruch S. Blumberg, "Astrobiology, Space and the Future Age of Discovery," *Philosophical Transactions of the Royal Society A* 369 (2011): 508-515; Carol E. Cleland, "Moving Beyond Definitions in the Search for Extraterrestrial Life," *Astrobiology* 19, no. 6 (2019): 722-729.

[25] Hisashi Hisabayashi, "An Encounter with Extraterrestrial Intelligence," *Biological Sciences in Space* 17, no. 4 (2003): 324-340.

assumption that only humans can be intelligent, then it is not impossible to find, just in case there are such organisms out there! This is but an easy problem because we actually have a fore-knowledge of what should constitute the object of the search! It may be noted here that search for signs of life dominates though.[26] Take for instance an infographic of a seven-level framework invented by National Aeronautics and Space Administration (NASA) to help people put "signs of alien life" discoveries in context.[27] The case is expected to be different when the search is narrowed to intelligent beings alone. This is because, eventually, there may be no necessary connection between being an extraterrestrial being and being intelligent.

One of the main technical implications of our findings in this paper is the strong possibility of multiples of appropriate implementing structure capable of intelligence. Even, the possibility of these intelligent structures existing around human domain is not ruled out. Curiously, this seems challenging in the sense that it may suggest that any object around us may be capable of implementing an abstract program to realize intelligence, once it possesses an appropriately implementing structure. Inclusively, the term *alien* may have to be redefined to include any non- human system with capable implementing structure, which is able to implement similar states as electrochemical structure. So, this is it! It follows that man's intelligence is a token realization of the abstract intelligence in the universe. Following this consistent implication, there is therefore no doubt to the possibility that multiples of other appropriate implementing structures, capable of implementing and then manifesting intelligence in the universe, exist. It may be, therefore, strongly inconsistent to presume, either that only human beings are intelligent, or that man's intelligence superintends over, or determines the nature of other *intelligence(s)*. Both conjuncts are assumptions which evidently run deep into the dungeon of anthropocentricism, the dungeon which implicitly inhibits man's freedom to really appreciate, investigate, and truly explore *what there is*. Therefore, absolute reliance on human conceptualizations, hypothetical conjectures, and methodology framework, defined signs, and properties of life, might significantly make it pretty difficult for human beings to ever correctly apprehend and appraise the nature of extraterrestrial intelligence and even life.

Though this hypothesis arguably supports the existence of multiple species (appropriate implementing systems) of intelligence, a pressing question is; can

---

[26] See Nathalie A. Cabrol, "The Coevolution of Life and Environment on Mars: An Ecosystem Perspective on the Robotic Exploration of Biosignatures," *Astrobiology* 18, no. 1 (2018): 1-27.

[27] Matthew Hart, "New NASA Chart Puts Signs of Aliens Reports into Context," *Yahoo! Entertainment*, https://www.yahoo.com/entertainment/nasa-chart-puts-signs-aliens-124335774.html?.tsrc=fp_deeplink.

human intelligence ever come across alien intelligence? From this hypothesis, though it may appear unsupported that a silicon- based structure might fully make sense of the neural workings of the metallic-based structure or electro-chemical-based structure, it may not be impossible, especially when we can *review* our methodology and *reorient* our technology. That means, there is, first, the need to deal with ourselves before launching our search or research out. This owes to the fact, if it is a fact, of asymmetrical nature and structures of various implementing systems. Electrochemical structure will define other structures by its own limited and narrow methodology and conceptualizations. Ditto for other implementing structures! But electrochemical structure is not the only structure that can realize intelligence in the universe. What turns out to be clear is that man's view about what counts as intelligence and its signs is still critical and could be redefined. This is to say that a great deal is still necessarily required for the search to produce the desired result.

V. Conclusion

Multiple realizability features so prominently in Putnam's abstract computational approach to the issue of the nature of mental states. Whereas this has been variously criticized and, in some cases, rejected by the identity theorists. They argue that the implementing physical structure is necessary in understanding the nature of mental states and any theory which ignores this is insufficient. In the relationship between the abstract states and the implementing structure, the paper demonstrates how multiple realizability shows that no particular implementing structure is necessary.

Arising from the hypothesis, if 'intelligence' is a function of the mental state, the paper deduces the possibility of multiple intelligent systems in the universe, where human intelligence is just a unit. The paper, therefore, challenges the basis of anthropocentricism which appears as an inhibiting and limiting factor in the search for the real nature of mental state and of reality. This paper argues that anthropocentricism could be overcome when we can redefine 'intelligence,' review our methodology, and reorient our technology to help the Search for Extraterrestrial Intelligence (SETI) to be more fruitful.

References

Adams, Fred, and Steve Beighley. "The Mark of the Mental." In *The Continuum Companion to Philosophy of Mind*, edited by James Garvey, 54-72. Continuum International, 2011.

Anderson, Michael, Susan Leigh Anderson, Alkis Gounaris, and Georgios Kosteletos. "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson." *Conatus –Journal of Philosophy* 6, no. 1 (2021): 177-202.

Bechtel, William, and Jennifer Mundale. "Multiple Realizability Revisited: Linking Cognitive and Neural States." *Philosophy of Science* 66, no. 2 (1999): 175-207.

Block, Ned. "The Computer Model of the Mind." In *Readings in Philosophy and Cognitive Science*, edited by Alvin Goldman, 819-832. MIT Press, 1993.

Blumberg, Baruch S. "Astrobiology, Space and the Future Age of Discovery." *Philosophical Transactions of the Royal Society A* 369 (2011): 508-515.

Cabrol, Nathalie A. "Alien Mindscapes – A Perspective on the Search for Extraterrestrial Intelligence." *Astrobiology* 16, no. 9 (2016): 661-676.

Cabrol, Nathalie A. "The Coevolution of Life and Environment on Mars: An Ecosystem Perspective on the Robotic Exploration of Biosignatures." *Astrobiology* 18, no. 1 (2018): 1-27.

Churchland, Patricia. "The Co-evolutionary Research Ideology." In *Readings in Philosophy and Cognitive Science*, edited by Alvin Goldman, 745-768. MIT Press, 1993.

Cleland, Carol E. "Moving Beyond Definitions in the Search for Extraterrestrial Life." *Astrobiology* 19, no. 6 (2019): 722-729.

Dick, Steven J. "NASA and the Search for Life in the Universe." *Endeavour* 30, no. 2 (2006): 71-75.

Fodor, Jerry A. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28 (1974): 97-115.

Francescotti, Robert. *Physicalism and the Mind*. Springer, 2014.

Gafvelin, Åke. "No God, No God's Eye: A Quasi-Putnamian Argument for Monotheism." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 83-100.

Gillett, Carl. "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences." *The Journal of Philosophy* 100, no. 11 (2003): 591-603.

Hart, Matthew. "New NASA Chart Puts Signs of Aliens Reports into Context." *Yahoo! Entertainment*. https://www.yahoo.com/entertainment/nasa-chart-puts-signs-aliens-124335774.html?.tsrc=fp_deeplink.

Hisabayashi, Hisashi "An Encounter with Extraterrestrial Intelligence." *Biological Sciences in Space* 17, no. 4 (2003): 324-340.

Jackson, Frank. "Representation and Narrow Belief." *Philosophical Issues* 13 (2003): 99-112.

Kant, Immanuel. *Critique of Pure Reason*. Translated by John M. D. Meiklejohn. J. M. Dent & Sons Ltd., 1978.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review* 83, no. 4 (1974): 435-450.

Oliver, Bernard M. "The Windows of SETI – Frequency and Time in the Search for Extraterrestrial Intelligence." *The Planetary Report* 7, no. 6 (1987): 23-25.

Pereboom, Derk, and Hilary Kornblith. "The Metaphysics of Irreducibility." *Philosophical Studies* 63, no. 2 (1991): 125-145.

Pereboom, Derk. "Robust Nonreductive Materialism." *Journal of Philosophy* 99, no. 10 (2002): 499-531.

Putnam, Hilary. "Minds and Machines." In *Mind, Language, and Reality: Philosophical Papers, Volume 2*, edited by Hilary Putnam, 362-385. Cambridge University Press, 1975.

Putnam, Hilary. *Representation and Reality*. MIT Press, 1988.

Quine, Willard Van Orman. *From a Logical Point of View: Logico-Philosophical Essays*. Harvard University Press, 1961.

Ross, Lauren N. "Multiple Realizability from a Causal Perspective." *Philosophy of Science* 87, no. 4 (2020): 640-662.

Searle, John R. "The Critique of Cognitive Reason." In *Readings in Philosophy and Cognitive Science*, edited by Alvin Goldman, 833-847. MIT Press, 1993.

Searle, John R. *Philosophy in a New Century: Selected Essays*. Cambridge University Press, 2008.

Shapiro, Lawrence. "Multiple Realizations." *Journal of Philosophy* 97, no. 12 (2000): 635-654.

Slijepcevic, Predrag, and Chandra Wickramasinghe. "Reconfiguring SETI in the Microbial Context: Panspermia as a Solution to Fermi's Paradox." *Biosystems* 206 (2021): 104441.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433-460.

Tye, Michael. *Consciousness Revisited: Materialism without Phenomenal Concepts*. MIT Press, 2009.

C. C. Ramos · *Conatus* 10, no. 1 (2025): 207-219
doi: https://doi.org/10.12681/cjp.38743

Understanding Love in Filipino Culture: An Examination of Indigenous Perspectives and Cultural Reflections

Christine Carmela Ramos
Mapua University, Philippines
E-mail address: christinecarmela68@gmail.com
ORCID iD: https://orcid.org/0000-0002-0851-7069

Abstract

This study explores Filipino cultural and political contexts through the lens of Filipino Indigenous thought, focusing on the philosophical underpinnings of the concept of "loob" and its influence on the understanding of love (pag-ibig). "Loob" is examined as a route to achieving harmony with others and nature, aiming for unity with the divine. The research distinguishes between two key dimensions: the analytic (interior) and the synthetic (holistic). The analytic dimension emphasizes the inherent goodness within individuals. In contrast, the synthetic dimension offers a holistic perspective, crucial for addressing Filipinos' multifaceted challenges in a diverse and pluralistic society. The paper highlights the importance of incorporating nonviolence into literature, the arts, and education, arguing that this integration fosters mature humanity within a rapidly evolving global consciousness. The articulation of nonviolence in these domains is presented as a crucial step toward achieving a more just and harmonious society. Furthermore, the study draws a philosophical comparison between the nonviolent efforts of Corazon Aquino and Mahatma Gandhi, examining their roles in pursuing freedom through nonviolence. Gandhi's application of ahimsa (doing no harm) as a tool for civil protest is analyzed in the context of the dynamic processes of societal control and justice he confronted. His leadership in resisting colonial rule, leading the Indian rebellion, and challenging discriminatory policies is contrasted with Aquino's efforts toward political and social change. The paper argues that Gandhi's philosophical approach aligns closely with Aquino's strategies for emancipatory political achievement and justice through nonviolence, underscoring their shared commitment to these enduring ideals.

Keywords: *Filipino indigenous philosophy; loob; pag-Ibig; love; non-violence; virtue*

I. Introduction

The author grew up surrounded by stories of rich cultural history in the Philippines. Her ancestors often told stories about traditions and beliefs that seem far removed from today's world. Throughout her academic career, the author realized the importance of combining this wis-

dom, which appears to be taken for granted in modern life. Societies in Southeast Asia, and especially the Philippines, are rapidly developing. Schools often struggle to make meaningful connections between the past and the present. This paper promotes indigenous wisdom to prepare students better to navigate and contribute to a more harmonious, inclusive, and sustainable world. Corporations, consumerism, mass media, anonymity, and individualism dominate our public culture, casting a shadow over the civic participation and common good that once flourished in our communities.

Writings by Leonardo Mercado, Florentino Timbreza, and Rolando Gripaldo popularized the quest to articulate Filipino indigenous thoughts. At the same time, several studies focused on the Filipino concept of love. Among the latter, though, there are distinct trends observed in the Philippines over the years, such as the debate on divorce and LGBTQ marriage. Those studies seek to defend claims that are important and illuminating. They highlight some of the unique challenges Filipinos face in a pluralistic society. The studies are worthwhile, address the viability of various claims, and navigate numerous difficulties.

In this work, the author investigates Filipino indigenous beliefs that serve as a humble means to avoid the irrelevance of Filipino culture and heritage in a fast-paced world.

i. This paper discusses and appreciates Filipino indigenous thoughts and how they shape our nation's history. Further, this paper explores the meaning of love (*pag-ibig*) in a globalized world based on the Filipino's Indigenous concept of "*loob.*" "Loob" encompassed Filipinos' humanity and daily experiences. It aspired to harmony with others and nature to be one with God, explaining the Filipinos' dualism in body-soul and emotional-rational. This work briefly discussed the two dimensions of "*loob"*: analytic (interior) and synthetic (holistic). The interior "loob" affirmed the innate goodness of the person. The holistic model represented the world as a whole entity and the world's non-dualistic perspective on crucial topics.

ii. Secondly, it is crucial to articulate nonviolence – locally, nationally, regionally, and globally – in the face of an intensifying global consciousness. In this vein, the author draws parallels between Corazon Aquino and Mahatma Gandhi in terms of nonviolent actions and a deep yearning for national liberation.

a. Statement of the problem

This paper explores the concept of "loob" through the author's reflections and experiences rather than as a traditional research report. The aim is to

provide a commentary on Filipino cultural and philosophical perspectives, focusing on how these reflections have reshaped the interpretation of research findings. The study is structured first to introduce "*loob,*" which is inclusive, encompassing everyday experiences and humanity. She mainly explores the definition of love (*pag-ibig*) through this concept. This approach does not reduce the depth of analysis but enhances the understanding of Filipino cultural values through personal insights and reflections.

b. Methodology

The Philippine Archipelago is one of the richest, and its tropical destinations are often voted the most beautiful on earth. As of 2022, there are over 7,641 islands in the archipelago, around 2,000 of which are inhabited. This study undertakes the need to re-interrogate the ethics of Filipinos to fully understand the cultural and political contexts.

Colonization has had a long-standing impact on Philippine culture and traditions. This impact goes far beyond language, food, and the many superstitions locals hold dear. With an Animist, pre-colonial past (with likely Hindu-Buddhist influences), followed by a successful conversion to Christianity, the Philippines claims ownership of a fascinating, eclectic set of beliefs.

The Augustinians preached the Gospel in Cebu, Panay, and other Visayan islands, while the Franciscans evangelized Manila and Bicol provinces. Conversely, the Recollects evangelized in Bataan, Mindoro, and other parishes on the Pacific coast of Luzon. In the 1500s, most churches, schools, asylums, and hospitals were erected; they existed after earthquakes and typhoons. These constructions are built in Manila, other towns, and provinces. In roughly three hundred years, 12,000 missionaries evangelized the Philippines. Some died as martyrs in the Philippines, China, Japan, and other Eastern countries.[1]

There are folk sayings in this work that need to be clarified. Some of the quotes in this study are grounded in regional beliefs. If the author mentions *Boholanos* in this paper, the term refers to those living in Bohol's island province. They are part of the wider Bisaya ethnolinguistic group. Four of their churches have been declared National Cultural Treasures for cultural, historical, and architectural importance.

While those living in Cebu are *Cebuanos,* the term *Cebuano* refers to the permanent residents on Cebu Island regardless of ethnicity. Magellan's priests' first recorded conversion occurred on this island.

[1] Anchi Hoh, *Catholicism in the Philippines During the Spanish Colonial Period 1521-1898*, July 10, 2024, https://blogs.loc.gov/international-collections/2018/07/catholicism-in-the-philippines-during-the-spanish-colonial-period-1521-1898/.

Figure 1. Magellan's Cross, 1521, Cebu

On the other hand, *Ilocanos* primarily reside within the Ilocos Region in the northwestern seaboard of Luzon. They speak Iloko or Iloco language. Ilocos attracts devotional tours for the faithful to basilicas such as St. Nicholas of Tolentino Parish Church, which existed for almost 440 years. On the other hand, *Ilonggos* refers to the people living in Iloilo province. Iloilo was also one of the most prominent episcopal during the Spanish colonial period.

Tagalogs are the largest cultural-linguistic group in the Philippines. They form the dominant population in the city of Manila. Finally, the *Kapampangans* live in the provinces of Pampanga.[2] Most were religious, though some of their shared beliefs survive today, such as spirits who reside in mounds (*nunu*) or nocturnal giants (*kapre*).

II. Moral, religious, and relational dimensions of "*loob*"

The condition of philosophy in the Philippines is divided into two stages. First, it is restricted in the academic setting of university classes. Second, it is in the process of formalization and mass adoption. There are two types of "loob": analytical concepts and synthetic concepts.

The analytical concept addresses morality, whereas the synthetic focuses on wholeness or the process of integration. Similar to other Eastern outlooks, Filipinos' "*loob*" is non-dualistic. There is emotional-rational and body-soul harmony, which aspires to unity with God. For example, the Indian doctrine

[2] Rosita Munoz-Mendoza, *The Kapampangan*, March 11, 2022, https://ncca.gov.ph/about-nc-ca-3/subcommissions/subcommission-on-cultural-communities-and-traditional-arts-sccta/northern-cultural-communities/the-kapampangan/.

of *Anatta*, which denies the inner self or the Atman, is intended to be in union with the Absolute.[3] Conversely, harmony between humanity, nature, and God is the essence of "*loob*."

"*Loob*" stresses being with others and sensitivity to their needs, inhibiting one's own personal and individual fulfillment.[4] "*Loob*" is connected to emotions.

Personalism and a high regard for smooth interpersonal relationships (SIR) are some of the stereotypical images of Filipinos. Personalism features the Filipino's emphasis on personalities and personal factors. For instance, politicians invite movie personalities to support their candidacy during election campaigns. Thus, meanings are not in words but in people.

Most Filipino voters prioritize the political theater and pomp associated with the candidate over their political agenda. The SIR manifests in reciprocity and "*Pakikisama*," or relationships.[5]

In times of need, Filipinos seek the support of relatives. Wealthier family members distribute their riches to neighbors and relatives out of "pakikisama." However, the giver is likely to predetermine what form should be of more excellent value than the original debt and should be repaid.[6]

Timbreza suggests that the authoritarian tradition that permeates our strong family system may be the reason for fully integrating the individual into the group.[7] The individual is advised what to do and follows the elders' advice. In short, the Filipino is group-oriented.[8]

"*Loob*" encompasses Filipinos' humanity, personality, theological perspective, and daily experiences. "*Loob*" aspires to harmony with others and nature to be in union with God. Thus, the interpersonalistic characteristics of "*loob*" explain the non-dualism in body-soul and emotional-rational of the Filipinos.[9]

In Philippine history, several Catholic orders strongly affected the notion of "loob," notably the Augustinians, Recollects, Franciscans, and Dominicans, who played critical roles in the archipelago's Christianization. These missionaries came to the Philippines not as colonists, as the region provided little money or spices. Nevertheless, Filipinos underwent substantial adaptation during this time, including acquiring new knowledge, languages, cultural

---

[3] Leonardo Mercado, *Elements of Filipino Philosophy* (Divine Word Publications, 1974), 3.

[4] Ibid., 107.

[5] Florentino Timbreza, *Mga Hugis ng Kaisipiang Pilipino* (Tomas Press, 1989), 167.

[6] Thomas Church, *Filipino Personality: A Review of Research and Writings* (DLSU Press, 1986), 38.

[7] Timbreza, 23.

[8] Emerita Quito, *Three Women Philosophers* (Diamond Jubilee Publications, 1986), 13.

[9] Mercado, 5.

practices, and belief system shifts. This historical context sheds light on the faith-based activities that influence these tremendous changes.

III. Reconceiving love: A philosophical inquiry into the Filipino indigenous concept of "*loob*"

In this section, the author will feature one of the central issues of the ethical views of Filipinos – the virtue of love. Broadly, love is divided into two types: authentic and inauthentic. Genuine love gives meaning to life. According to the Ilocanos, "Love is the comfort of life." (*Aliw sa buhay ang pag-ibig*.) Similarly, the Ilonggos believe that love is the salt of life (*Ang pag-ibig ay siyang asin ng buhay.*) Like a dish that needs to be seasoned, our lives need love to make it more fun and exciting. At this point, despite humanity's hardship and sorrow, there is happiness amid one's toil and suffering because of love.

Francisco Balagtas (1788-1862), the William Shakespeare of the Philippines, poetizes about the power of love in these words:

> Oh, love is so powerful...
> ...once you get in the heart of anyone,
> everything will be defied
> if only you give in to your wishes.[10]

The saying of Balagtas was based on his circumstances.[11] Those lines symbolize love's magnitude, which is overwhelming. The same sentiment is shared by Ilocanos, who believe that one endures to one who genuinely loves. (*Kung tunay kang umibig, tiisin mo ang hirap.*) A true lover ignores failure for the beloved's sake. Alfred Tennyson said, "It is better to fall in love and fail than never fall in love."

True love is sweet; it is fun and exciting. The Tagalogs say, "If love is real, it will be sweet until the end." (*Kung tunay rin lang ang pag-ibig, hanggang katapusa'y matamis.*) Filipinos also believe that not in word alone but in deed, can true love be assured. Further, Ilocanos say, "If love is genuine, there would be no gossip" (*Kung talagang tunay ang pag-ibig, wala ng marami pang satsat*).

Filipinos also have the concept of dishonest and inauthentic love. Hypocritical love will not last. The Kapampangans say, "Superficial and hypocritical love is like smoke" (*Ang pahapyaw at mapagkunwaring pag-ibig ay parang usok na maglalaho*). Phony love is unwilling to suffer; it is exploitative

---

[10] Originally: *O pagsintang labis ang kapangyarihan / Sampung mag-aama'y iyon nasasaklaw; / pag ikaw ay nasok sa puso ninuman, / hahamakin lahat masunod ka lamang.*

[11] Francisco Balagtas, *Florante at Laura* (Jiahu Books, 2016), 80.

and deceitful. It quickly disappears during hardship and deprivation. Fake love is false; it has evil intent and masquerades.

Love plays a part in Filipinos' values, the emotional aspect sometimes running more deeply than the intellectual dimension. Since Filipinos are passionate, a significant element of daily life is affected. Choices usually revolve around what a person likes or dislikes.

However, in Filipino culture, love involves friends, siblings, and parents. As Cebuanos says, "it is easier to block flood than love" (*Mas madaling pigilin ang alon (o baha) kaysa pag-ibig*). For instance, if one's circle of friends does not approve of the suitor, the suitor will not be treated well or might be belittled.

On the other hand, Filipinos can fall in love with the help of acquaintances or a circle of friends. For example, if a man likes a woman, he usually finds more information from her best friend or relatives (e.g., her favorite food and hobbies). Traditionally, a suitor's friend accompanies him as he woos the woman.

Further, love is usually about helping people in Filipino life, based on the synthetic concept of "loob." Love transforms. For instance, those who do not want to study will struggle out of love. Those who do not groom themselves will be inspired, too. Those who used to have many unpleasant habits will pray or attend mass.

Too often, too much love becomes harmful: it can be highly toxic. We read in headlines about people being killed out of jealousy. Whether fair or not, love strongly motivates a person to do good or bad. According to the Kapampangans, "Whoever is most loved, sacrifices most" (*Kung sino ang pinakamamahal ay siyang pinahihirapan*). Could the intense emotions involved in love lead to such extremes? Though not accepted by the woman's circle of friends, the suitor shows his loyalty by persevering.

However, the disapproval of the woman's friends might sway her to reject the suitor even though she may like him. Therefore, the relational dimension of "loob" becomes evident in this instance, e.g., camaraderie and shame. In this case, friendship influenced her decision rather than her personal feelings. Although she may have genuine affection for the suitor, the social pressure exerted on her decision is essential.

"Double standard" is also rampant among Filipinos. The father might admonish his son against adultery while committing it himself. The law of Christianity forbids adultery, but it is committed by the "poor and uneducated" and by the wealthy. The media's attention will be drawn, for instance, to a government official having liaisons and children out of wedlock. Perhaps the "double standard" can be based on a historical perspective.

Before the arrival of the Spaniards, it was recorded that our Filipino ancestors had many wives. However, the advent of Christianity in our country

introduced that having only one wife was pleasing to God and man, significantly altering the moral framework around marriage. However, in the author's observation, the media promotes adultery. For example, some films or shows portray a married police officer as having another woman. He is not necessarily despised by society. Instead, his masculinity ("machismo") is enhanced.

Further, love for Filipinos can be based on socio-economic standing. It is not a secret that the rich marry their kind, or actors marry fellow actors. Business or political beliefs bind them. If a rich man marries a woman of a lower class, he is sometimes frowned upon. Love then is measured by social class or outward appearance. Although ideal love is unparalleled, it is marketable, distinguished by wealth, power, or age.

Although Filipinos are staying married together, the new generation is becoming aware of the readily available transient love. However, such love for the Boholanos is *love that lies and quickly fades (Ang sinungaling na pag-ibig ay madaling kukupas*).

For the author, love's true meaning is that it directs our lives. If there is no love, there will be no civilization. We will surely perish if we do not love and understand. We learned to turn our energies to pursue other endeavors such as music, philosophy, and mathematics because of emotions, e.g., pity and joy. It becomes clear that love plays a critical role in shaping our pursuits and achievements.

At this stage, the author observes commonalities between the Filipino concept of love and Greek interpretations. Filipino love often embodies a blend of passion and community, similar to Greek *eros* (romantic love) and *philia* (friendship), where emotional intensity and social circles play crucial roles. However, Filipino love also reflects aspects of *agape* (selfless, unconditional love) through its emphasis on personal transformation and helping others. It shows how love permeates individual and communal dimensions in ways akin to the Greek ideal of loving one's neighbor selflessly.

VI. Filipinos' worldview on nonviolence: Rethinking liberty and greatness

The concept of love is not limited to romantic relationships. It also refers to relationships with family, friends, and God. For example, in a family with close siblings, a brother may exercise excessive protection, constricting his sister's suitor or boyfriend. Extreme love can contribute to why siblings may have conflicts. Therefore, the saying of the Ilocanos is true, "Too much tenderness makes the heart cry" (*Ang labis na lambing ay siyang nagpapaiyak sa puso*). Indeed, love should be moderated. Aristotle, the Buddhists, and Taoists say that any excess is wrong in fundamental relationships where balance is crucial.

Love is mighty. According to the Bible, love is the greatest, combined with faith and hope (John 3:16). True love lasts and understands. If there

are unclear things, love dispels them. Love makes us happy. Whether there are shortcomings or weaknesses, love conquers them. Indeed, love boosts a person's perspective and self-confidence. Love has continued to mark the path of human experience, guiding us through life's challenges and triumphs.

For Filipinos, the virtue of love is expressed through deeds. The Tagalogs say, "Love is in deeds, not words." Unlike Westerners, who are very vocal in expressing their thoughts, and for the cynical, the word "love" might be belittled. Indeed, love is like a spark or perfume that spreads fragrance to others. As "Pass it on" imparts:

It only takes a spark
to get the fire going
and soon, all those around
can warm up in its glowing.

Only in love can there be change – not in force or revolution. The seemingly serious problems are eased because of love. The person who knows how to love achieves what he wants because of sincerity or authenticity.

At this juncture, President Corazon Aquino is one of the most remarkable people in the Philippines' history because of her dedication to restoring the country's democratic status from President Ferdinand Marcos's dictatorship.[12] Aquino became an agent of reform by making a mark on society that will remain a legacy for future generations. Her efforts to promulgate nonviolence to combat all the odds inspired many. The power towards resolution is in the people's hands, who voice their rights and opinions to their oppressors. Because of her sacrifices to make a difference, she got more people to trust, love, and believe in her.

Aquino, the Philippines' first female president, vowed to the Filipinos that she would not fight fire with fire. On the contrary, she urged them towards a nonviolent revolution, popularly known as the EDSA People Power. She believes real power is not in the government's tanks, artilleries, or weapons. Instead, power belongs to the people, and they can defeat tyranny without the need for violence. The saying "Flowers defeat guns" (*Daig ng bulaklak ang baril*) during this dark period becomes apparent.[13] Accordingly, this sentiment guided Aquino's approach and inspired the nonviolent movement that defined her presidency.

Her desire to end the Marcos regime is rooted in the assassination of her husband, Senator Benigno 'Ninoy' Aquino Jr. It had been a very controversial

[12] Marcelo Ordonez, *People Power: A Demonstration of Emerging Filipino Identity* (Sampaguita Printing Press, 1986), 17.

[13] Christine Carmela Ramos, *Introduction to Philosophy* (Rex Bookstore Inc., 2019), 156.

occurrence in our history. Initially, Aquino showed no interest in entering politics, but the people supported her candidacy for the opposition. She envisioned becoming the agent of change, to liberate the Philippines from dictatorship. The country had a malicious 'Snap Elections' in which the results were manipulated and subjected to significant election fraud.

The Filipinos' collective effort mirrors the analytical and synthetic concepts of "loob." Unity will serve as people's power to overwhelm the odds and successfully win the battle. In February 1986, Filipinos marched along EDSA with church advocates and others who yearn for Philippine democracy. They held hands, prayed, and initiated civil rights movements.[14]

At this point, the author compares the similarities between Aquino and Mahatma Gandhi – their participation in nonviolent movements and a strong desire to attain freedom for their nation. Both supported peace despite challenging times, though in different circumstances. Notably, both figures are discriminated against in their passive resistance undertakings: Aquino for her gender and Gandhi for his race and religion. These prejudiced treatments did nothing but strengthen their will to advocate for rights. Further, due to their humane movements, many consider them role models for peace.

During Gandhi's thirty years as the leader of India's freedom struggle, he spent significantly more time educating the practice of nonviolence. He adopted the Buddhist, Hindu, and Jain religious ideal of *ahimsa* (doing no harm) as a nonviolent weapon for civil protest. He utilized it to oppose control over the region and societal issues like racial prejudice and injustices. Gandhi started massive protests against the colonial administration, coordinated Indian rebellion, and challenged anti-Indian policies in the judiciary. He became a prominent social figure and spread *Satyagraha*, a concept of nonviolent non-cooperation. Although Aquino was not imprisoned, she successfully led the People Power Revolution, ousting former President Marcos' two-decade dictatorship.

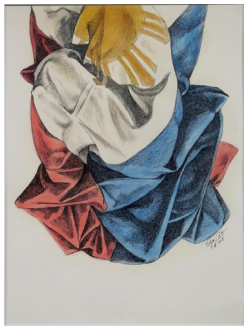


Figure 2. "Ode to the Flag," BenCab Museum, Philippines

[14] Dionisio Miranda, *Loob: The Filipino Within* (Divine Word Publication, 1989), 65.

The Filipinos' perseverance reinstated the lost democracy and the collapse of Marcos's regime.[15] Despite the success of her movement, Aquino is aware that applying the concept of nonviolence itself has limitations. Though the Marcos regime disintegrated, Marcos's loyalists censured the new Aquino government. These loyalists will spark threats to the peacefulness of the country. Aquino issued measures to regulate when to pull the trigger in using weapons in case of insurgence from the opposition. The Filipinos' trust in the nonviolent actions of the Aquino government has not been well established from the outset. Rebels and the opposition might pose a severe threat to peace and order. Aquino knows this fact, so she has set regulations and strictly monitored the overall condition of the country. After all, Aquino envisions a unified democratic government in which people have the means to live safe and sound in a 'nonviolent' society, which likewise purports to ensure the nation's stability and harmony. Accordingly, this vision argues for balancing nonviolent ideals with practical measures, making it relevant to this issue and guiding her actions throughout her presidency.

V. Conclusion: Bridging gaps of indigenous wisdom and global understanding

The memories of Aquino and Gandhi remain. Not how much they have given allows them to have more in return. Instead, it is about how they lived by sharing their resources, knowledge, and efforts to make a difference. They earned respect from the people who gave their trust to them. No matter how hard it is to succeed, remaining faithful to the principles, values, and viewpoints that inspired them to set their goals becomes essential. Additionally, they highlight non-violence, global understanding, cooperation, and core values in supporting social accountability.

Having virtue guides us to a better life.[16] Aristotle's concept of a virtuous existence is related to human connections cultivating mutual respect and a collective dedication to ethical principles. These connections are essential for an individual's ethical development and well-being. He envisions as the ultimate expression of human success what is beyond primal urges: profound ties of friendship and affection. In this way, love becomes a moral power.

It is a powerful virtue to choose and do good. It binds us to those who are close to us. Love makes us human. Since we can love, we are separated from the category of animals. In Filipino culture, love is influenced by oth-

[15] Leonardo N. Mercado, *The Filipino Mind – Philippine Philosophical Studies II* (The Council for Research in Values and Philosophy, 1994), 19.

[16] Purissima Emelda Egbekpalu, "Aristotelian Concept of Happiness (Eudaimonia) and it's Conative Role in Human Existence: A Critical Evaluation," *Conatus – Journal of Philosophy* 6, no. 2 (2021): 78.

er values, such as shame, camaraderie, "double standard," and "machismo." These values are closely related to how love is expressed in Filipino culture. However, excessive or unbalanced love can stray from the ideal, highlighting the need to moderate it.[17]

The Greek concepts of love – *eros* (romantic passion), *philia* (deep friendship), and *agape* (selfless love) – also illustrate the importance of balance in love.[18] Kobow suggests that erotic love acts as a channel for examining the profound aspects of human experience, providing a route to self-discovery and individual development. She recognizes the intrinsic contradiction in this pursuit: although *eros* propels us toward the eternal, it also confronts our mortality and the transience of existence. The intricate nature of our desires and ambitions mirrors both our fleeting nature and an expression of our inherent desire for lasting significance.

Eros and *philia* resonate with Filipino expressions of love, where emotional depth and social connections are paramount. Meanwhile, *agape* reflects the transformative and altruistic aspects of love seen in Filipino values of personal growth and social responsibility. These relationships strengthen ethical integrity and act as vital elements of human flourishing. The contradiction between the fleeting essence of desire and the human quest for permanence is how love links impermanence and meaning.

Which virtue truly matters? Would you live in a world without freedom and greatness for security and comfort? Human life is not just made for comfort. Liberty and excellence could not be achieved by being comfortable lest we seek less development. In the face of severe disagreements, love must be upheld. It lets us act morally.

References

Agapay, Ramon. *Ethics and the Filipino*. National Bookstore Inc., 1991.
Aguilar, Pido. *The Gift of Abundance*. Tahanan Books, 2010.

Alejo, Alberto. *Tao Po! Tuloy Kayo!* Ateneo de Manila University, 1990.

Balagtas, Francisco. *Florante at Laura*. Jiahu Books, 2016.

Bulatao, Jaime. *Split Level Christianity*. Ateneo University Press, 1992.

Cabrera, Benedicto. *Ode to the Flag*. Bencab Museum Philippines, 2015.

Church, Thomas. *Filipino Personality: A Review of Research and Writings*. DLSU Press, 1986.

[17] Ramos, 203.

[18] Beatrice Sasha Kobow, "The Erotic and the Eternal: Striving for the Permanence of Meaning," *Conatus – Journal of Philosophy* 6, no. 2 (2021): 216.

Egbekpalu, Purissima Emelda. "Aristotelian Concept of Happiness (Eudaimonia) and its Conative Role in Human Existence: A Critical Evaluation." *Conatus - Journal of Philosophy* 6, no. 2 (2021): 75-86.

Hoh, Anchi. "Catholicism in the Philippines During the Spanish Colonial Period 1521-1898." *4 Corners of the World International Collection*, July 10, 2022. https://ncca.gov.ph/about-ncca-3/subcommissions/subcommission-on-cultural-communities-and-traditional-arts-sccta/northern-cultural-communities/the-kapampangan/.

Kobow, Beatrice Sasha. "The Erotic and the Eternal: Striving for the Permanence of Meaning." *Conatus - Journal of Philosophy* 6, no. 2 (2021): 213-236.

Mclean, George. *Philosophy and Christian Theology*. Cultural Press, 1970.

Mercado, Leonardo N. *The Filipino Mind – Philippine Philosophical Studies II*. The Council for Research in Values and Philosophy, 1994.

Mercado, Leonardo. *Elements of Filipino Philosophy*. Divine Word Publications, 1974.

Miranda, Dionisio. *Loob: The Filipino Within*. Divine Word Publication, 1989.

Moore, Charles. *The Status of the Individual in the East and West*. University of Hawaii Press, 1983.

Munoz-Mendoza, Rosita. "The Kapampangan." *NCAA Website*, March 11, 2022. https://ncca.gov.ph/about-ncca-3/subcommissions/subcommission-on-cultural-communities-and-traditional-arts-sccta/northern-cultural-communities/the-kapampangan/.

Ordonez, Marcelo. *People Power: A Demonstration of Emerging Filipino Identity*. Sampaguita Printing Press, 1986.

Quito, Emerita. *Three Women Philosophers*. Diamond Jubilee Publications, 1986.

Ramos, Christine Carmela. *Introduction to Philosophy*. Rex Bookstore Inc., 2019.

Smith, Margaret. *The Way of the Mystics*. Oxford University Press, 1978.

Timbreza, Florentino. *Mga Hugis ng Kaisipang Pilipino*. Tomas Press, 1989.

Van Over, Richard. *Eastern Mysticism*. New American Library, 1977.

M. Sartzetaki et al. • *Conatus* 10, no. 1 (2025): 221-238
doi: https://doi.org/10.12681/cjp.38439

The Ecosystem of Ethical Decision Making: Key Drivers for Shaping the Corporate Ethical Character

Maria Sartzetaki,[1] Antonia Moutzouri,[2] Aristi Karagkouni,[3] and Dimitrios Dimitriou[4]

[1]*Democritus University of Thrace, Greece*
E-mail address: msartze@econ.duth.gr
ORCID ID: https://orcid.org/0000-0002-9564-5332

[2]*National and Kapodistrian University of Athens, Greece*
E-mail address: amoutzouri@philosophy.uoa.gr
ORCID ID: https://orcid.org/0000-0001-7078-3037

[3]*Democritus University of Thrace, Greece*
E-mail address: arkaragk@econ.duth.gr
ORCID ID: https://orcid.org/0000-0002-5744-5110

[4]*Democritus University of Thrace, Greece*
E-mail address: ddimitri@econ.duth.gr
ORCID ID: https://orcid.org/0000-0003-4596-8069

Abstract

In today's global economy, large corporations possess considerable power and exert a profound impact on the societies and the environments in which they operate. The effectiveness of Corporate Social Responsibility initiatives can widely vary, depending on how well they align moral principles and social expectations, thus enhancing their impact on societal welfare and environmental sustainability. This paper examines the practical application of ethical theories in the corporate world, moving beyond normative prescriptions to what defines a corporation's ethical character. The transition to applied ethics represents a shift towards actionable guidance and contextual relevance in a way that takes into account all the involved forces and the practical implications for various stakeholders. A conceptual framework is developed to depict the relationships and decision-making influences within a socio-economic system and outline an ethical zone where care and justice ethics converge, indicating that decisions are just and do not neglect the welfare of individual stakeholders, society and the environment.

Key-words: *corporate social responsibility; corporate ethics; ethics of care; well-being; ethical decision-making; virtue; social welfare; justice*

I. Introduction

Within the complex structure of today's global economy, large corporations possess substantial power and exercise significant influence,[1] operating across a landscape where Corporate Social Responsibility (CSR) and Environmental, Social, and Governance (ESG) initiatives are increasingly becoming indispensable. These companies are now more intertwined with the different communities and the diverse environments in which they conduct business, underscoring their growing importance in shaping a sustainable and socially responsible future.

Ethical corporate behavior signifies sustainable and responsible business practices that go beyond legal compliance and prioritize more than just profit-making. CSR has become a strategically important tool and an integral part of corporate operations, leading to the development of numerous models, frameworks, guidelines, and indicators.[2] However, it is still a concept whose meaning is highly debated, open to different interpretations, or lacking strategic alignment.[3]

At the heart of ethical business operations, there are two fundamental forces, namely, societal expectations and corporate imperatives. Societal expectations are embedded in the needs, beliefs, and values of local and broader communities[4] where companies operate in, and which call for business practices that not only carry a responsibility label but also deliver tangible and meaningful impacts. In this regard, CSR practices are not meaningful unless they resonate with the daily concerns and aspirations of the people, positively affecting their quality of life. In contrast, corporate imperatives typically focus on the strategic goals, operational needs, and ethical standards that corporations set for themselves. This dichotomy highlights a divergence between what companies claim to be ethical and what is perceived as such by various stakeholders. It is clearly demonstrated in cases where multinational

---

[1] Brian Roach, *Corporate Power in a Global Economy, An ECI Teaching Module on Social and Environmental Issues, Economics in Context Initiative* (Global Development Policy Center, Boston, University 2023), 10, https://www.bu.edu/eci/files/2023/09/Corporate-Power-Module.pdf.

[2] Dimitrios J. Dimitriou, "Corporate Ethics: Philosophical Concepts Guiding Business Practices," *Conatus – Journal of Philosophy* 7, no. 1 (2022): 36-37.

[3] Kasturi V. Rangan, Lisa Chase, and Sohel Karim, "The Truth About CSR," *Harvard Business Review*, February 29, 2024, https://hbr.org/2015/01/the-truth-about-csr; Alina Dizik, "Why Corporate Social Responsibility Can Backfire," *The University of Chicago Booth School of Business*, accessed April 5, 2024, www.chicagobooth.edu/review/why-corporate-social-responsibility-can-backfire.

[4] For a seminal discussion on the conception of community as established by shared values and goals among people of common social reality, see Babalola Joseph Balogun, "How not to Understand Community: A Critical Engagement with R. Bellah," *Conatus – Journal of Philosophy* 8, no. 1 (2023): 55-76.

corporations may be driven by a quest to develop a universally applicable ethical strategy grounded in global ethical norms but fail to be relevant at a local level and address local communities' culturally rich expectations.

Overall, the expectation for corporations to go beyond legal compliance and undertake a multifaceted approach to responsible behaviour entails navigating an interplay between divergent forces. The value of ethics is indisputable in strengthening the relevance and efficacy of CSR efforts and guiding organizations to make decisions that are both morally sound and socially responsible.

II. An instrumental stakeholderism

One pragmatic approach that seeks to align the stakeholders' interests with organizational success is the concept of *instrumental stakeholderism*. Its main point is that stakeholder relationships can be a strategic tool to contribute to organizational objectives and to maximize long-term shareholder value.[5]

When involving managers with a broader scope of corporate objectives, there is a risk of arbitrariness, inefficiency, or even corruption. Managers face significant challenges in prioritizing stakeholders, partly due to the need to rely on subjective information about their preferences and contributions to the firm.[6] Since corporations are not equipped to address broad social issues, the way to lead to better overall societal outcomes is by focusing on profit maximization.[7]

In the same spirit, the famous "Friedman doctrine" claims that social responsibility for any business is to increase its profits; profit maximization and shareholder primacy are means to social welfare.[8] Similarly to Adam Smith, who is considered by many the father of modern capitalism, it is only when firms focus on their own best interests that ultimately the best interests of all stakeholders are served.[9] As corporations seek profit, they inadvertently create benefits for

---

[5] Lynn S. Paine, "Corporate Leaders Say They Are for Stakeholder Capitalism – But Which Version Exactly? A Critical Look at Four Varieties," *Harvard Business School Working Paper,* No. 24-008 (2023): 4. For a nuanced philosophical approach, see Olga Kourtoglou, Elias Vavouras, and Nikolaos Sariannidis, "The Stoic Paradigm of Ethics as a Philosophical Tool for Objectifying the Concepts of Organizational Ethics, Corporate Social Responsibility, and Corporate Governance," *Conatus – Journal of Philosophy* 9, no. 2 (2024): 119-143.

[6] Marc Fleurbaey and Grégory Ponthière, "The Stakeholder Corporation and Social Welfare," *Journal of Political Economy* 131, no. 9 (2023): 2557.

[7] Thomas L. Carson, "Friedman's Theory of Corporate Social Responsibility," *Business and Professional Ethics Journal* 2, no. 1 (1993): 15.

[8] Eva Witesman et al., "From Profit Maximization to Social Welfare Maximization: Reclaiming the Purpose of American Business Education," *Futures* 150 (2023): 103152. Also, John R. Danley, "Polestar Refined: Business Ethics and Political Economy," *Journal of Business Ethics* 10 (1991): 916.

[9] Mark Buchanan, "Wealth Happens," *Harvard Business Review*, April 2002, https://hbr.org/2002/04/wealth-happens.

other individuals and parties, due to resources being shifted to where they are most valued, thereby economic growth benefiting society at large.

Utilitarianism, a moral foundation of stakeholder theory, epitomizes *consequentialism*, an ethical framework stating that the moral status of an action is determined solely by its consequences and not by the value of the act itself. The fundamental principle of *utilitarianism* – according to its 'founding fathers,' Jeremy Bentham and John Stuart Mill – is "the greatest amount of good for the greatest number."[10] Hence, at the heart of utilitarian thought is a pragmatic evaluation system: Actions are considered moral if they result in the best possible outcomes for the largest number of people.

When closely analyzing *utilitarianism*'s relationship with CSR, one finds that it is not without issues, even though utilitarianism is innately linked to the highest social welfare. Socially responsible actions are pursued with the anticipation of contributing to the company's economic prosperity. From this point of view, CSR programs can be seen as tactical instruments to eventually increase shareholder value. Both *utilitarian ethics* and instrumental CSR prioritize the outcomes and benefits of actions, evaluating the ethicality of decisions based on their contribution to the welfare of stakeholders and society at large rather than the inherent moral quality of the actions themselves.[11]

III. Classical stakeholderism

Next to consequentialism, *deontology* is another ethical tradition that focuses on *duty*. The most prominent figure in deontological ethics is Immanuel Kant, who argued in favor of a central, absolute, and unconditional principle of morality – what he calls a *categorical imperative*. The categorical imperative applies to all rational agents independently of the circumstances in which they act. Three main formulas of the imperative are universalizability, humanity as an end in itself, and kingdom of ends, all three of which have implications in the field *of business ethics*.[12] According to the first, one should only act according to principles that could be universally applied to everyone in similar circumstances without leading to logical inconsistencies or undesirable outcomes. The second formula emphasizes the intrinsic value of all human beings, stating that individuals should never be used solely as a

[10] Julia Driver, "The History of Utilitarianism," *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), eds. Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/.

[11] Dimitrios J. Dimitriou, Maria F. Sartzetaki, and Aristi G. Karagkouni, *Managing Airport Corporate Performance: Leveraging Business Intelligence and Sustainable Transition* (Elsevier, 2024), 127-151.

[12] Norman E. Bowie, "A Kantian Approach to Business Ethics," in *A Companion to Business Ethics*, ed. Robert E. Frederick (Wiley-Blackwell, 1999), 4.

means towards an end but should be valued and respected for their own sake. Finally, according to the third formula, rational moral agents form communities whose members develop moral relationships with each other as self-legislative members of a merely possible kingdom of ends. When it comes to business ethics, such an approach requires formulating guiding principles, respecting employee autonomy and dignity, cultivating a culture of growth, and giving room for humanitarian needs.[13]

At its core, Kantian ethics prioritizes ethical obligations over utilitarian calculations. In this respect, it can be related to *classic stakeholderism*. Classic stakeholderism values the well-being of stakeholders inherently, not just when it coincides with shareholder value.[14] In fact, this type of stakeholderism could be termed categorical stakeholderism, evoking the notion of a categorical imperative as put forth by Kant.[15] Just as Kant argued that moral actions are the ones that can be universalized as a law for everyone to follow, *classic stakeholderism* advocates that businesses should act in ways that respect the inherent rights of stakeholders. This means making decisions that are ethically defensible on their own merits, not just when they align with financial objectives. Respect for autonomy and dignity is central to both Kantian ethics and the classic stakeholder approach, demanding actions that protect the moral rights and inherent value of all affected parties.[16]

IV. Shaping corporate ethical character: The role of virtue ethics

The environment in which ethical decisions are made in business is dynamic, involving continuous interaction among various stakeholders, each with different levels of interest and influence. Large corporations, especially the ones operating across multiple geographies, tend to align their practices at a global level to maintain a consistent corporate identity while also adhering to international standards. In this respect, corporate governance follows a top-down approach to decision-making, with policies related to CSR and sustainability being developed at the top levels of the hierarchy and then disseminated throughout the organization.

However, an expanding body of research argues for the necessity to embrace a more comprehensive viewpoint, considering a broader framework and

---

[13] Jacquie L'Etang, "A Kantian Approach to Codes of Ethics," *Journal of Business Ethics* 11 (1992): 739-740.

[14] Paine, 15.

[15] Ibid.

[16] For a discussion on ethically unjustifiable practices such as exploitations, see Fausto Corvino, "Sweatshops, Harm and Exploitation: A Proposal to Operationalise the Model of Structural Injustice," *Conatus – Journal of Philosophy* 5, no. 2 (2020): 9-23.

striving for socially relevant outcomes.[17] Such an approach would typically follow a bottom-up approach whereby needs and priorities are determined locally, at the point where business activities take place. This approach genuinely values the insights of individuals and groups who are directly involved or affected by business operations. Examples of such an approach include the participation of employees in the decision-making processes and the inclusion of their perspectives in the shaping of company policies and practices. Furthermore, the role of the local community is strengthened by actions such as the support of community development projects, social initiatives, or engagement in partnerships that address local issues.

It can be argued that this approach emphasizes the cultivation of virtues such as honesty, fairness, and responsibility as ends in themselves. In this respect, virtues are not subordinate to consequences (as in *utilitarianism*) or duties (as in *deontology*) but are core elements of moral evaluation and decision-making. Complementary to *utilitarianism* and *deontology*, virtue ethics represents the third major approach in *normative ethical theory*.[18] While *consequentialists* view virtues as traits leading to positive outcomes and *deontologists* as qualities of duty-fulfilling individuals, *virtue ethicists* consider virtues and vices as central to the ethical framework.[19] Virtue ethics has a long history, dating back to Aristotle, who first formalized it as a theory.[20] According to virtue ethics, the right action is what a virtuous person would do in a given situation. One can achieve virtue by using reason to identify and implement a "golden mean," which is the desirable balance between two extremes, i.e., excess and deficiency. Aristotle also argues that virtues are character traits that lead to happiness, or eudemonia.[21] The philosophical insights of Aristotle regarding happiness, virtue, and the soul, can be applied to the corporate context by advocating for virtues such as honesty, fairness, and integrity in business practices. The soul, according to Aristotle, is the driving force behind life activities and the preservation of a living body.[22] What motivates

[17] Mihaela Constantinescu and Muel Kaptein, "Virtue and Virtuousness in Organizations: Guidelines for Ascribing Individual and Organizational Moral Responsibility," *Business Ethics, the Environment & Responsibility* 30, no. 4 (2021): 801–817.

[18] Rosalind Hursthouse and Glen Pettigrove, "Virtue Ethics," *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), eds. Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

[19] Ibid.

[20] Peter Simpson, "Contemporary Virtue Ethics and Aristotle," *The Review of Metaphysics* 45, no. 3 (1992): 503-524.

[21] Evangelos D. Protopapadakis, *Creating Unique Copies: Human Reproductive Cloning, Uniqueness, and Dignity* (Logos Verlag, 2023).

[22] See Pia Valenzuela, "Fredrickson on Flourishing through Positive Emotions and Aristotle's Eudaimonia," *Conatus – Journal of Philosophy* 7, no. 2 (2022): 37-61.

the actions and choices of an organization in a business context is what is commonly known as the company culture.[23] Like Aristotle's soul, an ethically grounded corporate culture directs the organization toward pursuits that are not only profitable but also beneficial to society.

Both beneficial and structural stakeholderism share similarities with virtue ethics. Beneficial stakeholderism explicitly aims to advance stakeholders' welfare,[24] suggesting an ethical commitment that aligns with virtue ethics' emphasis on intrinsic motivations. Structural Stakeholderism involves giving stakeholders more power in governance, integrating their perspectives and values into the corporate decision-making process,[25] a process which ensures that companies are motivated to act in inherently virtuous ways. This effectively shapes the company's character, in alignment with virtue ethics.

In summary, stakeholder-centered approaches enhance the application of ethics in the business environment as they move beyond the traditional shareholder-centric model and promote a holistic view of value creation that includes the well-being of all stakeholders.

V. Ethics of care: Emphasizing well-being through relationships

However, without systematically and thoroughly addressing the fundamental imbalances that exist between organizations and employees, programs under the umbrella of CSR only achieve to treat surface-level symptoms and, at times, even prioritize the corporate brand image over employee well-being. The *ethics of care* presents a compelling perspective, beyond *utilitarianism*, Kantian and Aristotelian *virtue ethics*. It integrates traditional moral concepts like justice and utility with a focus on care, adapting these ideas to emphasize mutual growth and relationships.[26] Elements that are occasionally overlooked in other frameworks, especially when it comes to promoting well-being through relational dynamics, are highly valued in this approach. This ethical theory has been applied across organizational studies with a focus on the importance of relationships for human development and ethical business practices.[27] A relational belief system highlights that true growth

---

[23] Purissima Emelda Egbekpalu, "Aristotelian Concept of Happiness (Eudaimonia) and Its Conative Role in Human Existence: A Critical Evaluation," *Conatus – Journal of Philosophy* 6, no. 2 (2021): 79.

[24] Paine, 27-28.

[25] Ibid., 40.

[26] Jing Xu and Hedley Smyth, "The Ethics of Care and Wellbeing in Project Business: From Instrumentality to Relationality," *International Journal of Project Management* 41, no. 1 (2023) 102431: 2.

[27] Jessica Nicholson and Elizabeth C. Kurucz, "Relational Leadership for Sustainability: Building an Ethical Framework From the Moral Theory of 'Ethics of Care,'" *Journal of Business Ethics*

and effective business outcomes are best achieved through connectedness, with moral actions being understood in the context of the human capital network and interpersonal bonds. From evaluating the consequences of actions, to the motivations behind them, to the character of the actor, and finally, to the web of relationships each individual is part of, there is evident the need for embracing a more systemic approach, acknowledging that ethical issues are embedded in broader social, technological, economic and environmental contexts.[28]

Modern organizational structures need increased flexibility to adapt to market changes and to efficiently bring forward innovations, thereby arranging human resources around particular projects or development plans. The need for stakeholder interaction and connections is being recognized more and more in project-based organizations.[29] Transportation, energy, tourism, telecommunications, and construction firms, as well as the supply chain in general, are examples of contexts where relationality and interdependence between suppliers and contractors is fundamental in the success of the undertaking projects. Moreover, empathy, mutual respect and nurturing of long-term relationships are crucial, fostering a holistic view of management that integrates concern for the well-being of all stakeholders.

VI. The Industry 5.0 era

The need for project-based structures to adopt more agile and adaptive methodologies becomes even more prominent in the *Industry 5.0 era*. The landscape of Industry 5.0 is being shaped by innovations in technology that seek to balance efficiency with environmental sustainability and a human-centric design. Technologies like Cognitive Cyber-Physical Systems and Cognitive Artificial Intelligence redefine the limits of machine autonomy and intelligence. With immersive and natural experiences, Human Interaction and Recognition Technologies and Extended Reality are enhancing the integration of human skills with computer systems. Industrial Smart Wearables and Intelligent Robots are pushing the boundaries of human augmentation and collaborative robotics.[30] Unlike its predecessors, *Indus-*

156, no. 1 (2017): 25-43.

[28] For an illuminating discussion on the extent to which ethical concerns are intertwined with contemporary market practices – particularly in relation to vulnerable target groups such as children – see Andrie G. Panayiotou and Evangelos D. Protopapadakis, "Ethical Issues Concerning the Use of Commercially Available Wearables in Children: Informed Consent, Living in the Spotlight, and the Right to an Open Future," *Jahr – European Journal of Bioethics* 13, no. 1 (2022): 9-22.

[29] Xu and Smyth, 1-2.

[30] Morteza Ghobakhloo et al., "Behind the Definition of Industry 5.0: A Systematic Review of Technologies, Principles, Components, and Values," *Journal of Industrial and Production*

try 5.0 acknowledges environmental, social, and fundamental rights as crucial determinants for the future industry.[31] Its holistic perspective requires a systemic approach of ethics in business, where the interdependence of economic, social, and environmental factors is recognized and addressed in decision-making processes.

Industry 5.0 emphasizes the importance of human well-being and ethical considerations in the use of technology. It "provides a vison of industry that aims beyond efficiency and productivity as the sole goals and reinforces the role and the contribution of industry to society."[32] A systemic ethical perspective helps consider the long-term impact of business operations and enhances quality of life and work, promoting sustainable practices that go beyond immediate financial gains. *Industry 5.0* encourages collaboration between humans and machines, as well as between different stakeholders, to innovate and create value.[33] As businesses incorporate advanced technologies like AI, big data, and robotics, ethical considerations become paramount to prevent misuse. A systemic approach guides the responsible development and application of technology while ensuring that the broader interests of society are served. Finally, such an ethical framework helps businesses navigate ethical dilemmas and make decisions that are robust and flexible in the face of uncertainty and change, making them more adaptable and resilient.

In agreement with these perspectives, researchers and business practitioners have been working towards structured frameworks that encourage an organization to holistically examine its ethical considerations. One such approach is illustrated in the Business Ethics Canvas, adapted from Alex Osterwalder's Business Model Canvas.[34] Firstly, the canvas requires businesses to identify and consider the interests and impacts of a wide range of stakeholders. It also encourages businesses to consider all aspects of their operations and how these interact with their ethical commitments. Unlike approaches that treat ethics as a separate or ancillary concern, the Business Ethics Canvas embeds ethical considerations directly into the strategic planning process.

Engineering 40, no. 6 (2023): 432-447.

[31] Ganesh Narkhede, Satish Chinchanikar, Rupesh Narkhede, and Tansen Chaudhari, "Role of Industry 5.0 for Driving Sustainability in the Manufacturing Sector: An Emerging Research Agenda," *Journal of Strategy and Management*, ahead-of-print (2024). Also, "Industry 5.0: Towards More Sustainable, Resilient and Human-Centric Industry," *European Commission*, Research and Innovation, January 7, 2021, accessed April 14, 2024, https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-01-07_en.

[32] "Industry 5.0 – A Transformative Vision for Europe," *Interreg Europe*, accessed April 14, 2024, https://www.interregeurope.eu/policy-learning-platform/news/industry-50-a-transformative-vision-for-europe.

[33] Ibid.

[34] "The Ethics Canvas," accessed April 15, 2024, https://www.ethicscanvas.org/.

Business Ethics Canvas has also been used in conjunction with Business Analytics Methodology to promote the inseparability of value and ethics in everyday practices. The canvas has been suggested to be designed around Markkula Center's five ethical principles: utility, rights, justice, common good, and virtue, with an added focus on stakeholder consideration.[35] Business Analytics can facilitate ethical decision-making in many ways. First, analytics can reveal hidden patterns, thus allowing decisions based on transparent and objective data. It can also assist in a more comprehensive stakeholder analysis whereby businesses can better understand and prioritize stakeholder needs. Furthermore, incorporating ethical considerations into analytics models, like strategic planning and management system frameworks which employ financial, customer or internal process metrics, can better measure and evaluate performance against ethical objectives and not just financial outcomes. Finally, analytics can also identify and quantify ethical risks, allowing corporations to devise strategies that minimize potential harm or legal implications.

VII. Ethics as a systemic imperative in business

Building on the foregoing discussion, ethics is an essential, integral part of the entire system rather than a peripheral or standalone consideration. This action framework for CSR aligns with an Aristotelian approach. Aristotle's philosophy, particularly his notion of *"phronesis"* or practical wisdom, emphasizes the importance of finding a balance or the "golden mean" between extremes,[36] aiming for virtuous conduct that benefits the individual and the community.[37] When combined with business ethics, particularly ethics of care, the analysis becomes oriented toward achieving the best outcomes for all stakeholders, reflecting the Aristotelian aim of promoting the common good. In practice, this integrated approach would mean that companies use intelligence tools to gather and analyze data on their operations, market conditions, and social and environmental impacts, while the ethical framework guides the interpretation and application of these data to make decisions.

---

[35] Richard Vidgen, Giles Hindle, and Ian Randolph, "Exploring the Ethical Implications of Business Analytics With a Business Ethics Canvas," *European Journal of Operational Research* 281, no. 3 (2020): 491-501.

[36] Richard Kraut, "Aristotle's Ethics," *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), eds. Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics.

[37] In this regard, the notions of effective altruism and affirmative action are often invoked. For a seminal discussion on effective altruism, see Iraklis Ioannidis, "Shackling the Poor, or Effective Altruism: A Critique of the Philosophical Foundation of Effective Altruism," *Conatus – Journal of Philosophy* 5, no. 2 (2020): 25-46, and Sooraj Kumar Maurya, "A Reply to Louis P. Pojman's Article 'The Case Against Affirmative Action,'" *Conatus – Journal of Philosophy* 5, no. 2 (2020): 87-113, respectively.

This Aristotelian systemic approach, rooted in practical wisdom, would encourage businesses to act in ways that are just and caring, taking into account the well-being of all stakeholders and leading to sustainable and ethical business practices. Thus, the analytical capabilities of business intelligence tools, combined with the guidance of business ethics, can indeed be seen as a modern application of the Aristotelian systemic approach to achieving virtuous and effective CSR practices. Business intelligence and business ethics can serve as a robust framework for CSR design and implementation. Business Intelligence, with its emphasis on utilizing data and analytics to make strategic choices, can help organizations identify and prioritize CSR projects that are closely connected with their strategic goals and stakeholder interests.[38] Business ethics introduces a moral dimension to this procedure, guaranteeing that the choices taken are not only financially advantageous but also just and fair in a wider societal framework.

VIII. Integrating justice and care in business ethics

Ethical considerations in business can be viewed as residing on a spectrum; on the one end justice-oriented practices focus on the adherence to regulatory standards and aim to ensure fairness while on the other end, care-oriented practices emphasize the importance of social values, nurturing relationships and fostering a positive organizational culture. This framework provides a comprehensive understanding of how businesses can approach ethical decision-making and responsibility in various dimensions of their operations.

Effective justice-oriented ethics emphasize compliance with environmental and health and safety regulations. Companies are increasingly held accountable for their environmental footprint and are required to comply with stringent regulations aimed at decreasing pollution, managing waste, or even ensuring sustainable resource utilization.[39] Strict compliance with environmental standards not only mitigates potential legal risks but may also encourage the adoption of sustainable practices. Beyond the environment, health and safety standards are essential for preventing occupational hazards while ensuring a safe working environment. Such measures also create conditions for products and services that are safe for public use. Companies that give priority to such regulations demonstrate a commitment to ethical responsibility and effective risk management.

At the other end of the spectrum, there is care-oriented ethics, in which a more comprehensive view encompasses a wide range of activities essential for

[38] Dimitriou, Sartzetaki, and Karagkouni, 127-151.

[39] Aristi Karagkouni and Dimitrios Dimitriou, "Sustainability Performance Appraisal for Airports Serving Tourist Islands," *Sustainability* 14, no. 20 (2022): 13363.

sustaining and promoting well-being.[40] This end aligns with Peter Singer's perspective on the responsibility of not only individuals but also organizations to contribute to the broader societal well-being.[41] In the context of integrating long-term social values into business practices, just as individuals are urged to make choices that benefit society, businesses can also incorporate enduring social principles into their operations. The concept of care aligns with CSR frameworks and practices. The value of care, viewed as part of the organization's identity and culture, points towards morally appropriate relationships based on trust and interdependence.[42] Companies at this end of the spectrum increasingly recognize the personal and professional development needs of employees while also acknowledging the need to foster well-being and prioritize the holistic development of individuals.

There can be an ideal zone in the spectrum of ethical decision-making where formal aspects of ethics and social values converge. In this zone, organizations conform with regulations and at the same time embrace a deep consideration for social values such as community engagement and employee well-being. The optimum ethical zone encompasses a harmonious equilibrium between justice and care principles. By integrating both ends of the ethical spectrum, corporations can establish a comprehensive ethical framework that supports sustainable success. Organizational justice can serve as a critical basis, sustaining the organization's credibility and standards while also reinforcing ethical practices and supporting CSR initiatives.

IX. The ecosystem of ethical decision-making

Justice principles are often prioritized by stakeholders like governmental organizations, associations, and NGOs. These entities typically focus on broader societal impacts, urging for fairness, rights, and equality across communities and often at a policy level. They typically focus on establishing and enforcing guidelines that maintain the standards of justice for the broader society.

Values of care are primarily associated with stakeholders in fields that directly engage with individuals, such as healthcare providers or social services. In these domains, the focus is on empathy, compassion, and the well-being of specific individuals or groups. However, it's also relevant in Occupational Health and Safety (OHS) contexts, where the primary concern is ensuring the physical and psychological safety of workers, emphasizing values within the actual workplace.

---

[40] Berenice Fisher and Joan Tronto, "Toward a Feminist Theory of Caring," in *Circles of Care*, eds. Emily K. Abel and Margaret K. Nelson (SUNY Press, 1990), 40.

[41] Peter Singer, *The Life You Can Save: Acting Now to End World Poverty* (Random House, 2009).

[42] Denis G. Arnold and Roxanne L. Ross, "Care in Management: A Review and Justification of an Organizational Value," *Business Ethics Quarterly* 33, no. 4 (2023): 622, 628.

Within this care-justice spectrum, ethical considerations are not separate from but intrinsic to the main function of business, economic, and social entities. Considerations are balanced with primary objectives, such as financial or operational performance. A conceptual framework depicting the relationships and decision-making effects within a socio-economic system appears in Figure 1. In the top section of the diagram, including entities like Equity Funds, Retail Banks, Regional government, and multinational or smaller enterprises, decisions are primarily driven by economic criteria and financial metrics. A focus on fair and equitable market practices, promoting transparency and accountability, resonates with the perspective of justice. In the bottom section of the diagram, R&D Institutions, Industry Associations, and Development Authorities prioritize the social impact and the immediate needs of individuals and communities, aligning with a care ethics approach. The "Ethical Zone" in the center of the framework represents the area of convergence between care and justice ethics, highlighting a balanced approach between humane and formal aspects. Ethical business practices ensure that care-based decisions do not compromise economic stability and that just-based decisions do not neglect the welfare of individual stakeholders and the broader society. Governments, Social Institutions, NGOs and Financial Associations cross both domains, having a more dominant role in fostering an environment which is both just and caring.

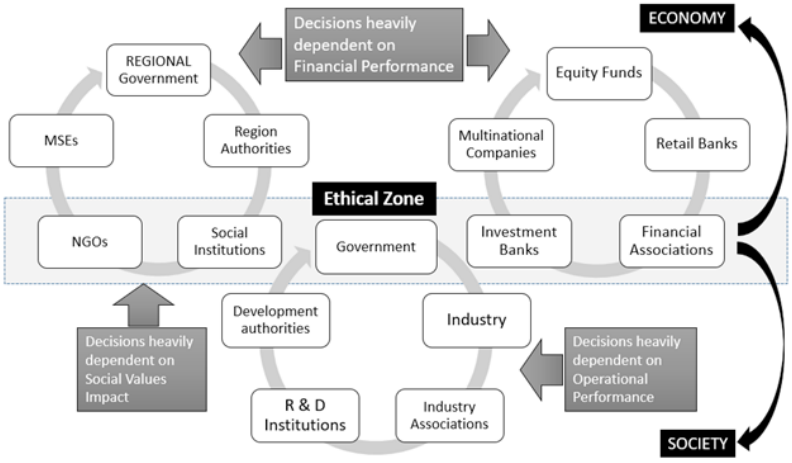


Figure 1. Conceptual Framework illustrating how different sectors balance ethical considerations with their primary objectives

Ethical business practices would ideally ensure that care-based decisions do not undermine financial sustainability and that justice-based decisions do not

neglect the welfare of employees, customers, and the broader society. The "Ethical Zone" represents the practical application of ethical principles to the decision-making processes of organizations across economic and social sectors. It is, thus, crucial that theoretical ethics meet the ground reality of business, government, and society influence. The ultimate goal is for ethics to be actively interwoven into the operational core of an organization and to manifest in both decisions and actions.

Such an all-encompassing approach fundamentally revolves around the concept of well-being, either individual, community, or of the environment. Key aspects include practices that address the physical and mental health of employees as well as their continuous development while extending to sustainable efforts and the ethical use of natural resources. Complex interdependencies increasingly emerge as businesses strive to address the ethical dimensions of pressing issues with solutions that are both caring and just and both equitable and sustainable.

Focusing on well-being in such a comprehensive manner, from the level of the individual to larger societal and environmental systems, brings forward the role of bioethics and shifts the discussion from a purely instrumental and normative approach to one that is more empowering, culturally integrated, and caring. Indeed, the normative approach is ineffective in generating a sense of caring in the implementation of well-being practices and programmes.[43] Employees may even perceive the promotion of healthy behavior in the workplace by managers as an overreach of authority, which can be demoralizing and – against expectations – counterproductive.[44] To put it another way, it might be the case that organizations manage to fulfill the instrumental compliance goals of legislation and industry norms but not fundamentally cultivate well-being.

X. Concluding remarks

Corporations need to make ethics foundational and pervasive across the spectrum of their operations, effectively addressing the complexity and interconnectedness of the various stakeholders and business ecosystem components. Corporate ethical character encapsulates the essence of the organization's moral and ethical identity. Essentially, what it stands for and how it behaves both internally and towards the external world.

Imagining ethical practices on a continuum, one end could represent care, focusing on compassion, individual well-being, and relational ethics, while the other end could represent justice, which emphasizes fairness, rights and systemic equality. This spectrum is an all-encompassing concept of Corporate

[43] Xu and Smyth, 1.

[44] Torkild Thanem, "More Passion Than the Job Requires? Monstrously Transgressive Leadership in the Promotion of Health at Work," *Leadership* 9, no. 3 (2013): 404-408.

social Responsibility. The cultivation of direct relationships and the concern for the requirements of stakeholders, including employees, customers, and related communities, are essential. Every company's approach will likely draw from both ends of this spectrum to different degrees, depending on corporate values, goals, and the expectations of their stakeholders. What is important is to avoid a type of uncritical promotion of practices just magnifying leaders' moral views and facilitating managerial aims.

Overall, the current era urges for progression from theoretical foundations through practical applications and emphasizes the integration of diverse ethical theories and stakeholder considerations into corporate decision-making.

Author contribution statement

All authors have contributed equally to the conception and design of the work, the drafting and revising of the manuscript, and the final approval of the version to be published.

References

Arnold, Denis G., and Roxanne L. Ross. "Care in Management: A Review and Justification of an Organizational Value." *Business Ethics Quarterly* 33, no. 4 (2023): 617-54.

Balogun, Babalola Joseph. "How not to Understand Community: A Critical Engagement with R. Bellah." *Conatus – Journal of Philosophy* 8, no. 1(2023): 55-76.

Bowie, Norman E. "A Kantian Approach to Business Ethics." In *A Companion to Business Ethics*, edited by Robert E. Frederick (Wiley-Blackwell, 1999), 3-17.

Buchanan, Mark. "Wealth Happens." *Harvard Business Review*, April 2002, https://hbr.org/2002/04/wealth-happens.

Carson, Thomas L. "Friedman's Theory of Corporate Social Responsibility." *Business and Professional Ethics Journal* 12, no. 1 (1993): 3-32.

Constantinescu, Mihaela, and Muel Kaptein. "Virtue and Virtuousness in Organizations: Guidelines for Ascribing Individual and Organizational Moral Responsibility." *Business Ethics, the Environment & Responsibility* 30, no. 4 (2021): 801-17.

Corvino, Fausto. "Sweatshops, Harm and Exploitation: A Proposal to Operationalise the Model of Structural Injustice." *Conatus – Journal of Philosophy* 5, no. 2 (2020): 9-23.

Danley, John. R. "Polestar Refined: Business Ethics and Political Economy." *Journal of Business Ethics* 10, no. 12 (1991): 915-933.

Dimitriou, Dimitrios J. "Corporate Ethics: Philosophical Concepts Guiding Business Practices." *Conatus – Journal of Philosophy* 7, no. 1 (2022): 33-60.

Dimitriou, Dimitrios, Maria Sartzetaki, and Aristi G. Karagkouni. *Managing Airport Corporate Performance: Leveraging Business Intelligence and Sustainable Transition*. Elsevier, 2024.

Dizik, Alina. "Why Corporate Social Responsibility Can Backfire." *The University of Chicago Booth School of Business*. Accessed April 5, 2024. www.chicagobooth.edu/review/why-corporate-social-responsibility-can-backfire.

Driver, Julia. "The History of Utilitarianism." *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), edited by Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/.

Egbekpalu, Purissima Emelda. "Aristotelian Concept of Happiness (Eudaimonia) and Its Conative Role in Human Existence: A Critical Evaluation." *Conatus – Journal of Philosophy* 6, no. 2 (2021): 75-86.

Fisher, Berenice, and Joan Tronto. "Toward a Feminist Theory of Caring." In *Circles of Care*, edited by Emily K. Abel and Margaret K. Nelson, 35-62. SUNY Press, 1990.

Fleurbaey, Marc, and Grégory Ponthière. "The Stakeholder Corporation and Social Welfare." *Journal of Political Economy* 131, no. 9 (2023): 2556-2594.

Ghobakhloo, Morteza, Mohammad Iranmanesh, Ming-Lang Tseng, Andrius Grybauskas, Alessandro Stefanini, and Azlan Amran. "Behind the Definition of Industry 5.0: A Systematic Review of Technologies, Principles, Components, and Values." *Journal of Industrial and Production Engineering* 40, no. 6 (2023): 432-47.

Hursthouse, Rosalind, and Glen Pettigrove. "Virtue Ethics." *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), edited by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

Interreg Europe. "Industry 5.0 – A Transformative Vision for Europe." Accessed April 14, 2024. https://www.interregeurope.eu/policy-learning-platform/news/industry-50-a-transformative-vision-for-europe.

Ioannidis, Iraklis. "Shackling the Poor, or Effective Altruism: A Critique of the Philosophical Foundation of Effective Altruism." *Conatus – Journal of Philosophy* 5, no. 2 (2020): 25-46.

Karagkouni, Aristi, and Dimitrios Dimitriou. "Sustainability Performance Appraisal for Airports Serving Tourist Islands." *Sustainability* 14, no. 20 (2022): 13363.

Kourtoglou, Olga, Elias Vavouras, and Nikolaos Sariannidis. "The Stoic Paradigm of Ethics as a Philosophical Tool for Objectifying the Concepts of Organizational Ethics, Corporate Social Responsibility, and Corporate Governance." *Conatus – Journal of Philosophy* 9, no. 2 (2024): 119-143.

Kraut, Richard. "Aristotle's Ethics." *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), edited by Edward N. Zalta and Uri Nodelman, https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

L'Etang, Jacquie. "A Kantian Approach to Codes of Ethics." *Journal of Business Ethics* 11 (1992): 737-744.

Maurya, Sooraj Kumar. "A Reply to Louis P. Pojman's Article 'The Case Against Affirmative Action.'" *Conatus – Journal of Philosophy* 5, no. 2 (2020): 87-113.

Narkhede, Ganesh, Satish Chinchanikar, Rupesh Narkhede, and Tansen Chaudhari. "Role of Industry 5.0 for Driving Sustainability in the Manufacturing Sector: An Emerging Research Agenda." *Journal of Strategy and Management*, ahead-of-print.

Nicholson, Jessica, and Elizabeth C. Kurucz. "Relational Leadership for Sustainability: Building an Ethical Framework from the Moral Theory of 'Ethics of Care.'" *Journal of Business Ethics* 156, no. 1 (2017): 25-43.

Paine, Lynn S. "Corporate Leaders Say They Are for Stakeholder Capitalism – But Which Version Exactly? A Critical Look at Four Varieties." *Harvard Business School Working Paper, No. 24-008* (2023): 1-60.

Panayiotou, Andrie G., and Evangelos D. Protopapadakis. "Ethical Issues Concerning the Use of Commercially Available Wearables in Children: Informed Consent, Living in the Spotlight, and the Right to an Open Future." *Jahr – European Journal of Bioethics* 13, no. 1 (2022): 9-22.

Protopapadakis, Evangelos D. *Creating Unique Copies: Human Reproductive Cloning, Uniqueness, and Dignity*. Logos Verlag, 2023.

Rangan, V. Kasturi, Lisa Chase, and Sohel Karim. "The Truth About CSR." *Harvard Business Review*, February 29, 2024. doi: https://hbr.org/2015/01/the-truth-about-csr.

Research and Innovation. "Industry 5.0: Towards More Sustainable, Resilient and Human-Centric Industry." January 7, 2021. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-01-07_en.

Roach, Brian. *Corporate Power in a Global Economy. An ECI Teaching Module on Social and Environmental Issues, Economics in Context Initiative*. Global Development Policy Center, Boston University, 2023. https://www.bu.edu/eci/files/2023/09/Corporate-Power-Module.pdf.

Simpson, Peter. "Contemporary Virtue Ethics and Aristotle." *The Review of Metaphysics* 45, no. 3 (1992): 503-524.

Singer, Peter. *The Life You Can Save: Acting Now to End World Poverty*. New York: Random House, 2009.

Thanem, Torkild. "More Passion than the Job Requires? Monstrously Transgressive Leadership in the Promotion of Health at Work." *Leadership* 9, no. 3 (2013): 404-408.

Valenzuela, Pia. "Fredrickson on Flourishing through Positive Emotions and Aristotle's Eudaimonia." *Conatus – Journal of Philosophy* 7, no. 2 (2022): 37-61.

Vidgen, Richard, Giles Hindle, and Ian Randolph. "Exploring the Ethical Implications of Business Analytics with a Business Ethics Canvas." *European Journal of Operational Research* 281, no. 3 (2020): 491-501.

Witesman, Eva, Bradley Agle, Justin Ames, Steven Christenson, Max Moore, and Rachel Pankey. "From Profit Maximization to Social Welfare Maximization: Reclaiming the Purpose of American Business Education." *Futures* 150 (2023): 103152.

Xu, Jing, and Hedley Smyth. "The Ethics of Care and Wellbeing in Project Business: From Instrumentality to Relationality." *International Journal of Project Management* 41, no. 1 (2023) 102431: 1-6.

J. Slater · *Conatus* 10, no. 1 (2025): 239-254
doi: https://doi.org/10.12681/cjp.35215

Political Decision-making, Lottocracy, and AI

Joe Slater
University of Glasgow, United Kingdom
E-mail address: joe.slater@glasgow.ac.uk
ORCID iD: https://orcid.org/0000-0003-2269-7235

Abstract

This article examines an argument for single-issue legislatures (SILs) as an alternative to typical decision-making procedures in representative democracies. It is argued that if this argument was successful, it could be extended to endorse decision-making processes utilizing advanced artificial intelligence. However, it is noted that this argument neglects an important feature of decision-making: authorization. It is important for its autonomy that a society must authorize certain decisions. If decisions were delegated to SILs or AI, this would undermine the autonomy of the group. This does not entail that these alternatives have no role in policy-crafting. It is argued that so long as a community can authorize a decision, perhaps by a vote, it need not undermine autonomy. An important caveat of this, however, is that the decision must be explicable to the community. For AI usage, this motivates the need for explainable AI (XAI).

Keywords: *lottocracy; AI; explanation and responsibility; authorization; autonomy*

I. Introduction

In representative democracies, elected officials are charged with crafting and enacting policies. Myriad problems have been long-documented for this system of policy-making. For example, given that the electorate is often uninformed about the issues, the politicians that get elected may be the ones who are most *persuasive* rather than the most competent. One suggestion, courtesy of Alexander Guerrero, is instead to use single issue legislatures selected by lot, which he terms "lottocracy."[1] He argues persua-

[1] Alexander Guerrero, "Against Elections: The Lottocratic Alternative," *Philosophy & Public Affairs* 42, no. 2 (2014): 135-178. Another lottocratic system is discussed in Claudio López-Guerra's "The Enfranchisement Lottery," *Politics, Philosophy & Economics*, 10, no. 2 (2010): 211-233.

sively that such a system may fare better than our current system in terms of its responsiveness to public opinion and good governance. However, if we explicitly express these as the values we want to promote, this may open the possibility of using sophisticated AI to develop policy instead, and this could be even more successful according to those metrics.

In this piece, I first consider Guerrero's argument for lottocracy, and show how (if we accept his assumptions), this argument could generalize and be used to defend policy-decisions being made by AI. I also note how similar defenses might be made for other existing (and proposed) uses of AI. Second, I propose that both of these systems fail to manifest something that is important about important decisions, namely that we *authorize* them. I point to several reasons why this authorization is valuable, drawing upon research in autonomy and responsibility. Finally, I investigate the implications of taking the value of authorization seriously. Crucially, I claim that we cannot authorize a decision unless we are sufficiently informed about it. This highlights the need for XAI in any domain where authorization is important.

II. Guerrero on lottocracy

Complaints about democracy are not new. In *The Republic* Plato argues for rule by "Philosopher Kings,"[2] viewing democracies as susceptible to the rule by the popular, rather than rule by the wise. In contemporary political philosophy, similar arguments are used by those endorsing "epistocracy" – the notion that experts should make the decisions.[3] Alexander Guerrero has also argued that representative democracies have a variety of problems. Notably, he points to the possibility for politicians to be controlled by powerful interested parties. Because individual people cannot practically be fully informed on the range of complicated issues, people with power and resources can influence the political process in their favor, leading to policies that benefit them at the expense of others. This phenomenon – which he calls *capture* – undermines democracy's core principles of equality and fairness, as it allows a particular group to wield excessive political power and influence decision-making. It also makes it less likely that policies that

López-Guerra suggests that a system of universal suffrage is superior to the version he discusses. An extreme version of this, where a lottery decides one person who will make political decisions, is discussed in Isaac Asimov's short story, "Franchise" (1955). See Isaac Asimov, "Franchise," in *If: Worlds of Science Fiction*, ed. James L. Quinn, 2-15 (Quinn Publishing, 1955).

[2] Plato, *Republic,* trans. Robin Waterfield (Oxford University Press, 1993), V:473c.

[3] Jason Brennan, *Against Democracy* (Oxford University Press, 2016).

benefit the electorate as a whole get enacted, as captured politicians will promote policy which favors those interest groups.

As an alternative, Guerrero proposes that legislation be determined by groups selected by lot – *lottocracy*. In Guerrero's suggestion, people would be randomly selected to a legislature about a particular domain of government, e.g., agriculture, tax policy or defense spending. These single-issue legislatures (SILs) would have about 300 people, demographically reflecting society as a whole, with legislators serving staggered three-year terms (each year, one hundred being replaced). Guerrero notes that this system would seriously reduce the risk of capture. In a representative democracy, paying a powerful politician (or someone who is expected to become a powerful politician) can be a good investment, as they could have a significant influence on decisions for a long time. Members of a SIL only have a small influence over one policy area, and for a mere three years, so if they could be influenced, this would be a less valuable investment for wealthy would-be manipulators.

The members of the SIL, Guerrero imagines, would meet several times a year, and be able to call experts to give advice. They would be well-compensated for their time, and encouraged to take their responsibility very seriously. This, Guerrero hopes, would enable discussions that would result in legislation that reflect the views of the general public, and would effectively pursue their policy goals, i.e., the system would be *responsive*. He also supposes that, freed from the interests of powerful parties, the resulting legislation would be better, however we might think that this could be understood from an objective vantage point, (e.g., may lead to fairer policies, or higher levels of average welfare), i.e., the system would result in *good governance*. In these two respects – responsiveness and good governance – Guerrero argues that lottocracy may perform better than representative democracy.

Of course, much can be said about whether Guerrero is correct about this claim. For instance, Lachlan Umbers argues that it is at best inconclusive whether these instrumental benefits would actually obtain. For instance, while members of the SIL might not be a suitable investment for capture, the experts who would be likely to provide advice would be, so wealthy people may provide extravagant support to fund favorable research. Furthermore, wealthy elites will retain lobbying power; they may be able to distort the information that SIL members receive, and SIL members may be more susceptible to such influences due to their relative inexperience in positions of political decision-making.[4]

[4] Lachlan Montgomery Umbers, "Against Lottocracy," *European Journal of Political Theory* 20, no. 2 (2021): 312-334. Umbers also argues that lottocracies establish "objectionable social

For my purposes, I will assume that Guerrero is correct, i.e., that his version of lottocracy could yield more responsive policy decisions and improvements in improved governance. Even if this is the case, however, we might feel that there would be something important absent if we were to switch from representative democracy to lottocracy in our decision-making. Even though the laws arrived at might do a good job in terms of *reflecting* the values and goals of the population, the role of the general public is severely diminished. Christina Lafont touches upon this concern when she points out that lottocracy seems to endorse a "rule of the minority,"[5] even if this minority is constituted by a representative sample and can accurately express public preferences. In the next section, I'll suggest one element that I think is lacking in such a system, namely, the *public authorization* of the policies.

Before that, consider the following thought experiment. A team of psychologists, political theorists, computer scientists and philosophers team together to create an AI system to design legislation (either for a particular policy or set of policies). This AI – call it 'Landru'[6] – draws upon vast quantities of data from the whole population, perhaps via monitoring social media or surveys to representative samples (put aside for now questions about how to ethically collect such data or the logistics of doing so). Once Landru is operational, the policies it proposes are generally agreed to be far superior to the policies that our actual politicians arrive at. Perhaps this is because it is able to include in its calculations far more factors. After a while, politicians start to simply utilise policies proposed by Landru, rather than developing their own. Sometime later, the politicians decide (or perhaps Landru suggests!) getting rid of the politicians altogether, possibly because they have become glorified (and expensive) intermediaries. Why would anyone agree to this? I think there could be multiple motivations.[7] Imagine a scenario where Landru's recommended policy decisions have been taken for a long time. Eventually, a politician – to the surprise and outrage of the citizens – makes a decision contrary to the guidance, and this leads to some undesirable consequences. To

---

and political inequalities."

[5] Cristina Lafont, "A Militant Defence of Democracy: A Few Replies to my Critics," *Philosophy and Social Criticism* 47, no. 1 (2021): 70.

[6] This name is taken from the *Star Trek* episode "The Return of the Archons" (1967), where a society has allowed all policy to be determined by an advanced computer (Landru). It was also revisited in the *Star Trek: Lower Decks* episode, "No Small Parts" (2020).

[7] I don't think the details of *how* a society would arrive at such an arrangement are so important – what I'm really concerned with is whether the society that results would be deficient in some way.

avoid this, it might be demanded that faith be placed in Landru, which might be viewed as more trustworthy, or as a better judge of how to weigh the interests of competing groups equally. So a referendum is held on the future of policy-making, and a decision is reached to automatically defer to Landru. After this point, Landru makes all policy decisions. It continues to draw upon the wishes and values of the people in directing public policy.

If we were to assume that it was possible to develop an AI like Landru – I don't want to make grand claims about this here, but it at least seems like we can imagine developing an AI that would serve similar functions – how would this fare in terms of responsiveness and good governance?[8] In terms of responsiveness, it seems like this system could be more effective than current representative democracies. This is because, perhaps by feedback mechanisms like regular surveys, or monitoring social media, it could detect changes in popular opinion and make policy changes at any point during the political cycle (not just when elections occur), could register viewpoints from every member of the population (and give them equal weight) and would be able to instigate policy changes immediately. It could also prove superior to the SILs in a lottocracy, because the members of the public summoned to service in the SIL would need quite a lot of time and effort to develop sufficient expertise about the policy areas. In terms of good governance, it also seems that Landru could perform very well here, however we think good governance should properly be understood. Once developed, it also looks like Landru would be a much cheaper option (and require less labor).

Now we are in a position to consider a generalization argument. Guerrero notes the importance of responsiveness and good governance. If (1) these features should determine our system of legislating,[9] and (2) we developed a Landru-like AI (one that can meet the above conditions), then we ought to use it.

In what follows, I will assume that it is possible for us to develop AI like this. I take this strategy because I want to demonstrate the

[8] In this thought experiment, I simply stipulate that programming Landru to make ethical decisions (e.g., decisions that do not violate rights, accompanied by no desires to dominate humanity). For some discussion about the difficulties in creating an AI of this sort, see Michael Anderson et al., "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson," *Conatus - Journal of Philosophy* 6, no. 1 (2021): 177-202.

[9] Guerrero does not endorse this claim, which is something of an oversimplification, but one I take to be worth examining. Rather, he says that these two factors present serious problems for electoral democracy, and that he thinks lottocracy is normatively attractive because it is superior in these dimensions.

failings of the first condition. When the failure of the first condition is revealed, this cuts of the inference from the potential instrumental successes of lottocracy (or Landru-cracy) to the claim that these systems should be implemented. Furthermore, observing how this argument fails may highlight issues with other AI uses, which perhaps we should be suspicious of, for instance, the use of algorithms in sentencing, or evaluation of résumés.

III. The Importance of authorization

In both of the alternatives to representative democracy considered in the previous section, the role of the public in decision-making has been diminished. Importantly, if there is a problem with these systems, it isn't that bad policies are being selected (by stipulation), or that the legislation passed is not what the public would have selected. Each system is designed to result in legislation that the public would be in favor of, if ideal conditions obtained (e.g., that they were fully informed about the issues, had time to deliberate, and had no cognitive limitations in foreseeing consequences). There are still various different ways we might view these systems as lacking. On one hand they could be seen as deficient in terms of accountability.[10] Such a system may also have undesirable consequences with regard to the public becoming disengaged with political matters.[11] A further missing feature, I argue, is *authorization*.

To illustrate what I mean by "authorization," consider someone employed as an assistant. There may be standing duties that are always part of this job. Some of these may not be fully determined in a variety

[10] Hubertus Buchstein, "Lottocracy and Deliberative Accountability," *Philosophy and Social Criticism*, 47, no. 1 (2021): 40-44; see also Kate Crawford and Jason Schultz, "AI Systems as State Actors," *Colombia Law Review* 119, no. 7 (2019): 1941-1972. They discuss the 'accountability gap,' which currently haunts algorithmic decision-making. See Michael Townsen Hicks, James Humphries, and Joe Slater, "ChatGPT is Bullshit," *Ethics and Information Technology* 26, no. 38 (2024). Hicks et al. raise the problem of accountability for the falsehoods uttered by large language models. See also Joe Slater, James Humphries, and Michael Townsen Hicks, "ChatGPT Isn't 'Hallucinating' – It's Bullshitting," *Scientific American* 34, no. 1 (2025): 46-47.

[11] Cristina Lafont, *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy* (Oxford University Press, 2020). Because no such system exists, and the notion has not inspired considerable debate, there is not a wealth of discussion about disengagement as a result of AI use in policy decisions. However, the reasons this may occur are the same as with lottocratic systems, namely, that if individuals see themselves as having no direct role in decision-making, they may lack the motivation to remain inform on political issues. This is essentially the counterpart of an argument made by Mill for democracy, namely that an empowered citizen is "called upon ... to weigh interests not his own; to be guided, in cases of conflicting claims, by another rule than his private partialities." John Stuart Mill, *Considerations on Representative Government* (Parker, Son, and Bourn, 1861).

of respects; i.e., they permit some discretion. For example, if not specified, documents might be formatted in a variety of ways, or lunch could be purchased at different venues. But there are also likely to be a variety of tasks that an employer will request to be done, e.g., to arrange a meeting or order new furniture. If the assistant was to perform these tasks without the employer's request (or being told that they could do them), the employer could appropriately feel wronged by this, even if the employer would have made these requests later anyway.

To embellish the example further, and make it more similar to a legislative decision-making process that citizens have voted for, we could imagine that the employer has instructed their assistant that they can make decisions on a wide range of matters, including important decisions normally seen within the employer's remit. In this case, when the secretary makes the decisions, they have not wronged the employer. However, the employer might still feel, when these decisions are enacted, like something is missing from their involvement. They haven't made the decision themselves (even if they would have made the same decision).

Why should it matter for political decision-making whether the citizens *brought about* the decisions, as opposed to the decisions merely reflecting what they would opt for when fully informed? I suggest that the answer lies in autonomy. In ethical theory, autonomy is seen as having central importance by philosophers of varying different stripes, as seen in Kant,[12] Mill,[13] or Rawls.[14] Mill sees respecting autonomy as conferring certain benefits, such as promoting utility. Others, like Kant, regard autonomy as intrinsically valuable; an individual's being autonomous confers upon them a special moral status, which commands respect.[15] Regardless of why we see it as valuable, *that* we value autonomy is uncontroversial.[16]

What *is* controversial is exactly how we should understand autonomy. There are three broad types of views of autonomy: reason-based

[12] For a discussion of Kant's conception of autonomy see Janis Schaab, "Kant on Autonomy of the Will," in *The Routledge Handbook of Autonomy*, ed. Ben Colburn, 44-54 (Routledge, 2023).

[13] This is especially clear in John Stuart Mill, *On Liberty* (Broadview Press, 1859).

[14] For example, Rawls notes that his first principle of justice can be given an interpretation based on Kant's notion of autonomy. See John Rawls, *A Theory of Justice, Revised Version* (Harvard University Press, 1999), 221.

[15] For a discussion of these views see Mark Piper, "Justifying Respect for Autonomy," in *The Routledge Handbook of Autonomy*, ed. Ben Colburn, 293-302 (Routledge, 2023): 293-302.

[16] Even Sarah Conly, who argues that we place too much value on autonomy, accepts that we do value autonomy. See Sarah Conly, *Against Autonomy* (Cambridge University Press, 2014).

views, motivation-based views, and self-creation/self-authorship views.[17] Notorious among the reason-based views is the Kantian picture. This holds that the notion of self-legislation is important: with the use of reason, one can recognize the unconditional authority of the moral law. The rational agent will then make themselves act only on maxims that are consistent with the moral law, expressed by the categorical imperative.

On the motivational view, expressed by Gerald Dworkin, the structure of one's motivations was emphasized. Autonomy on such a view involves having authentic preferences that do not have some alien or external source.[18] Preferences being authentic requires that they are those of one's "true self" or "highest-order self," i.e., not only what one wants, but what one *wants* to *want*.

Self-creation views were advocated by Joseph Raz. These highlight the importance of the person *determining* how their life goes. Raz sees autonomy as "an ideal of self-creation or self-authorship."[19]

Strictly speaking, the positions mentioned in the aforementioned paragraphs are views of *personal* autonomy. However, we also talk about *group* autonomy. We may ask, for instance, if some policy violates the autonomy of an indigenous group, or whether some treaty diminishes the autonomy of a nation. As Wellman notes: "Group autonomy exists when the group as a whole, rather than the individuals within the group, stands in the privileged position of dominion over the affairs of the group."[20]

Just as we regard considerations of *personal* autonomy as providing us with reasons for or against some decision, so too do we with regard to these kinds of groups.[21] It is reasonable to wonder how well views of personal autonomy could map onto accounts of *groups*. This is particularly evident when we consider puzzles of group agency. Can a group *really* be an agent? Does it make sense to say that a group *believes* X, or *wants to do* Y? This is a rich debate, which I cannot do justice to here.[22] What *is* clear is that we at least talk *as if* groups can have

---

[17] Ben Colburn, *Autonomy and Liberalism* (Routledge, 2010).

[18] Gerald Dworkin, "Autonomy and Behavior Control," *Hastings Center Report* 6, no. 1 (1976): 23-28.

[19] Joseph Raz, *The Morality of Freedom* (Oxford University Press, 1988), 369.

[20] Christopher Heath Wellman, "The Paradox of Group Autonomy," in *Autonomy. Social Philosophy and Policy*, eds. Ellen Franken Paul, Fred D. Miller Jr, and Jeffrey Paul, 265-285 (Cambridge University Press, 2003), 273.

[21] Exactly *which* kinds of groups are such that their autonomy is a morally relevant consideration is a difficulty question.

[22] For an overview of some of the discussions in this area, see Katherine Ritchie, "The Meta-

such states and can act as groups. For instance, we may say, perfectly felicitously, that the UK voted the Tories out in the 2024 general election. We do, in everyday discourse, speak as if groups are agents that can rationally consider options, make decisions, and act accordingly.[23]

Given that we can talk about groups having the same kinds of properties as are required in the accounts of personal autonomy, then – if we think that group autonomy is simply the group equivalent of personal autonomy – it seems natural to think of group versions of the accounts of personal autonomy, i.e., that we can talk about reason-based views, motivation-based views and self-creation/self-authorship views of group autonomy. So, for my purposes, I assume that similar conditions will be applicable.

On each of these accounts of autonomy (or at least, on certain versions of these accounts), I contend that a persuasive story can be told about the value of authorization with regards to important decisions.[24] When we consider *personal* autonomy, it is important that we authorize certain decisions about how our own lives go. When we consider the autonomy of a *group* or *nation*, similarly, the group must authorize certain decisions about what affects them. In what follows, I will pay particular attention to self-authorship views, which are particularly suitable for such an explanation, and specifically examine Ben Colburn's.[25]

In Colburn's view, "autonomy requires both deciding for oneself what is valuable and also living in accordance with that decision by making decisions that make one responsible for the shape one's life takes."[26] An important part of this view is *making your life go the way you decide*. This is something that we care about, and perhaps see as part of us living a good life. To be autonomous in this way, we do need to be *responsible* for the shape of our life; we need to actually bring it about.

For individuals or groups to be autonomous, they must actually *be* the authors of their lives. The autonomy critique of lottocracy/Landru-cracy, however, holds that there is simply a problem because the citizens themselves have forfeited important controls over their lives.

---

physics of Social Groups," *Philosophy Compass* 10, no. 5 (2015): 310-321.

[23] For a discussion of how group intention is possible, see Margaret Gilbert, "Shared Intentions and Personal Intentions," *Philosophical Studies* 144, no. 1 (2009): 167-187.

[24] Colburn himself makes a similar move in discussions of assisted dying, in Ben Colburn, "Autonomy, Voluntariness and Assisted Dying," *Journal of Medical Ethics* 46, no. 5 (2020): 316.

[25] Colburn's view is described in detail in *Autonomy and Liberalism*.

[26] Colburn, "Autonomy," 316.

They are no longer masters of their lives, forging their own destinies.

It is perhaps unsurprising that these alternative systems fail to manifest conditions of autonomy. Autonomy is often characterized as self-rule or self-governance, and this is precisely what is sacrificed when decisions are entirely delegated to a randomly selected lot, or to an AI. To put it in terms of Lincoln's proverb, the governing may still be *for* the people, but in an important sense, it is no longer *by* the people.

When we attend to considerations of autonomy, this can elucidate issues that may have been obscured in various analyses of policy-making systems. This can be seen in the case of a pure lottocracy, where policy decisions were all determined by randomly selected legislators. Guerrero suggests that this could result in policies more responsive to citizens' values and objectively better at whatever we would desire government to do. But as these policy decisions are important to our lives – perhaps to our very identities – our being excluded from participation in the decision-making process constitutes a violation of our autonomy. And if autonomy is one of our fundamental values, this would be a bitter pill to swallow.

Thus far, I have argued that moving from a democracy – either direct or representative (voting for representatives, as well as for particular policy proposals can be regarded as a form of authorization) – to a system where laws are designed by a representative sample of the population or by a sophisticated AI fails to respect the value of autonomy. Finally, I reflect upon what the requirement for authorization really entails, and what implications this has for our policy-making decisions and other areas we might utilize AI.

IV. Authorization, AI and XAI

If we accept, as I have suggested we should, that it is important for our autonomy that we authorize important decisions, this leaves open several questions. For instance, we might wonder what form the authorization must take, how often it must occur, or how fine-grained acts of authorization must be. If explicit authorization for every aspect of every policy must occur, then seemingly no system but a direct democracy with compulsory voting would suffice. On the other end of the spectrum, we might imagine that our imagined society does have a vote every fifty years to decide whether or not to continue allowing Landru to determine all policy decisions. If something like this – voting to continue with the AI-system once a generation (or an alternative whereby citizens voted whether or not to continue a lottocratic system at similar intervals) – would constitute sufficient authorization then

rule by lot or AI would presumably be acceptable, so long as these kinds of "approval elections" did take place sufficiently regularly.

I do not propose to give a precise answer to such questions here, but will note a couple of important features that the authorization should have.

First, while it is possible to talk about a certain *decision* being made autonomously, autonomy is typically viewed as a *global* property, i.e., it applies over a long period of time, not merely with regards to particular decisions.[27] It does admit of degrees, but above a certain threshold we will attribute the property. So to have this property, the entity in question (be it a person or a group) must have enough decisions of enough importance. With this in mind, for public to act autonomously in deciding its policy goals, its role cannot be a one-time affair. It must persist over a course of time, e.g., with regular votes. This may also allow for systems where citizens can participate in a variety of other ways – I take no stand on this here.

Second, when we consider how fine-grained an instance of authorization must be in order, it could again be helpful to consider the value we are trying to promote. Recall the claim that autonomy requires, as Colburn puts it, "deciding for oneself what is valuable and also living in accordance with that decision by making decisions that make one responsible for the shape one's life takes." Highlighting *responsibility* here can be fruitful. When Colburn discusses this,[28] he notes two senses we can be responsible: attributability and substantive responsibility. Colburn sees both of these as necessary for the required responsibility.

For our actions to be attributable to us, "it must be recognizably our choices and actions that make our lives the way they are."[29] To be substantively responsible, we must be "liable for the consequences of the things attributable to us."[30] Exactly what this responsibility requires may be very context-dependent. In some contexts, the public indicating a general direction to go on might be all that is necessary for them to meet these conditions. In others, perhaps something more specific could be needed. What does seem important is that the public has been *informed*. To be liable for consequences of decisions, clearly one needs to understand (or have *the ability* to understand) at least something about the content of those decisions. Again, being informed might re-

[27] Ibid., 4.

[28] When discussing responsibility, Colburn draws upon a discussion from Thomas M. Scanlon's book. See Thomas M. Scanlon, *What We Owe to Each Other* (Harvard University Press, 1998), 148-251.

[29] Colburn, "Autonomy," 32.

[30] Ibid., 32.

quire different degrees of detail in different scenarios, but complete ignorance about policies selected and reasons that count in favor of them seems incompatible with this.

Now we are in a position to say something about what authorization must look like. Because autonomy is a global condition, authorization must be regular. The public must have more than merely occasional say about policy-matters. Furthermore, in order for the policies to be properly attributable to the public, they must be sufficiently *informed*.

The alternative decision-making systems can now be considered again. Lottocracy does not respect autonomy, because the public does not authorize the decisions. However, if the lottocratic method of crafting legislation is combined with suitable public participation to ensure authorization, this problem can be overcome. Members of the legislator can explain their decisions to members of the public, and then a vote can be used to ratify the decision. Interestingly, this is compatible with Buchstein's proposal,[31] and with various ways citizens' assemblies have been used recently, whereby the assemblies determine what questions in a referendum should be put to the public.[32]

With regards to legislation crafted by AI, a problem becomes apparent. One worry with machine learning algorithms is that they essentially constitute "black boxes,"[33] i.e., we may know what goes in and what comes out, but the process is left something of a mystery to us. If Landru, our policy-making AI, was of this sort, then it would not be possible for us to authorize the decisions it arrives at, because the public could not be informed.

However, just as the lottocracy case could be adjusted, so can the AI case. The public would need some say on the policies proposed. But more than that, they would need access to the *reasons* why these policies are favored over alternatives. This would require explainable AI (XAI), i.e., "systems that explain how the algorithms reach their conclusions or predictions."[34] This explanation would need to be suitable for

---

[31] Buchstein, "Lottocracy."

[32] For example, Ireland's use of a citizens' assembly, which set the question that would be put to a referendum on abortion. See "The Irish Abortion Referendum: How a Citizens' Assembly Helped to Break Years of Political Deadlock," *Electoral Reform Society*, accessed June 9, 2019, https://www.electoral-reform.org.uk/the-irish-abortion-referendum-how-a-citizens-assembly-helped-to-break-years-of-political-deadlock/.

[33] For some discussion of "black box" issues with algorithms see Avihay Dorfman and Alon Harel, "Why Not Artificial Intelligence?" in *Reclaiming the Public*, ed. Avihay Dorfman and Alon Harel (Cambridge University Press, 2024), 171-187.

[34] Ashley Deeks, "The Judicial Demand for Explainable Artificial Intelligence," *Columbia Law*

at least some human beings to understand. It need not be put in terms so simple that every member of the public can understand (at least not in principle), as some members of the public will be able to take on the task of explaining to others.

Worries about use of algorithms has led to the demand for XAI in other areas. For instance, algorithms used in sentencing criminals,[35] to decide outcomes of loan applications or to determine which students are admitted into universities.[36] Often, the concern with such uses is that the algorithms are giving weight to factors that should not be relevant, and if the algorithm is a black box, parties disadvantaged by this have no grounds for complaint. For example, companies have used algorithms to evaluate CVs of candidates, but some of these algorithms have systematically disadvantaged women, or paid undue attention to a candidate's name.[37] These issues are concerning even putting aside considerations of autonomy of the organizations using them, and has led to calls for a right to protest decisions made by AI.[38] Yet even if the algorithms did not exhibit these significant failings with respect to fairness, if organizations value their autonomy, they may see these algorithms as independently objectionable. If no individual members of the organization – be it an employer making a hiring decision, a university considering who to admit, or a court deciding on sentencing – have knowingly authorized the decision, it is dubious to what extent the decision can appropriately be attributed to the organization.

V. Conclusion

I have argued that, when we consider political decision-making, focusing on features like responsiveness and efficiency in obtaining results

Review 119, no. 7 (2019): 1829.

[35] Some of the issues with the use of algorithms in sentencing are discussed by Duncan Purves and Jeremy Davis. See Duncan Purves and Jeremy Davis, "Should Algorithms that Predict Recidivism Have Access to Race?" *American Philosophical Quarterly* 60, no. 2 (2023): 205-220.

[36] Gyorgy Denes, "A Case Study of Using AI for General Certificate of Secondary Education (GCSE) Grade Prediction in a Selective Independent School in England," *Computers and Education: Artificial Intelligence* 4 (2023): 100-129.

[37] "Companies Are on the Hook if Their Hiring Algorithms are Biased," *Quartz*, accessed August 13, 2023, https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased.

[38] Problems with algorithm use in hiring are also discussed by Pauline T. Kim and Matthew T. Bodie. See Pauline T. Kim and Matthew T. Bodie, "Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy," *ABA Journal of Labor and Employment Law* 289, no. 35 (2021): 289-315. Margot Kamiski and Jennifer Urban, "The Right to Contest AI," *Columbia Law Review* 121, no. 7 (2021): 1957-2048.

can obscure some of what we value. Specifically, this approach fails to appreciate the value we place on our decisions being *autonomous*. We want to be the authors of our own lives; we hold in high esteem our "privileged position of moral dominion"[39] over our affairs, and see alternative ways of life which deprive us of this as problematic. This is highlighted by an unease some of us feel regarding lottocracy. It can also be seen in the argument I constructed about a policy-making AI. If we accept that a Landru- like AI would create good policy decisions, which are extremely responsive to public values, we can arrive at a disjunctive conclusion. Either i) AI-legislation (at least regarding some domains of decisions) is superior than both democratic *and* lottocratic methods, or ii) the values of political decisions are not exhausted by the two main factors Guerrero considers. I argue in favor of ii), suggesting that there is value in the *authorization* of a decision. This feature is lacking in lottocratic and AI-crafted legislation. The lack of authorization inhibits the public's autonomy.

In order for a society's decisions to be made autonomously, it must authorize those decisions. This entails that the public – as a whole, not merely a small subset – has a say (e.g., elections), and that the justifications for the decisions are made clear. The easiest ways this can be done, utilising new methods of legislation-crafting, would still involve elections, perhaps with the matters put to an electorate determined not merely by politicians. Those excited by the potential for new technological developments to improve our decision-making (and, as a result, our lives) are right to be excited, but if we value our autonomy, we should be careful not to sacrifice that along the way.

References

Anderson, Michael, Susan Leigh Anderson, Alkis Gounaris, and George Kosteletos. "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 177-202.

Asimov, Isaac. "Franchise." In *If: Worlds of Science Fiction*, edited by James L. Quinn, 2-15. Quinn Publishing, 1955.

Brennan, Jason. *Against Democracy*. Oxford University Press, 2016.

Buchstein, Hubertus. "Lottocracy and Deliberative Accountability." *Philosophy and Social Criticism* 47, no. 1 (2021): 40-44.

[39] Wellman, "The Paradox of Group Autonomy," 263.

Colburn, Ben. *Autonomy and Liberalism*. Routledge, 2010.

Colburn, Ben. "Autonomy, Voluntariness and Assisted Dying." *Journal of Medicine and Philosophy* 46, no. 5 (2020): 316-319.

Conly, Sarah. *Against Autonomy*. Cambridge University Press, 2013.

Crawford, Kate, and Jason Schultz. "AI Systems as State Actors." *Colombia Law Review* 119, no. 7 (2019): 1941-1972.

Deeks, Ashley. "The Judicial Demand for Explainable Artificial Intelligence." *Columbia Law Review* 119, no. 7 (2019): 1829-1850.

Denes, Gyorgy. "A Case Study of Using AI for General Certificate of Secondary Education (GCSE) Grade Prediction in a Selective Independent School in England." *Computers and Education: Artificial Intelligence* 4 (2023): 100-129.

Dorfman, Avihay, and Alon Harel. *Reclaiming the Public*. Cambridge University Press, 2024.

Dworkin, Gerald. "Autonomy and Behavior Control." *Hastings Center Report* 6, no. 1 (1976): 23-28.

Gershgorn, Dave. "Companies Are on the Hook if Their Hiring Algorithms Are Biased." *Quartz,* October 22, 2018. https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased.

Gilbert, Margaret. "Shared Intentions and Personal Intentions." *Philosophical Studies* 144, no. 1 (2009): 167-187.

Guerrero, Alexander. "Against Elections: The Lottocratic Alternative." *Philosophy & Public Affairs* 42, no. 2 (2014): 135-178.

Hicks, Michael Townsen, James Humphries, and Joe Slater. "ChatGPT is Bullshit." *Ethics and Information Technology* 26, no. 38 (2024): 1-10.

Kamiski, Margot, and Jennifer Urban. "The Right to Contest AI." *Columbia Law Review* 121, no. 7 (2021): 1957-2048.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals. A German-English Edition*. Translated by Mary Gregor and Jens Timmermann. Cambridge University Press, 2023.

Kim, Pauline T., and Matthew T. Bodie. "Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy." *ABA Journal of Labor and Employment Law* 289, no. 35 (2021): 289-315.

Lafont, Cristina. *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*. Oxford University Press, 2020.

Lafont, Cristina. "A Militant Defence of Democracy: A Few Replies to my Critics." *Philosophy and Social Criticism* 47, no. 1 (2021): 69-82.

López-Guerra, Claudio. "The Enfranchisement Lottery." *Politics, Philosophy & Economics* 10, no. 2 (2010): 211-233.

Macleod, Christopher. "Mill on Autonomy." In *The Routledge Handbook of Autonomy*, edited by Ben Colburn, 75-84. Routledge, 2023.

Mill, John Stuart. *On Liberty*. Broadview Press, 1859.

Mill, John Stuart. *Considerations on Representative Government*. Parker, Son, and Bourn, 1861.

Oshana, Marina. *Personal Autonomy in Society*. Routledge, 2006.

Piper, Mark. "Justifying Respect for Autonomy." In *The Routledge Handbook of Autonomy*, edited by Ben Colburn, 293-302. Routledge, 2023.

Plato. *Republic*. Translated by Robin Waterfield. Oxford University Press, 1993.

Purves, Duncan, and Davies Jeremy. "Should Algorithms that Predict Recidivism Have Access to Race?" *American Philosophical Quarterly* 60, no. 2 (2023): 205-220.

Rawls, John. *A Theory of Justice: Revised Edition*. Harvard University Press, 1999.

Raz, Joseph. *The Morality of Freedom*. Oxford University Press, 1988.

Ritchie, Katherine. "The Metaphysics of Social Groups." *Philosophy Compass* 10, no. 5 (2015): 310-321.

Scanlon, Thomas M. *What We Owe To Each Other*. Harvard University Press, 1998.

Schaab, Janis. "Kant on Autonomy of the Will." In *The Routledge Handbook of Autonomy*, edited by Ben Colburn, 44-54. Routledge, 2023.

Slater, Joe, James Humphries, and Michael Townsen Hicks. "ChatGPT Isn't 'Hallucinating' – It's Bullshitting." *Scientific American* 34, no. 1 (2025): 46-47.

Umbers, Lachlan Montgomery. "Against Lottocracy." *European Journal of Political Theory* 20, no. 2 (2021): 312-334.

Wellman, Christopher Heath. "The Paradox of Group Autonomy." In *Autonomy. Social Philosophy and Policy*, edited by Ellen Franken Paul, Fred D. Miller Jr, and Jeffrey Paul, 265-285. Cambridge University Press, 2003.

B. Yelubayev & C. Olay · *Conatus* 10, no. 1 (2025): 255-274
doi: https://doi.org/10.12681/cjp.34144

Locke and Rousseau: From Natural Freedom to The Social Contract

Bainur Yelubayev
ELTE Eötvös Loránd University, Hungary; Al-Farabi Kazakh National University, Kazakhstan
E-mail address: yelubayev.bainur@kaznu.kz
ORCID iD: https://orcid.org/0000-0002-5438-9950

Csaba Olay
ELTE Eötvös Loránd University, Hungary
E-mail address: olay.csaba@btk.elte.hu
ORCID iD: https://orcid.org/0000-0001-9713-1418

Abstract

John Locke and Jean-Jacques Rousseau are two eminent proponents of the contractual tradition, which asserts that political power is artificial, and its legitimacy stems from individual consent. The fundamental and common feature of all classical social contract theories is that the agreement concluded by all its participants is considered the basis of a true political body. Accordingly, only a political association based on the concept of a contract can create a form of government that binds naturally free people. The primary purpose of this work is to analyse and compare the contractual views of Locke and Rousseau. Thus, in the first chapter, we will explore Locke's main contractual ideas, developed in his book Two Treatises of Government, emphasising the concepts of the law of nature and private property. In chapter two, we will examine Rousseau's political ideas, particularly on human nature and the general will. Then, in the end, we will attempt to outline the differences and similarities between their views about the social contract.

Keywords: *Locke; Rousseau; social contract; state of nature; private property*

I. Introduction

The social contract theory is widely considered one of the most influential and consequential theoretical traditions in the history of political philosophy. The concepts developed within this tradition have significantly impacted the evolution of ideas related to the normative dimension of politics. Moreover, some authors even consider this theory as the quintessence of modern political philosophy.[1] The contractual tradition claims that political legitimacy, power, and obligations stem from individual consent, and the state is governed only by the consent of the people. Hence, government and society have an artificial nature, and their legitimacy and even their existence depend on an act of individual will rather than on the tenets of theocracy, patriarchy, or the naturalness of political life.[2]

The origin of contract theory can be traced back to classical antiquity[3] and the Middle Ages,[4] where the first discussions about the contractual nature of society and law can be found. Over time, these early ideas evolved, and from the seventeenth century onwards, political thought became dominated by voluntarism, which emphasised individual will. As a result, consent came to be seen as the main criterion of political legitimacy, mainly due to the advent of Christianity in Western thought, which gradually replaced the ancient quasi-aesthetic doctrines of the good political order and the naturalness of human sociality with an approach to politics based on the model of "good acts." Thus, just as good acts necessitated knowledge of the good and the will to pursue it, politics now necessitated moral consent and human participation in politics through one's own free will. The freedom to voluntarily submit to absolute norms has always been essential in Christianity, stressing the importance and merit of individual good will.[5]

In the modern period, Thomas Hobbes, John Locke, Jean-Jacques Rousseau, and Immanuel Kant, who are considered the foundational fathers of the tradition, developed the main concepts and approaches of the theory. We can

---

[1] Deborah Baumgold, *Contract Theory in Historical Context: Essays on Grotius, Hobbes, and Locke* (Koninklijke Brill NV, 2010), ix.

[2] Patrick Riley, *Will and Political Legitimacy*: *A Critical Exposition of Social Contract Theory in Hobbes, Locke, Rousseau, Kant, and Hegel* (Harvard University Press, 1982), 1-2.

[3] For discussions of the origins of contract theory in ancient Greek thought, see C. C. W. Taylor, "*Nomos* and *Phusis* in Democritus and Plato," *Social Philosophy and Policy* 24, no. 2 (2007): 1-20; as well as Rachel Barney, "The Sophistic Movement," in *A Companion to Ancient Philosophy*, eds. Mary Louise Gill and Pierre Pellegrin (Blackwell Publishing, 2006), 77-97.

[4] See David G. Ritchie, "Contributions to the History of the Social Contract Theory," *Political Science Quarterly* 6, no. 4 (1891): 656-676; and John Wiedhofft Gough, *The Social Contract: A Critical Study of Its Development* (The Clarendon Press, 1957), 22-49.

[5] Riley, 2-3.

outline a unified conceptual approach to understanding human nature in this classical period. The concept of the natural state of humanity with its natural rights, such as the right to life, freedom, and property, emerges. Accordingly, the "natural man" becomes the primary agent of classical contractual theories.

The central and common characteristic of all concepts of the social contract is that the agreement concluded by all its participants is considered as the basis of the genuine political body. Accordingly, the social contract is not an agreement between the ruler and the people but between individuals to establish a government. Thus, individuals move from the "state of nature" to the "civil state." In this respect, the development of the idea of equality throughout the history of humankind eventually found political expression in the concept of the social contract, and the Protestant Reformation gave great momentum to the formation of this idea.[6] Therefore, it is not surprising that all classical contractarians were Protestants. The concept of equality implies that all people are equally free. Only a political body based on the concept of contract can create a form of government that binds naturally free people. Any other types of legitimacy, such as the divine right of kings, charisma, and physical strength are no longer valid.[7]

Thus, when Enlightenment ideas began to challenge the validity of traditional moral systems, philosophers turned to social contract theory as an alternative to outdated ethical doctrines, and the principle of the divine right of kings was among the first traditional elements to be called into question.[8] The idea that the monarch held the throne by divine grant lost its relevance even among those who supported the institution of kingship. Thus, while monarchs were ordinary men and women who inherited their extraordinary

---

[6] It is believed that the social contract ideals found their most distinct expression, especially in Calvinism. For example, Calvinists believe that all of humanity is imperfect because of original sin, and therefore no one person or elite can be trusted with unqualified authority. Power is accountable and self-correcting only when it is widely distributed among people. For a more detailed discussion on this topic, see J. Philip Wogaman, "Protestantism and Politics, Economics, and Sociology," in *The Blackwell Companion to Protestantism*, eds. Alister E. McGrath and Darren C. Marks (Blackwell Publishing, 2004), 287-298.

[7] Murray Forsyth, "Hobbes's Contractarianism. A Comparative Analysis," in *The Social Contract from Hobbes to Rawls*, eds. David Boucher and Paul Kelly (The Taylor & Francis e-Library, 2005), 37-38.

[8] It is worth noting that the Enlightenment, alongside the contractual approach, gave rise to various views on political legitimacy and human nature, one of the most prominent of which was Edmund Burke's traditionalism. He argued that a political society is an organism that develops through the incremental accumulation of experience and wisdom, implying that the human will cannot create it. Accordingly, he was highly critical of the abstract and universalistic conception of the "rights of man" put forward by the French Revolution, advocating instead for inherited rights. For more details on this subject and the role of "enthusiasm" in political change, see Christos Grigoriou, "'Enthusiasm' in Burke's and Kant's Response to the French Revolution," *Conatus – Journal of Philosophy* 7, no. 1 (2022): 61-77.

position through chance, the question of legitimacy arises: how can it be justified that some individuals rule while others are ruled when everyone is inherently equal by nature?[9]

Accordingly, the first contract theorists were mainly interested in this specific inquiry: What is the source of our political obligations to those ordinary men and women in power? Their answer can be summarized as follows: In the absence of any natural or divine obligation to obey the rulers, such an obligation can be created through the voluntary act of promising to obey. This appeals to people's personal obligation to fulfil their promises and thus creates a legitimate basis for political obligations.[10]

Thus, unlike his compatriot Hobbes,[11] the prominent English philosopher John Locke described the state of nature positively, calling it the "rule of reason." According to him, people in the state of nature had a "moral consensus" that allowed them to have individual property. It was the violation of this state of affairs that prompted people to create a civil society that legally enshrines their rights to private property. John Locke introduced fundamentally new ideas into the contractual tradition, including popular sovereignty and the right of people to resist governments that fail to act on the trust placed in them. He also argued that absolute power is incompatible with the concept of civil society, and reasonable people would not enter into such an absolute contract.[12]

Jean-Jacques Rousseau was another influential follower of the contractual tradition who was periodically condemned for laying the theoretical foundation for various radical ideologies and regimes. He agrees with Hobbes[13] that in the state of nature, there are no concepts of law, rights, and morality, so people do not have a natural predisposition to follow the moral law.

---

[9] Will Kymlicka, "The Social Contract Tradition," in *A Companion to Ethics*, ed. Peter Singer, 186-196 (Blackwell Publishers, 2000), 186-187.

[10] Ibid., 187.

[11] The first classical theorist of the contractual tradition, Thomas Hobbes, describes the natural state of man extremely pessimistically, calling it a "war of all against all." According to him, people in the "state of nature" are isolated and concerned only with self-preservation, and since everyone is equal and has a natural right to everything, people inevitably get into conflicts over various resources. Therefore, he asserts that there can be no morality in the state of nature: "To this warre of every man against every man ... nothing can be Unjust. The notions of Right and Wrong, Justice and Injustice, have no place." Thomas Hobbes, *Hobbes's Leviathan: Reprinted from the Edition of 1651. With an essay by W.G. Pogson Smith* (Clarendon Press, 1909), 98.

[12] Baumgold, 17.

[13] Hobbes contends that without morality, human life is "Solitary, Mean, Nasty, Brutish and Short." He argues that morality is necessary for a peaceful life, and his First Law of Nature requires that we pursue it, and only if we cannot obtain it do we have a right to resort to the aid and benefits of war. For more details on Hobbes' stance on war, see Jan Narveson, "War: Its Morality and Significance," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 445-456.

However, unlike Hobbes and Locke, he believes that people normally try to avoid causing any harm to others, not because they consider it immoral, but because they have a natural aversion to harm, even if it is directed at others. Consequently, people naturally sympathize with others and get upset when they witness suffering.[14]

In this paper, we will argue that although both philosophers agree that legitimate political authority should be based on the voluntary consent of individuals, they diverge in their understanding of the nature of political power and the role of individuals in political communities. As we will demonstrate, Locke favours limited government, arguing that power should be limited to protecting individual rights and that people have the right to remove a government that fails in its original duties. On the other hand, Rousseau advocates absolute subordination of the individual will to the general will for the common good, believing that the general will is infallible and indivisible. Thus, by examining these different views, this paper seeks to emphasise the contrast between these two philosophers' ideas on the social contract and the nature of political power.

II. John Locke and the law of nature

John Locke considers freedom as one of the most important aspects of human nature. According to him, in the pre-political period, there was freedom and equality between people within the framework of natural law. However, this is not a state of disorder and permissiveness, but a law of nature comprehended by reason, which tells people to respect the natural rights of others to life, freedom and property. He writes:

> In transgressing the law of nature, the offender declares himself to live by another rule than that of reason and common equity, which is that measure God has set to the actions of men for their mutual security.[15]

Thus, Locke contends that the law of nature has a divine origin; in other words, it is the command of God. He directly links natural law with divine law and states that reason does not constitute natural law; reason can only help us to find it.

Locke describes people in their natural state as social beings who live

[14] Jonathan Wolff, *An Introduction to Political Philosophy* (Oxford University Press, 2006), 25.

[15] John Locke, "The Second Treatise: An Essay Concerning the True Original, Extent, and End of Civil Government," in *Two Treatises of Government and A Letter Concerning Toleration*, ed. Ian Shapiro, 100-201 (Yale University Press, 2003), 103.

in peace and community despite periodic conflicts. Over time, these people gradually evolve into large families and tribes, but without political organization.[16] However, the people of this non-politicized society are not Stone Age people. Locke believes that all people who could not unite into a political society are in the state of nature, and he considers some people of his time, such as Indians and Peruvians, to be still in this natural state.[17]

Thus, people living in a state of nature have the liberty to defend their natural right to life, freedom, and property and punish those who try to violate them. Consequently, the right to punish violators belongs not to a particular group of people or the sovereign but to all people:

> And if any one in the state of nature may punish another for any evil he has done, every one may do so: for in that state of perfect equality, where naturally there is no superiority or jurisdiction of one over another, what any may do in prosecution of that law every one must needs have a right to do.[18]

Thereby, the natural state is a state of freedom and equality in which individuals encounter each other without internal authority over each other in their shared status as God's creatures. They are equal in their common position in the normative order of creation. If they violate this order, they forfeit their normative status of equality. When their normative status descends to the level of the lowest members of the order, they become normative beasts that can be treated accordingly by others.[19] Consequently, people may destroy such violators of the order,

> for the same reason that he may kill a wolf or a lion; because such men are not under the ties of the common law of reason, have no other rule but that of force and violence, and so may be treated as beasts of prey, those dangerous and noxious creatures, that will be sure to destroy him whenever he falls into their power.[20]

As it became clear, even though the Lockean people live in peace in the state of nature, there are still partial conflicts between them because of their disobedience to the law of nature and the punishment of those who do not com-

---

[16] Ibid., 133.

[17] Ibid., 106.

[18] Ibid., 103.

[19] John Dunn, *The Political Thought of John Locke: An Historical Account of The Argument of The 'Two Treatises of Government'* (Cambridge University Press, 1969), 106-107.

[20] Locke, *Two Treatises*, 107.

ply with it. Nevertheless, this state of conflict is not the war of all against all as described by Hobbes. According to Locke, the reason for these conflicts is the absence of a superior judge to resolve disputes among people. Therefore, the vacuum of authority to discipline those who transgress the natural law was a serious problem for people in the state of nature. Further, Locke also believes that people cannot be judges in their own cases because if they do, it will lead to disproportionate punishments, since:

> self-love will make men partial to themselves and their friends: and, on the other side, that ill-nature, passion, and revenge will carry them too far in punishing others; and hence nothing but confusion and disorder will follow.[21]

According to Locke, the original abundance of land eventually turned into scarcity not because of population growth but because of greed and the invention of money since there was no reason to take more land than necessary for the survival of families before its appearance. If you produced a surplus, it would just go to waste unless you could exchange it for something more permanent. However, when money was invented, it became possible to accumulate an enormous amount of money without the risk that it would go spoiled. This gave people strong motivation to cultivate more land in order to produce more goods for sale. This condition, in turn, created pressure on the land, which then became scarce. Thus, because of the shortage of land, there were more conflicts and other inconveniences in the state of nature.[22]

Correspondingly, Locke argues that since people in their natural state are always vulnerable to encroachments from others, their possession of natural rights is insecure. This condition, in turn, creates fear and anxiety among people and forces them to leave their natural state to form a political alliance with others who pursue the same goal of protecting natural rights. All this ultimately leads to the birth of the commonwealth.[23]

Thus, by a social contract, individuals transfer their right to enforce the law of nature to the government, thereby creating a political association. As a result of the contract, Locke believes that people become one body politic:

> For when any number of men have, by the consent of every individual, made a community, they have thereby made that community one body, with a power to act as one body, which is only by the will and determination of the majority; for that which acts

[21] Ibid., 105.

[22] Wolff, 23.

[23] Locke, *Two Treatises*, 141-142.

> any community being only the consent of the individuals of it, and it being necessary to that which is one body to move one way.[24]

Consequently, this implies that the state was created to guarantee people's observance of natural law and institutionalize punishment, which, in turn, means that people only transfer the right to punishment to the state and retain their other natural rights.

Accordingly, people always reserve the right to change the government they no longer trust. Moreover, Locke contends that resistance to repressive power is a requirement of human nature:

> For when the people are made miserable, and find themselves exposed to the ill usage of arbitrary power, cry up their governors as much as you will, for sons of Jupiter; let them be sacred and divine, descended, or authorized from heaven; give them out for whom or what you please, the same will happen. The people generally ill-treated, and contrary to right, will be ready upon any occasion to ease themselves of a burden that sits heavy upon them.[25]

In this way, Locke regards resistance as morally justified and calls on all humankind to resist tyrannical forces.[26]

Hence, Locke argues that a government can be overthrown by its subjects if it violates natural law by using force without right. Thus, natural law can also help to determine whether a government is acting legitimately or not. Additionally, he asserts that natural law can only be directly applied in politics in exceptional cases, such as with foreigners or when rulers put themselves in a state of war with their subjects.[27]

Thus, Locke opposes absolute monarchy, arguing that it is incompatible with civil society because it lacks a common authority to resolve conflicts between ruler and subjects. Therefore, the absolute state is even worse than the state of nature because, in it, subjects are deprived of the right to punishment:

> Now, whenever his property is invaded by the will and order

[24] Ibid., 142.

[25] Ibid., 199.

[26] Ibid., 166.

[27] Riley, 68-69.

> of his monarch, he has not only no appeal, as those in society ought to have, but, as if he were degraded from the common state of rational creatures, is denied a liberty to judge of, or to defend his right: and so is exposed to all the misery and inconveniencies, that a man can fear from one, who being in the unrestrained state of nature, is yet corrupted with flattery, and armed with power.[28]

Therefore, Locke argues that civil society needs a peaceful electoral policy and believes that parliamentary sovereignty is the most suitable one. He writes that:

> the people finding their properties not secure under the government as then it was (whereas government has no other end but the preservation of property), could never be safe nor at rest, nor think themselves in civil society, till the legislature was placed in collective bodies of men, call them senate, parliament, or what you please.[29]

According to Locke, the state has three powers: the legislative, the executive, and what he calls the federative power, which is the power to conduct international relations. He defines legislative power as the power that has the right to determine how the state's power is used to protect society and individual rights. However, he also says that people with the power to make laws may also have the power to enforce them, creating the opposite situation where laws are made according to their private interests. Therefore, Locke believes that in well-ordered states, legislative power should be given to various people who seek the public good and disband after the law has been passed and are themselves subject to it.[30]

Hence, Locke argues that societies not governed by declared laws are no different from communities living in a state of nature. However, this does not imply that Locke, like Hobbes,[31] held a positivist view of the nature of law.

---

[28] Locke, *Two Treatises*, 139.

[29] Ibid., 141.

[30] Ibid., 164.

[31] Hobbes adheres to the so-called positivist position regarding the nature of law. He contends that regardless of a law's content and how unfair it may seem if it was prescribed by the sovereign, then it is the law. In other words, law is determined exclusively by the will of the sovereign: "I define Civill Law in this manner, CIVILL LAW, IS to every Subject, those Rules, which the Common-wealth hath Commanded him, by Word, Writing, or other sufficient Sign of the Will, to make use of, for the Distinction of Right, and Wrong; that is to say, of what is contrary, and what is not contrary to the Rule." Thomas Hobbes, *Hobbes's Leviathan: Reprinted*

He believes that regardless of whether power is concentrated in one or many hands, the limit of power is determined by the state's laws, enacted strictly in accordance with the public interest and the law of nature, which is the will of God. In other words, every law enacted by the state must be in accordance with natural law, which commands not to harm the life, health, freedom, or property of others.[32]

Accordingly, in Locke's view, natural law defines only general moral obligations, which is insufficient to formulate political obligations; therefore, consent and social contract are necessary to define political rights and obligations. In other words, natural law defines common moral goods and vices; it cannot legally define what constitutes an offence in the commonwealth.

Thus, Locke argues that originally equal, free, and independent people give up their natural freedom to come together in community for a safe, comfortable life and secure ownership of property. Respectively, Locke considers property to be one of the natural rights of individuals and derives it partly from God, who gave the earth to people, and partly from human labour. He says,

> Though the earth, and all inferior creatures, be common to all men, yet every man has a property in his own person: this nobody has any right to but himself. The labour of his body, and the work of his hands, we may say, are properly his. Whatsoever then he removes out of the state that nature hath provided, and left it in, he hath mixed his labor with, and joined to it something that is his own, and thereby makes it his property.[33]

Ultimately, it can be concluded that concepts such as natural law and natural rights are crucial to fully understand Locke's concept of rights. Locke argues that people create a political system by consent and contract to guarantee natural rights derived from natural law. However, some scholars criticise Locke's natural law by arguing that natural law must stem solely from reason, whereas Locke, in order to make his natural law a real law, used divine rewards and punishments based on immortality, which reason cannot confirm.[34] Accordingly, for Locke, natural law only defines general moral obligations, which is not enough to formulate political obligations, so consent and social contract are necessary to define political rights and obligations. In other

from the Edition of 1651. With an essay by W. G. Pogson Smith (Clarendon Press, 1909), 203.

[32] Locke, *Two Treatises*, 159-160.

[33] Ibid., 111-112.

[34] Riley, 61-62.

words, natural law defines general moral goods and vices; it cannot legally define what constitutes an offence in the commonwealth.[35]

Thus, having carefully examined Locke's central concepts, our attention will now shift to another eminent philosopher and one of the key social contract theorists, Jean-Jacques Rousseau. Close consideration of Rousseau's fundamental ideas on the essence of human beings and the nature of sovereignty will prove particularly useful for further comparing the two authors' views and identifying commonalities and differences in their respective contractual theories.

III. Jean-Jacques Rousseau and popular sovereignty

Rousseau believes that such qualities as greed, pride, oppression and desires, which Hobbes attributes to natural man, actually characterize social man.[36] According to him, the savage man has few desires and needs, which are usually satisfied through hunting and gathering, rather than by attacking others. The savage is a solitary creature who rarely interacts with others and desires only food, sexual satisfaction, and sleep. As for children, they would leave their mothers as soon as they could survive on their own. Hence, there are no families because, according to Rousseau, compassion is not a strong enough feeling to create family ties. Moreover, at this stage, the savage had not yet developed a language, so he was extremely limited in forming and transmitting thoughts and ideas.[37]

Thus, we can observe that all the motives, such as gain, security and reputation, which Hobbes claimed to lead to war, are invalid and defused in Rousseau's natural state. The key point in his thought is that people have two distinctive characteristics: free will and the ability to self-develop, which, according to him, are the sources of both human advancement and misery.[38]

Rousseau contends that the skills and abilities that people developed over time as a result of the progress of their minds eventually led to technological progress. As people began to work and produce, the division of labour and progress led to increased interdependence between individuals. However, this also increased inequality as people learned to compare and compete with each other. Consequently, due to the reality that talented individuals produce more, the division of skills and abilities between people revealed strong and weak, in other words, rich and poor people. The absolute equality and liberty of individuals from nature were irreversibly limited:

---

[35] Ibid., 64.

[36] Jean-Jacques Rousseau, *Emile: Or on Education*, trans. Alan Bloom (Basic Books, 1979), 132.

[37] Ibid., 146.

[38] Ibid., 140-141.

> the moment one man needed the help of another; as soon as it was found to be useful for one to have provisions for two, equality disappeared, property appeared, work became necessary.[39]

Hence, Rousseau argues that the appearance of property opened a chasm between people and created dominant relationships between them. This situation resulted in an insecure and restless social order characterized by a master-slave relationship. He refers to this order as an aggregation of individuals, not association, because there is no political unity or public good.[40]

According to Rousseau, the creation of genuine civil society provides conditions for the moral improvement of people, and the totality of individual wills and freedoms united as a result of a social contract creates a political organism, the so-called "general will," which is infallible, indivisible and cannot be represented. This general will is collective decision-making, which is universal or the most popular and which must be followed by all citizens for the common good and harmony in the state. The general will manifests itself in the voting, the results of which serve as a guide to action. Rousseau writes,

> Each of us puts his person and all his power in common under the supreme direction of the general will; and we as a body receive each member as an indivisible part of the whole.[41]

Thus, when individuals become part of a political body, they unconditionally fall under the subordination of the general will, and this is not a matter of individual choice but a matter of duty[42]:

> In order therefore that the social pact should not be an empty formula, it contains an implicit obligation which alone can give force to the others, that if anyone refuses to obey the general

[39] Ibid., 167.

[40] Jean-Jacques Rousseau, *Discourse on Political Economy and the Social Contract* (Oxford University Press, 1999), 53.

[41] Ibid., 55.

[42] In this regard, Rousseau has often been accused of laying the ideological foundation for many repressive and radical movements and regimes, from the terror era of the French Revolution to the right and left totalitarian regimes of the twentieth century. Especially his idea of the general will has been criticized by scholars as abstract Platonism, which establishes the dictatorship of the state and rejects basic human rights. Some authors believe that all of Rousseau's authoritarian passages are only a restatement of arguments that can be found in French absolutist thought. Jeremy Jennings, "Rousseau, Social Contract and the Modern Leviathan," in *The Social Contract from Hobbes to Rawls*, eds. David Boucher and Paul Kelly, 117-134 (Routledge, 1994), 118.

> will he will be compelled to do so by the whole body; which means nothing else than that he will be forced to be free.[43]

Rousseau argues that democracy is the best form of government for free people. However, he rejects elective democracy and favours direct democracy because it alone can provide the conditions for citizens to act genuinely freely. Without freedom, it is impossible to imagine the emergence of moral citizens because unfree people primarily think about their needs and self-preservation rather than what should be done. Therefore, Rousseau argues that it is only through self-government that people can achieve freedom. By giving up the right to make laws through direct participation, people give up freedom, which means giving up basic needs and human duty:

> To renounce our freedom is to renounce our character as men, the rights, and even the duties, of humanity. No compensation is possible for anyone who renounces everything. It is incompatible with the nature of man; to remove the will's freedom is to remove all morality from our actions.[44]

Correspondingly, just as power is a constitutive characteristic of a person's physical side, so will is a constitutive characteristic of their moral side. As an individual who cannot legally transfer their will to another person, such as in the case of slavery, similarly, a collective body cannot transfer its collective will to others. Thus, for Rousseau, people, as a collective body, not citizens as individuals, become enslaved, transferring their legislative rights to others.[45]

Thus, in a representative democracy with an elected government, people lose their freedom by transferring it to the will of others as elected representatives cannot know the general will and are not obliged to follow it. Instead, they act according to individual will and make laws based on the values and beliefs of groups and individuals rather than on the interests of the whole population.[46] Rousseau contends, "The moment that a people provides itself with representatives, it is no longer free; it no longer exists."[47]

However, Rousseau recognizes that direct democracy can only be effective in geographically small states with homogeneous and unified populations. In exten-

---

[43] Rousseau, *Discourse on Political Economy and The Social Contract*, 58.

[44] Ibid., 50.

[45] Robin Douglass, "Rousseau's Critique of Representative Sovereignty: Principled or Pragmatic?" *American Journal of Political Science* 57, no. 3 (2013): 740.

[46] Hope Sweeden, "Technology and The Social Contract: Is a Direct Democracy Possible Today?" *Susquehanna University Polititcal Review* 7 (2016): 32-33.

[47] Rousseau, *Discourse on Political Economy and The Social Contract*, 129.

sive and densely populated states, the importance of individual will in governance loses its force and relevance. Small ones make it easier for people to legislate and govern because a small and homogeneous population means greater unity in beliefs, values, and ideas. Therefore, Rousseau argues that an increase in territory and population leads to a decrease in the objectivity of governance and the substitution of the interests and will of all citizens for the will of groups and individuals.[48]

Hence, for Rousseau, the state is legitimate only when the people are sovereign and laws are made in accordance with the general will. Rousseau calls this type of regime a Republic. However, the state still needs executive power to enforce the laws passed. In this case, the government can be organised in the form of a monarchy (one magistrate), in the form of an aristocracy (a small number of private citizens), or in the form of a democracy (the entire population or a majority of people). Rousseau argues that all these government forms are legitimate and appropriate in different contexts.[49]

Rousseau goes on to condemn modern political life for the lack of common morality, virtue, and civic religion. Instead, he revered ancient political systems for their high unity, which encouraged people to entirely socialize and be truly political. Rousseau believed that in ancient polities such as Sparta, with its morality of the common good, civic religion, moral use of fine and military arts and lack of individualism, people felt part of a larger entirety. He regarded it as an example of a proper political society and argued that modern people have lost this ancient spiritual vigour due to extreme selfishness.[50]

Thus, Rousseau sought to adhere to both the position that the ancient highly organized political community is the best kind of political system and the idea that all political society is conventional, which is possible solely due to individual will and social contract.[51] Nonetheless, he does not think that the ancient polities were created by a social contract. Instead, he contends that they were created by the genius of legislators such as Moses and Lycurgus.[52]

---

[48] Ibid., 94.

[49] Pedro Abellan Artacho, "Rousseau, Democracy, and His Ideological Intentions: Conceptual Arrangements as Political Devices," *Revista De Estudios Políticos* 186 (2019): 47-48.

[50] Riley, 100-102.

[51] Riley points out that the will, which Rousseau considers the source of all political obligations, is at the same time the cause of everything he hates in modern society. Moreover, he says that the absence of the idea of individual will made possible unified ancient states with common morality. He suggests that Rousseau's idea of a general will was an attempt to combine the generality of ancient morality (unity) with the will of modernity (consent, contract). However, Riley believes that the concepts of generality and will are mutually exclusive, and the will can be considered general only metaphorically. The general will that Rousseau admired in ancient communities is not the general will but the political morality of the common good, where the individual will does not appear with objections to society. Riley, 108-113.

[52] Ibid., 106-107.

Rousseau, thereby, seeks to bring the individual will into line with the general will through the role of the great legislator. He tries to replace the lack of morality of the common good with the wisdom of great legislators. It should be said that Rousseau rejected natural law and believed that the will should correspond to ancient perfection. This creates a contradiction since the ancient standard is non-voluntarist; the standard that gives the will its object is in itself a negation of voluntarism.[53]

Nevertheless, Rousseau's novelty consisted in denying authority identification with one individual. Sovereignty was based on the will of all those who made up the political body. Thus, the theory of absolute monarchy has been altered into an alternative democratic version of absolute popular sovereignty. For Rousseau, sovereignty is an inalienable possession of human beings and part of their essence. It is this idea that distinguishes him from his predecessors, who viewed sovereignty as a temporary possession that had to be transferred to the appropriate authority.

Ultimately, Rousseau ascribes to the people not only the origin but also the exercise of sovereignty. Thus, for him, there was only one contract of association, no contract of subordination, and no losses, only benefits. The individual is "doubly committed" to his contract partners and as a citizen to the sovereign. Thus, the role of a government is to execute the general will expressed by the people as sovereign.[54]

IV. Comparison

As demonstrated, both authors employed the concept of the natural state to develop their contractual theories. However, their accounts have significant differences: Locke's natural man is a social being, while Rousseau's natural man is an antisocial and solitary being. Both depict the natural man as good, free and equal, but in Locke, he is free and equal in the community, while in Rousseau, he is free because he lives alone and is equal only when he encounters others. Thus, for Rousseau, the state of nature

> is the state in which the care for our own preservation is least prejudicial to the self-preservation of others, it follows that this state was the most conductive to peace and the best suited to mankind.[55]

---

[53] Ibid., 115-121.

[54] Jennings, 119.

[55] Jean-Jacques Rousseau, "The Discourses and Other Early Political Writings," in *Cambridge Texts in the History of Political Thought*, ed. V. Gourevitch, 134-160 (Cambridge University Press, 1997), 151.

According to Locke, the law of nature declares that no one has the right to infringe on the life, liberty, and property of others. This can be applied by anyone; in other words, everyone can punish violators of the law of nature. In general, the natural state was characterized by happiness, freedom and equality:

> To understand political power right, and derive it from its original, we must consider what state all men are naturally in, and that is, a state of perfect freedom to order their actions and dispose of their possessions and persons, as they think fit, within the bounds of the law of nature; without asking leave, or depending upon the will of any other man.[56]

Rousseau, for his part, argues that there was true equality in the natural state and that the differences that existed between people were not so significant that they depended on each other, unlike modern civilized society based on illusory equality. Therefore, he asserts that in the natural state before the social contract, our emotions were genuine, and our traditions were crude but natural. According to Rousseau, modern man is born, lives and dies in slavery:

> At his birth he is sewed in swaddling clothes; at his death he is nailed in a coffin. So long as he keeps his human shape, he is enchained by our institutions.[57]

Thus, despite being born free, modern man finds himself bound everywhere, and even those who consider themselves masters of others cannot escape the reality of being slaves.

According to Locke, a person's ownership of oneself presupposes that everything they do with their body and hands belongs to them. Therefore, anything a person creates through their labour becomes their private property. He argues that the transformation of products into property results from people's industriousness and that God has provided for property acquisition as a reward for hardworking and intelligent individuals. Hence, Locke considers the right of ownership as the ownership obtained through human labour, emphasizing labour's role in the acquisition of property:

> It being by him removed from the common state nature hath placed it in, it hath by this labour something annexed to it that

[56] Locke, *Two Treatises*, 101.

[57] Rousseau, *Emile: Or on Education*, 42-43.

> excludes the common right of other men. For this labour being the unquestionable property of the labourer, no man but he can have a right to what that is once joined to, at least where there is enough, and as good, left in common for others.[58]

On the contrary, Rousseau thinks that property resulting from labour ended peace among people. He believes that as people became enlightened, they became industrious. They stopped falling asleep under the first tree or in the first cave and started inventing tools, building huts and covering them with mud to improve their living conditions: "This was the epoch of a first revolution, which established and distinguished families, and introduced a kind of property, in itself the source of a thousand quarrels and conflicts."[59]

Both authors consider the main reason for the transition to a political society to be the emergence of private property and the subsequent need for protection from related violence cases. For Locke, the reason for the transition is the absence in the state of nature of an authorized judge who could resolve disputes that arise, as well as an inadequate ratio between guilt and punishment. Although there are rules based on reason in the natural state, there is no neutral authority that would ensure justice in case of disobedience. Consequently, Locke maintains that while human nature is inherently good, property acquired through labour is still vulnerable to encroachment. Therefore, he believes that the property a person possesses in a state of nature necessitates the establishment of a political association formed through a social contract. Thus, Locke argues that all rights inherent in the natural state are preserved in a political society.

Whereas for Rousseau, the central reason for the transition to a political society is the desire to overcome the state of war that arises as a result of the emergence of private property. In this sense, the social contract is a way to address this issue, preserving some rights during the transition period, while denying others, particularly the right to property, which is the original cause of the problem. Thereby, property is transferred to the sovereign with subsequent redistribution.

According to Rousseau, private property is the basis of civilization. He contends that any progress in civilization is also progress in human inequality. In his opinion, all social and political institutions that have emerged due to civilizational progress have changed their essence and initial purpose, and property is the root cause of this.[60]

Ultimately, both philosophers believed that legitimate political authority should be based on the voluntary consent of the people. Locke, however,

---

[58] Locke, *Two Treatises*, 112.

[59] Rousseau, *The Discourses and Other Early Political Writings*, 164.

[60] Ibid., 161.

favoured limited government and argued that power should be limited to protecting individual rights, maintaining that the people had the right to remove a government if it failed in its original duty. In contrast, Rousseau advocated absolute subordination of the individual will to the general will for the common good because he believed that the general will was infallible and indivisible. Therefore, Rousseau also opposed representative democracy, arguing that the moment people elect representatives, they are no longer free.

V. Conclusion

In conclusion, it has become evident that the concept of the state of nature is one of the integral parts of the philosophical justification of Locke's and Rousseau's contract theories. However, there are significant differences in their views on human nature and the concept of private property, even though both describe the natural man favourably. For Locke, the natural man is a social being living in complete freedom and equality, while Rousseau's natural man is an anti-social and solitary being whose natural state is characterised by true equality.

Locke believed that the natural law governing all things declares that all men are equal and free and that no one should infringe on the life, freedom, and property of others. Moreover, he considered labour mixed with natural resources as the basis of legitimate property ownership. On the other hand, Rousseau argued that private property destroyed harmony between people, and the main reason for the transition to a political society was to overcome the state of war which resulted from the emergence of private property. He believed that concentrated property fuelled social contradictions and favoured a more equitable distribution. Thus, Locke believed that property was necessary and that the social contract preserved all the rights that existed in the state of nature. On the other hand, Rousseau saw private property as the source of inequality and the cause of many of the problems of modern society.

Moreover, it became clear that the issue of sovereignty was central to both authors. While Locke's conception of sovereignty is based on individual rights and government with limited power, Rousseau's understanding refers to the general will reflecting society's common interests. Thus, it is safe to say that the ideas of Locke and Rousseau continue to influence political philosophy today, as people will always strive to create societies that are both just and harmonious.

Author contribution statement

Both authors have contributed equally to the conception and design of the work, the drafting and revising of the manuscript, and the final approval of the version to be published.

References

Artacho, Pedro Abellán. "Rousseau, Democracy, And His Ideological Intentions: Conceptual Arrangements As Political Devices." *Revista de Estudios Políticos* 186 (2019): 45-71.

Barney, Rachel. "The Sophistic Movement." In *A Companion to Ancient Philosophy*, edited by Mary Louise Gill and Pierre Pellegrin. Blackwell Publishing, 2006.

Baumgold, Deborah. *Contract Theory in Historical Context: Essays on Grotius, Hobbes, and Locke*. Koninklijke Brill NV, 2010.

Douglass, Robin. "Rousseau's Critique of Representative Sovereignty: Principled or Pragmatic?" *American Journal of Political Science* 57, no. 3 (2013): 735-747.

Dunn, John. *The Political Thought of John Locke: An Historical Account of The Argument of The 'Two Treatises of Government.'* Cambridge University Press, 1969.

Forsyth, Murray. "Hobbes's Contractarianism. A Comparative Analysis." In *The Social Contract from Hobbes to Rawls,* edited by David Boucher and Paul Kelly. The Taylor & Francis e-Library, 2005.

Gough Wiedhofft, John. *The Social Contract: A Critical Study of Its Development*. The Clarendon Press, 1957.

Grigoriou, Christos. "'Enthusiasm' in Burke's and Kant's Response to the French Revolution." *Conatus – Journal of Philosophy* 7, no. 1 (2022): 61-77.

Hobbes, Thomas. *Hobbes's Leviathan: Reprinted from the Edition of 1651. With an essay by W.G. Pogson Smith*. Clarendon Press, 1909.

Jennings, Jeremy. "Rousseau, Social Contract and the Modern Leviathan." In *The Social Contract from Hobbes to Rawls,* edited by David Boucher and Paul Kelly. The Taylor & Francis e-Library, 2005.

Kymlicka, Will. "The Social Contract Tradition." In *A Companion to Ethics*, edited by Peter Singer, 186-196. Blackwell Publishers, 2000.

Locke, John. "The Second Treatise: An Essay Concerning the True Original, Extent, and End of Civil Government." In *Two Treatises of Government and A Letter Concerning Toleration*, edited by Ian Shapiro, 100-201. Yale University Press, 2003.

Narveson, Jan. "War: Its Morality and Significance." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 445-456.

Riley, Patrick. *Will and Political Legitimacy*: *A Critical Exposition of Social Contract Theory in Hobbes, Locke, Rousseau, Kant, and Hegel*. Harvard University Press, 1982.

Ritchie, David G. "Contributions to the History of the Social Contract Theory." *Political Science Quarterly* 6, no. 4 (1891): 656-676.

Rousseau, Jean-Jacques. "The Discourses and Other Early Political Writings." In *Cambridge Texts in the History of Political Thought*, edited by V. Gourevitch. Cambridge University Press, 1997.

Rousseau, Jean-Jacques. *Discourse on Political Economy and The Social Contract*. Oxford University Press, 1999.

Rousseau, Jean-Jacques. *Emile: Or on Education*. Translated by Alan Bloom. Basic Books, 1979.

Sweeden, Hope. "Technology and The Social Contract: Is a Direct Democracy Possible Today?" *Susquehanna University Political Review* 7 (2016): 27-45. https://scholarlycommons.susqu.edu/supr/vol7/iss1/3/.

Taylor, C. C. W. "*Nomos* and *Phusis* in Democritus and Plato." *Social Philosophy and Policy* 24, no. 2 (2007): 1-20.

Wogaman, Philip J. "Protestantism and Politics, Economics, and Sociology." In *The Blackwell Companion to Protestantism*, edited by Alister E. McGrath and Darren C. Marks. Blackwell Publishing, 2004.

Wolff, Jonathan. *An Introduction to Political Philosophy*. Oxford University Press, 2006.

A. Zhangalieva et al. • *Conatus* 10, no. 1 (2025): 275-295
doi: https://doi.org/10.12681/cjp.37337

The Problem of Space and Time in Kazakh Falsafa

Ainur Zhangalieva,[1] Garifolla Yessim,[2] Beken Balapashev,[3] Manifa Sarkulova,[4] and Aigul Tursynbayeva[5]

Abstract

The research aims to analyse the relationship between space and time in different directions, presenting the concepts and historical factors that influence these concepts. The methods used to conduct the research are questionnaire, analysis, and comparison. The proposed methods provide an opportunity to identify multiple opinions and discussions about space and time, to unite them by common features, and to make an in-depth analysis of the factors that influence them. The main results of the study are to identify the characteristic features of space and time in the country, to describe and analyse the concepts and historical events influencing them, to identify the main problems and to formulate recommendations for future generations. The conducted research has provided a deeper understanding of the significance of space and time for modern society and can serve as a basis for further study of this issue. The practical significance of studying space and time concepts includes enhancing understanding among academics and the public, developing educational programs to appreciate cultural identity, facilitating cross-cultural exchanges, and contributing to societal sustainable development.

Key-words*: space and time; unique concepts; perceptions; historical factors; contemporary society; sharing experiences*

[1] L. N. Gumilyov Eurasian National University, Republic of Kazakhstan; West Kazakhstan Marat Ospanov Medical University, Republic of Kazakhstan. E-mail address: ain.zhangalieva@gmail.com. ORCID iD: https://orcid.org/0009-0002-9862-7372.
[2] L. N. Gumilyov Eurasian National University, Republic of Kazakhstan. E-mail address: garifolla_yessim@outlook.com. ORCID iD: https://orcid.org/0009-0005-8523-3202.
[3] L. N. Gumilyov Eurasian National University, Republic of Kazakhstan. E-mail address: bbalapashev@gmail.com. ORCID iD: https://orcid.org/0009-0000-0268-9778.
[4] L. N. Gumilyov Eurasian National University, Republic of Kazakhstan. E-mail address: sarkulova3@outlook.com. ORCID iD: https://orcid.org/0009-0000-5540-4254.
[5] L. N. Gumilyov Eurasian National University, Republic of Kazakhstan. E-mail address: tursynbayeva-a@hotmail.com. ORCID iD: https://orcid.org/0009-0005-1086-6013.

I. Introduction

The examination of space and time plays a crucial role in the philosophical discourse on cultural heritage in Kazakhstan. The study helps analyse these concepts, consider different philosophers' points of view, and prepare programmes for further development. It helps define and preserve national identity, promotes dialogue between different cultures, identifies factors that interact with space and time, and suggests options for future research. The limited number of studies of ideas about space and time and the heterogeneity of the philosophical tradition, which contains different points of view, is problematic. The presence of such influencing factors as the culture, traditions, and history of the Kazakh people in space and time contributes to the understanding of the uniqueness of this nation. The study should consider the fact that philosophical ideas about space and time presented earlier may no longer be relevant now, so it is necessary to properly assess their relevance for modern society and determine how they can be supplemented with new knowledge. Comparison of concepts about space and time with other peoples will help to identify common and distinctive features and to find the potential for discoveries and interpretations.

Next, we examine the perspectives of prominent Kazakh philosophers Raushan Sartayeva and Askar Battalov on the concepts of space and time. These philosophers have made significant contributions to the discourse, providing insights that are deeply rooted in Kazakh cultural and historical contexts. Raushan Sartayeva[1] defined the relationship between space and time because they express orderliness in the world. The author analysed these concepts from different points of view and concluded that time orders events that follow each other, and space – those that follow one after another. According to Askar Battalov,[2] space and time are closely interrelated with the history of the Kazakh people. The researcher analysed its influence on the consciousness and quality of human life; and concluded that it is important to study history based on the experience of previous generations because it helps to determine the identity of the nation. By focusing on the works of these key philosophers, we aim to illuminate the unique ways in which space and time are conceptualized in Kazakh philosophical thought, offering a culturally specific perspective that enriches the broader philosophical discourse.

Following Galia Temirton,[3] the perception of space and time is closely related to the culture of the people, which can interpret these concepts in different

---

[1] Raushan Sartayeva, "Abay's Teaching 'Tolyk Adam' and Modern Tendencies in Solving the Problem of the Whole Man," *Adam Alemi* 89, no. 3 (2021): 76-91.

[2] Askar Battalov, "Current Problems of Teaching the History of Kazakhstan in Schools," in *Materials of the International Scientific-Practical Conference "School – Teacher – Innovations in the Modern World,"* ed. Zhanbol Zhilbaev, 249-254 (Pavlodar Pedagogical University, 2021).

[3] Galia Temirton, "The Role of National Traditions in Cultural Integration," *KazNU Bulletin. Series: "Historical and Socio-Political Sciences"* 3, no. 66 (2020): 1-7.

ways. The author found that cultural values and norms have a significant impact on behaviour and perception of time and space, which in some are strictly controlled, while others have a more relaxed approach. Alisher Ismailov and Lazzat Omarbaeva[4] studied the influence of spatial and temporal characteristics that have a significant impact on the formation, development, and preservation of traditions. The authors determined that rituals and crafts, passed from generation to generation, provide continuity in time, and the space where traditions develop influences their form and meaning. The researchers concluded that traditions associated with a particular space and time can be unique and reflect local culture and identity.

The study of Seytkali Duisen and Kayrken Adiyet[5] was devoted to the relationship between space and time and politics. The scientists analysed their influence on territorial, historical and cultural aspects. They concluded that space and time determine the borders of states, periods of government, and interactions between countries, as well as affect the identity of people, and the formation of political systems and ideologies. Akmaral Dalelbekkyzy and Nurlan Yildiz[6] analysed the influence of space and time on the identity of the people. The authors found that all these determine the unique cultural, historical, and social features that unite the people into a single whole and form their own national belonging and personal consciousness.

The research aims to find and analyse different views on space and time, identify the interaction with the culture, traditions, and history of the people, determine the relevance of the topic under study and complement, and compare it with other ideas on these concepts. The assignments are to examine and analyse previous research, design, and conduct a questionnaire, and formulate further perspectives to increase the relevance of the topic, contribute to the academic field and bring Kazakh philosophy to the attention of the international community.

II. Materials and methods

Theoretical and empirical methods were used to study the problem of space and time in the falsafa of Kazakhstan. In the empirical part of the study, a questionnaire survey was used to obtain structured and quantitative information about opinions, views and understanding of the problem. In the questionnaire, 200 people aged between 20 and 60 years took part. The period of comple-

[4] Alisher Ismailov and Lazzat Omarbaeva, "Spiritual Revival – Moving Force, Changing Traditions and World Ideology of the Kazakh," *Scientific Journal "Auezov University"* 4, no. 56 (2020): 188-191.

[5] Seytkali Duisen and Kayrken Adiyet, "Foreign Policy of the Republic of Kazakhstan in the Works of Domestic Researchers," *KazNU Bulletin. Series: Historical and Socio-Political Sciences* 72, no. 1 (2022): 164-176.

[6] Akmaral Dalelbekkyzy and Nurlan Yildiz, "The Influence of the National Code on Historical Consciousness," *KazNU Bulletin. Series: Philological* 184, no. 4 (2021): 156-166.

tion was 3 weeks. The participants are from different regions of Kazakhstan. They are students of higher educational institutions, teachers and researchers, representatives of academic institutions and scientific organisations, and members of the Kazakh intellectual community. An equal ratio of women and men was achieved. Respondents represented different socio-economic groups. The questionnaire survey was conducted at universities, online and other research institutions. Participants were ensured confidentiality and ethical behaviour.

To achieve an objective and comprehensive understanding of the topic, participants provided their demographic details, such as age, gender, occupation, education, and region. They were asked to share their perceptions of the relationship between time and space and the relevance of this issue in contemporary Kazakh philosophy. Participants also reflected on how the younger generation's understanding of space and time differs from that of previous philosophers and thinkers, identifying significant ideas and principles in both past and present works. The survey explored whether the problem of space and time is of universal significance or influenced by Kazakhstan's cultural, political, and historical contexts. Respondents noted any philosophical sources or theories they had studied, assessed the impact of culture on the development of philosophical ideas about space and time, and discussed its reflection in Kazakh literature, art, and religion. They also addressed the challenges and contradictions in the current study of this problem, its significance for Kazakh identity and national consciousness, and the effects of modern science and technology on contemporary philosophical thought. Finally, participants identified the main challenges and proposed solutions to these issues.

The questionnaire method is an effective tool for obtaining data on the problem of space and time of Kazakh falsafa. It was used to identify the reasons and motivations that underlie opinions and views on the topic under study and to identify common trends for different regions and participants. The theoretical part of the study conducted a detailed analysis of the questionnaire, comparing it with the results obtained by other scholars. The method of analysing data from the empirical part was used to identify the main themes and trends related to the problem of space and time. It was used to identify interrelationships and patterns that are unnoticeable upon initial examination, identifying differences and similarities of multiple perspectives. The method of analysis was used to draw conclusions and make recommendations for further research, educational programmes, or other actions. The method of comparing the data obtained in the study was a powerful tool for in-depth and objective analysis of opinions and views on the problem of space and time in falsafa. It contributed to a more comprehensive understanding of the topic and the development of informed approaches to it. This method provides a systematic comparison of data, which helps to better understand

the perceptions of the topic and to identify important aspects that may be missed if the data are analysed without comparison.

III. Results

Statistical data collection in the study was carried out utilizing filling in the proposed questionnaire for the participants. The questionnaire aimed to collect opinions, views and perceptions of scientists, philosophers, and researchers on the relationship between space and time in Kazakh philosophy, as well as to identify and analyse the main topics, approaches and concepts related to this issue. Data on the relevance of the researched topic is presented by the number of answers from respondents (Figure 1).

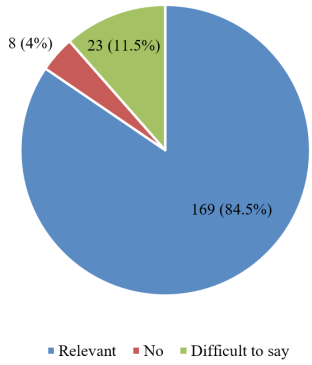


Figure 1. Relevance of space and time

It can be concluded that in modern society the problem of space and time is relevant and requires comprehensive study and analysis. The questionnaire data helped to determine the relationship of the studied direction with other factors (Figure 2).

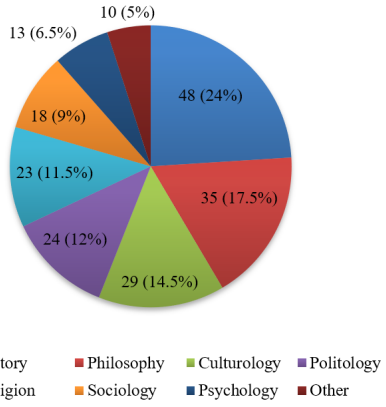


Figure 2. Relationship of space and time with different directions

The results of the questionnaire also allow to present the views on space and time of Kazakhs and formulate common views for all respondents. They are closely interconnected with each other (Figure 3).

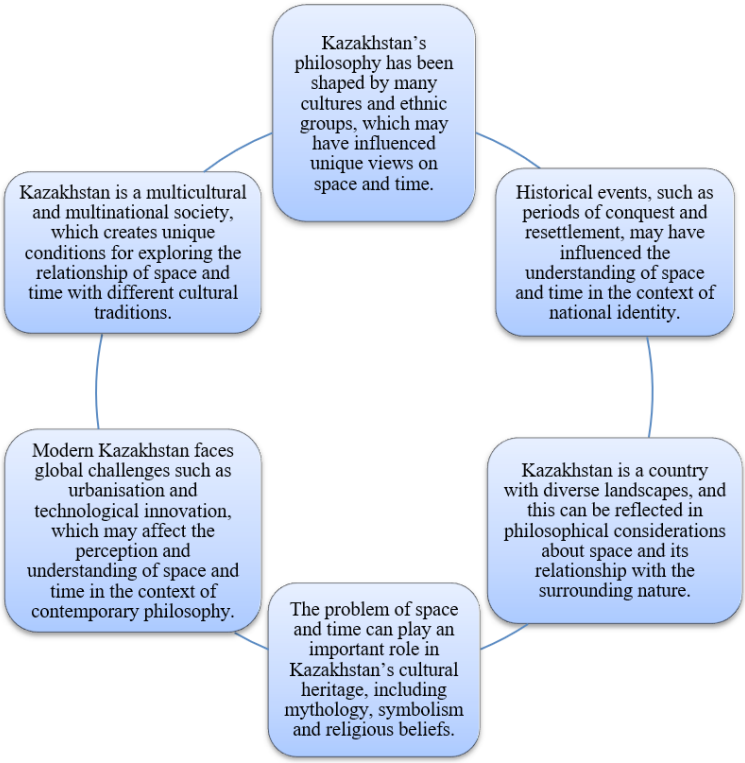


Figure 3. Specific features of the problems of space and time in Kazakhstan

The study shows that the perception of space and time in Kazakhstan is deeply rooted in the national culture, history, and worldview, and continues to evolve under the influence of modern changes and challenges.

i. Concept of time and space

There are many concepts of space and time in philosophy that reflect a variety of views:

1. Space and time as an illusion. Space and time as illusion is a concept that suggests that they exist only in human perception and consciousness and have no objective reality. People perceive them through their senses and experiences, but they do not exist independently of them. Phenomenalism ar-

gues that space and time are merely phenomena arising from perception. Constructivism sees them as constructs of the human mind to interpret the world. Idealism says that they exist only in consciousness. In Eastern philosophical traditions, they are considered maya, a cosmic illusion.[7]

The idea that space and time are illusions stems from several philosophical traditions that argue these concepts do not possess objective reality but exist primarily in human perception and consciousness. In philosophical phenomenalism, exemplified by thinkers like Kant, space and time are seen as mere phenomena; they do not exist independently but are ways in which we structure our experience of the world. Kant's idea that time and space are a priori forms of intuition suggests that they frame how we perceive rather than correspond to external realities. In the context of Eastern philosophy, which has influenced Central Asian thought including that within Kazakhstan, space and time are often viewed through the lens of Maya in Hinduism and Buddhism. Maya represents the deceptive, illusory quality of the world, which includes temporal and spatial dimensions. This perspective resonates with the Kazakh spiritual and philosophical heritage, where the impermanence and illusion of worldly existence are acknowledged in folk traditions and oral literatures.

Constructivism, another relevant approach, suggests that our understanding of space and time is shaped by cognitive processes that interpret sensory input. This aligns with modern psychological and cognitive sciences, which demonstrate how deeply our cultural and personal backgrounds influence our perception of reality. Within the Kazakh context, this could be linked to the nomadic heritage, where the perception of space as vast and unbounded and time as cyclical rather than linear reflects an adaptation to the life and challenges in the steppes.

2. Space and time as fundamental structures of reality. Space and time are considered as fundamental structures of reality, possessing objective existence beyond human perception. They are closely related and form the basis for understanding the universe and its laws. In the theory of relativity, they are combined into a four-dimensional space-time that depends on gravity and the motion of objects. Space and time are considered fundamental concepts, but they are also abstractions used to describe the world and conduct scientific research. As structures of reality, they play a crucial role in the organisation and functioning of the universe, becoming an integral part of the underlying structure, and forming the basis for understanding the world.[8]

---

[7] Manfred Poser, *Time and Consciousness: Why Time is an Illusion* (Crotona Verlag, 2020); Tamara Nedoshovenko, "Time and Space as Tools of Cognition of the External World in the Fine Arts," *Kultura i Suchasnist: Almanakh* 1 (2020): 139-145.

[8] Walter Bloch, *Mysteries of Space and Time: Synchronicity and Nonlocality* (Crotona Verlag, 2020).

3. Space and time as a relationship. The concept of space and time as relations implies that they do not exist independently but are ways of organising and measuring phenomena in the universe. Space and time are defined by the relationships between objects or events. This abstract concept helps to understand the relationships between objects and events, making space and time key elements for understanding the world.[9] Phenomenology of space and time is a philosophical approach focusing on the experience and perception of these aspects. It views them as subjective and constructive categories, dependent on individual experience and openness to different perspectives. This approach allows to better understand the nature of man and the world through comprehending the direct experience of space and time.

4. Space and time as subjective and cultural constructs. The concept of space and time as subjective and cultural constructs implies that these aspects of reality are influenced by cultural and social contexts and are also shaped by the subjective perceptions of each individual. Different cultures may have different perceptions of space and time, and socio-cultural factors influence the perception and interpretation of these aspects. The approach shows how local and individual contexts shape the understanding of the world through subjective perceptions of space and time.[10]

The notion that space and time are subjective and culturally constructed argues that different societies and individuals can experience and understand these concepts in diverse ways. This approach emphasizes the role of cultural, historical, and social contexts in shaping our perceptions of space and time. In the Kazakh context, this is particularly evident in the traditional nomadic lifestyle, which engendered a specific understanding of space as something fluid and expansive. The seasonal migrations (transhumance) across vast steppes shaped a unique spatial perception that contrasts markedly with the sedentary, bordered conceptions prevalent in more urbanized societies. Similarly, the traditional Kazakh sense of time, which revolves around natural cycles and events (e.g., seasons, animal behaviors), contrasts with the Gregorian calendar and clock time that dominate in industrialized contexts.

Cultural anthropologists and sociologists studying Kazakh society have noted how these perceptions influence everything from social organization to music and poetry, where space and time are often fluid rather than fixed dimensions. Philosophers like Mircea Eliade have discussed how such cyclical and natural conceptions of time in traditional societies differ fundamentally from the linear, historical time perceived in much of Western philosophy. By incorporating these richer, culturally infused perspectives into our under-

[9] Rudolf Meer, "Alois Riehl's Theory of Space and Time in the Context of Realistic Interpretations of Kant's Transcendental Idealism," *Kant-Studien* 113, no. 3 (2022): 459-486.

[10] William Sahakian, *History of Philosophy* (Barnes & Noble, 2022).

standing of space and time, it becomes clear that these are not merely neutral or universal dimensions but are deeply intertwined with the fabric of individual and collective life. This helps elucidate how Kazakhs, historically and in contemporary settings, navigate and interpret their world, offering a deeper, more nuanced understanding of these philosophical concepts.[11]

5. Space and time as infinity and finitude. The concept of space and time as infinity and finitude represents a philosophical approach that considers their extent and duration. The infinity of space and time implies their unlimitedness, the absence of limits in extension and duration. Spatial dimensions can be infinite, and time can be eternal, extending into the past and future without end. On the other hand, the finiteness of space and time implies boundedness and the existence of limits. Space can be limited, related to the size of the Universe and its structure, and time can have a beginning and an end, forming a time continuum.[12]

These concepts demonstrate the richness and versatility of philosophy, in which space and time are important components of the formation of cultural heritage and national identity. They provide a better understanding of the relationship of culture to the world around it, history and modernity, and the formation of unique values and enduring traditions.

ii. The impact of historical events on the philosophical outlook of Kazakh people

Migration has played a pivotal role in the historical and cultural evolution of Kazakhstan. These movements across the Kazakh landscape have influenced the region's geographical, socio-cultural, and political realms, fostered a rich tapestry of diverse cultural traditions and promoting the exchange of knowledge among various ethnic groups. This dynamic has crucially shaped the cultural heritage and collective mentality of the Kazakh people. The introduction of Islam deeply impacted the Kazakh worldview, intertwining religious rituals, holidays, and traditions with daily life, and reinforcing a unified national identity. Islamic practices introduced precise temporal frameworks for rituals like prayers and fasting, while Islamic philosophy brought new perspectives on time and cosmology, shaping the spiritual and philosophical discourse in Kazakhstan.[13]

---

[11] Raikhan Doszhan, "The Problem of Existence in the Heritage of Al-Farabi and Its Continuity with Modern Scientific Knowledge," *Arts Academy. Series: Social and Human Sciences* 3, no. 7 (2023): 50-64.

[12] Sanem Kulatti, *David Hume's Critique on Infinite Divisibility of "Space" and "Time"* (MA Thesis, Adnan Menderes University, 2019).

[13] Zabira Myrzatayeva, "Historical Memory," *KazNU Bulletin. Series "Historical and Socio-Political Sciences"* 2, no. 65 (2020); Ibrahim Ozdemir, Kalmurat Subanbayev, and Ayman Keldinova,

During the Soviet era, the landscape of Kazakh philosophy was dramatically altered by industrialization, collectivization, and atheistic propaganda. These changes not only reconfigured the spatial and social structures but also redefined the relationship with the historical past and religious traditions, significantly influencing the philosophical understanding of time and space. The post-independence national revival has been instrumental in reinforcing Kazakhstan's cultural identity, emphasizing the preservation of traditional customs, language, and folklore. This resurgence has not only highlighted historical continuity in national consciousness but also reshaped the philosophical perceptions of space and time, focusing on the nation's heritage and its implications for the future.[14]

In the contemporary context, globalization and technological advancements are reshaping perceptions of space and time, diminishing physical distances and integrating diverse cultures. This global connectivity brings new challenges and opportunities, influencing philosophical thought in Kazakhstan by introducing novel ideas while also prompting reflections on cultural preservation amidst global integration.[15]

Adapting Western philosophical traditions to Kazakh culture poses challenges in interpreting concepts of space and time, due to limited research and academic support in this field. The task of preserving traditional philosophical conceptions amidst contemporary changes is critical, yet complex, due to the influence of cultural, ethnic, and religious factors on Kazakh philosophy. Furthermore, integrating these studies into educational programs is essential for raising public awareness but remains a hurdle. Developing new philosophical approaches to space and time in Kazakh philosophy requires interdisciplinary cooperation and an understanding of the unique identity of Kazakh philosophical tradition, while still acknowledging global philosophical influences. To address these challenges, funding for research and the establishment of academic programs and chairs dedicated to this field are vital. Hosting scientific conferences and public events can facilitate knowledge exchange and public engagement, enhancing understanding across disciplines. Re-

"Issues of Continuity of Kazakh Philosophy," *Al-Farabi* 78, no. 2 (2022): 72-85; Oleksandr Demenko, "The Impact of Historical Traditions on the Foreign Policy Priorities of the Republic of Kazakhstan," in *Trends in International Relations and Problems of European Security: A Collection of Scientific Works*, ed. Serhiy Tolstov, 120-137 (Institute of World History of the National Academy of Sciences of Ukraine, 2019).

[14] Ajar Shaldarbekova, "On the Current State of National Identity in Kazakhstan: Historical Aspect," *Asian Journal "Steppe Panorama"* 6, no. 2 (2019): 328-332; Vedat Karagun and Said Aras, "Globalization and the Information Age," *Dicle Akademi Dergidi/Journal of the Dicle Academy* 2, no. 1 (2022): 32-40.

[15] Lyudmila Gotz, "Distinction of the Categories 'Synchrony, Diachrony' and 'Chronotope' in the Methodology of Cultural Studies," *Almanac Culture and Contemporaneity* 2 (2021): 45-50.

search that bridges traditional Kazakh ideas with contemporary philosophical thought can both preserve and evolve Kazakh philosophy. Analyzing Kazakh literature, art, and religious texts will clarify the unique aspects of space and time within the Kazakh context. Collaboration with international scholars and institutes, and the creation of research centers and databases, are key to enriching and disseminating knowledge in this field, contributing significantly to the cultural and scientific development in Kazakhstan.

IV. Discussion

The analysis of the questionnaire results provides a deep understanding of the relationship between culture, history, philosophy and the concepts of space and time. The results helped to identify general trends in the perception of Kazakh culture, as well as specific features related to historical, cultural, and philosophical contexts. In addition, the analysis helped to shed light on contemporary challenges and perspectives that affect the understanding of space and time in contemporary Kazakh falsafa, including the impact of globalisation and technological progress. This may contribute to a better understanding of the nation's cultural heritage and identify new research directions in this field.

The findings of the current study on the problem of space and time in Kazakh falsafa align with several key themes in contemporary philosophical discourse. Jacob Andrew Bell's[16] exploration of existential meaning as an experiential and holistic phenomenon resonates with the way Kazakh falsafa perceives space and time not just as abstract concepts but as deeply embedded in the lived experiences and cultural heritage of the Kazakh people. The current study highlights that space and time in Kazakh philosophy are intricately tied to the cultural, historical, and social context, reflecting Bell's argument that meaning emerges from the relationship between human beings and their world. Babalola Joseph Balogun's[17] critical engagement with the concept of community underscores the importance of shared spaces and collective identity, which is reflected in Kazakh philosophy's emphasis on the communal aspects of space and time. The study shows that Kazakh thinkers often perceive space and time through the lens of community and tradition, highlighting how these concepts are not only personal but also collective, influencing and being influenced by the shared cultural and historical experiences of the Kazakh people. This communal perspective aligns with Balogun's call for an understanding of community that incorporates shared spaces and collective values.

---

[16] Jacob Andrew Bell, "The Reinstatement and Ontology of Meaning," *Conatus – Journal of Philosophy* 8, no. 1 (2023): 77-86.

[17] Babalola Joseph Balogun, "How Not to Understand Community: A Critical Engagement with R. Bellah," *Conatus – Journal of Philosophy* 8, no. 1 (2023): 55-76.

Omobola Olufunto Badejo[18] addresses the problem of persistence in metaphysics, proposing the need for new theories beyond Endurantism and Perdurantism. Similarly, the current study identifies the necessity of re-evaluating and possibly redefining traditional Kazakh philosophical concepts of space and time to ensure their relevance in modern society. The dynamic and evolving nature of space and time in Kazakh philosophy reflects Badejo's call for alternative theories that can better capture the diachronic identity of concepts. Thiago Pinho's[19] proposal of an object-oriented social theory that incorporates the insights of philosophers of life and process aligns with the study's findings on the influence of modern science and technology on contemporary Kazakh philosophical thought. The integration of global philosophical influences and the impact of technological advancements highlight the evolving nature of space and time in Kazakh philosophy, mirroring Pinho's emphasis on the benefits of incorporating diverse philosophical traditions into social theory.

Recep Yilmaz[20] examined the relationship between space, time, and cultural studies. The author analysed the historical epochs of Islamic civilisation and how each of them influenced its traditions, symbolism, literature, language, and architecture. The author concluded that understanding the timeline and historical events of a nation is important for its cultural identity in the modern world. Cultural heritage transmitted through temporal and spatial contexts can shape collective memory and a sense of belonging to a particular nation or ethnicity.[21] Analysing the relationship between space, time and culture can contribute to a better understanding of national characteristics and the overall picture of societal development.[22] However, to gain a more complete picture, a multicultural perspective, the influence of social, economic, and political factors, and interaction with other regions must be considered. The study of the development of art, music, the impact of technology and globalisation also plays an important role in understanding the formation and preservation of a nation's cultural heritage and identity. Combining all these aspects allows to penetrate deeper into the essence of

[18] Omobola Olufunto Badejo, "The Persisting Problem of Persistence: A Call for an Alternative Theory," *Conatus – Journal of Philosophy* 7, no. 1 (2022): 9-31.

[19] Thiago Pinho, "Six Steps Towards an Object-Oriented Social Theory (O.O.S.T)," *Conatus – Journal of Philosophy* 8, no. 1 (2023): 263-283.

[20] Recep Yilmaz, "Perception of Historical Time in the West and Turkish-Islamic Culture," *Journal of Social Sciences of Mus Alparslan University* 10, no. 2 (2022): 899-912.

[21] Alima Auanasova, Kamilla Auanassova, and Ganizhamal Kushenova, "Alash Party and Issues of National Statehood of Kazakhstan," *Nuova Rivista Storica* 108, no. 2 (2024): 535-549.

[22] Alima Auanasova and Kamilla Auanassova, "The Struggle for Kazakh Statehood in 1917-1918 through the Prism of the History of Constitutionalism," *Investigaciones Historicas* 44 (2024): 641-661.

culture and to understand how space and time influence its development and evolution in modern society.

Gökhan Çinkara[23] presented the influence of space and time on the country's politics. He investigated how they are reflected in Turkey's domestic and foreign policies. He concluded that the relationship of space and time in this area helps to understand the complex interactions between policy decisions and the context in which they are made and allows anticipating potential challenges and opportunities for countries and international communities. The relationship between world political events and global issues such as climate change and environmental crises is also worth noting. Global political actions and decisions can have a significant impact on sustainable development and the future of the planet. However, it is also important to recognise that political decisions are not always made with long-term consequences or global impact in mind. Some political forces may focus on short-term interests or narrow group priorities, which can lead to a disregard for sustainability and the environment.[24]

Kizil Ömür and Cengiz Donmez[25] presented the relationship between space, time, and history, which forms the basis for understanding the evolution of human civilisation. They determined that time and space have become a key factor in the transmission of knowledge and cultural heritage, the formation of collective memory and the identification of lessons from the past. Developed recommendations for more effective teaching of history in educational institutions. This view emphasises the importance of analysing historical phenomena in terms of context, time, and geographical location, which enriches understanding of past events and processes. Additional research options may include the study of migrations, climate change, trade, comparative analysis of civilisations, historical conflicts and peace processes, and the role of space and time in the formation of cultural identities. These studies enrich knowledge of the past and provide a better understanding of the present.

Jeremy Walton and Neena Mahadev[26] focused on the relationship of space and time with religion, which is an important aspect of human culture

---

[23] Gökhan Çinkara, "The Geopolitical Context of Nationalism and the Transformation of Political Elites in Turkey: Memory, Identity, and Space," *Akademik Hassasiyetler/The Academic Elegance* 10, no. 21 (2023): 549-581.

[24] Ganizhamal Kushenova, Alima Auanasova, Albina Maxutova, and Ainur Kairullina, "Kazakh History in British Periodicals: Interpretations and Historical Accuracy," *Bylye Gody* 20, no. 1 (2025): 129-137.

[25] Kızıl Ömür and Cengiz Donmez, "The Model of Synchronic Approach and Scaled Synchronological Charts to Teaching Historical Subjects," *International Journal of New Approaches in Social Studies* 4, no. 2 (2020): 285-308.

[26] Jeremy Walton and Neena Mahadev, "Introduction: Religious Plurality, Interreligious Pluralism, and Spatialities of Religious Difference," *Religion and Society: Advances in Research* 10, no. 1 (2019): 81-91.

and history. They analysed how religious places and temples; time frames and historical narratives influence the formation of beliefs. The authors concluded that the interaction of time and space determines the religious identity of peoples and cultures and influences historical events. Spatial features define sacred places, rituals, and symbolism.[27] Time cycle's structure religious life and festivals.[28] Research in this area helps to understand the role of beliefs in culture and society and their impact on ecology and migration. Possible other areas of study: the role of place and pilgrimage, rituals and calendars, architecture and symbolism, the influence of beliefs on social organisation, religion and ecology, beliefs, and migration.

Michael Strand and Lyn Spillman[29] focused on the connection of space and time with sociology, which allows to study the interaction of society with the surrounding world and time frames. They analysed the evolution of society under the influence of a sequence of events, and changes in social norms and values. Developed recommendations for predicting and adapting to changes in the socio-cultural space. There are also other important aspects to study such as social structure, the impact of globalisation, the role of virtual space and crises on society, and the psychology of space and time. This broadens the understanding of sociology and helps to predict changes in society.

Mehmet Volkan Demirel[30] identified the relationship of space and time with psychology and highlighted interesting aspects of human perception, cognition, and behaviour. He noted that individual perception of time is also related to psychological characteristics such as motivation and meaning in life. He analysed how spatial and temporal factors can be reflected in behaviour, e.g., deadlines and time constraints can affect decision-making and stress levels. He developed recommendations for increasing a person's psychological comfort level based on the influence of space and time on emotional state and mood. Analysing the psychology of spatial perception and temporal perspectives allows for a better understanding of how people form their perceptions of themselves and society.[31] It is also important to study the impact of architecture, design, and cultural differences on human psychological well-being.

---

[27] Gulnara Jumabekova, Galiya Bazarbayeva, Victor Novozhenov, Elina Altynbekova, Anton Gontscharov, and Aliya Manapova, "'Sun and Steppe – Eternal Entities': A Museum Reconstruction Model With a Mace From the Trans-Tobyl Region (Northern Kazakhstan)," *Archaeology of Kazakhstan* 26, no. 4 (2024): 143-180.

[28] Viktor Novozhenov, "Vehicles in the Bronze Age Petroglyphs of Kazakhstan: Mobility and Elitism," *Archaeology of Kazakhstan* 24, no. 2 (2024): 70-99.

[29] Michael Strand and Lyn Spillman, "Cultural Sociology," in *The Cambridge Handbook of Social Theory*, ed. Peter Kivisto, 43-62. Cambridge University Press, 2020.

[30] Mehmet Volkan Demirel, *Time, Aspect and Mood/Modality in Language* (Pegem Academy, 2021).

[31] Andrei Efremov, "The Psychology of Faith and Religious Identity: How Theology Shapes Worldview and Self-Perception," *Pharos Journal of Theology* 106, no. 3 (2025): 1-15.

Zhenci Xu et al.[32] determined that technology allows for to reduction of spatial and temporal distances by providing fast communication and information transfer. The authors analysed telecommunications and the Internet, which allow the exchange of information online and affect the perception of space and time, making it fast and dynamic. Suggested options for predicting future trends in technology and its impact on space, and time. Technologies shorten distances by providing fast communication and information transfer, change the perception of space and time, influence socio-cultural aspects, and raise new ethical issues.[33] The attention should be also paid to the influence of virtual reality in the formation of new spatial and temporal impressions.

Kenan Mutluer[34] presented the important role of space and time in philosophy. He focused on understanding the nature of the universe and its spatio-temporal organisation. He concluded that philosophical considerations about space and time offer profound perspectives for understanding the nature of reality, human experience, and the relationship to the world around us. Philosophy plays a key role in the study of space and time, which helps to gain a deeper understanding of the characteristics of the universe and human existence.[35] It can also deal with epistemology, ethics, social philosophy, and metaphysics, revealing various aspects of the relationship with space and time.

Stefan Berger[36] studied the relationship between time, space, and philosophical ethics, which is manifested in thinking about moral aspects. He determined that spatial aspects are related to the distribution of resources, responsibility for the environment, fairness in the distribution of benefits, and temporal aspects are related to issues of long-term consequences of actions, moral evaluation of historical events and changes in moral norms. This view of the relationship between time, space, and philosophical ethics is an important one and provides an in-depth look at the influence of these aspects on the moral reasoning and decisions.[37] It highlights the importance of spatial and temporal contexts in shaping ethical beliefs and behaviour.

---

[32] Xu Zhenci et al., "Assessing Progress Towards Sustainable Development in Space and Time," *Nature* 577, no. 7788 (2020): 74-78.

[33] Oleksiy Polunin, "Modelling Explanation in the Space of Multiple Representations of the Flow of Time," *Humanities Studios: Pedagogy, Psychology, Philosophy* 13, no. 1 (2025): 83-97.

[34] Kenan Mutluer, "An Inquiry on Space and Time as a priori Forms of Intuition," *ETHOS: Dialogues in Philosophy and Social Sciences* 13, no. 2 (2020): 203-224.

[35] Zdzisław Kieliszek and Ewa Gocłowska, "The Tragedy of Ismena's Fate and Character With the "Theban Trilogy" of Sophocles as the Realization of Aristotle's Catharsis Theory," *Studia Warminskie* 56 (2019): 7-26.

[36] Stefan Berger, "History Making and Ethics – An Integral Relationship?" *History and Theory* 62, no. 1 (2023): 161-173.

[37] Zdzisław Kieliszek, "Assessment of the Rationality of Gender Studies from the Perspective of Bocheński's Concept of Philosophical Superstition," *Philosophia* 50, no. 2 (2022): 581-594.

Mehmet Bilgili[38] investigated the relationship of space and time with geography. He noticed how these concepts influence geographical processes, seasonal cycles, and climate changes. He concluded that geography studies the spatial distribution and interaction of natural and socio-cultural phenomena on Earth, while time represents historical processes, temporal changes, and consequences. Exploring the relationship between space and time in geography may also include analysing the impact of technology and transport infrastructure on the mobility and accessibility of different regions.[39] The relationship between climate change and seasonal variations with societal and economic activity in specific spatial contexts can also be examined.

Jacek Wozny[40] devoted his study to exploring the relationship between space and time with archaeology, which helps to understand the human past. He analysed the chronology of finds in the country, which can be used to reconstruct historical events, lifestyles of previous civilisations and cultural changes. Additional areas of research may include analysing archaeological finds in a culturally and ethnically sensitive manner, as well as using new technologies to more accurately date and interpret materials. It is also worth considering the impact of archaeological discoveries on the formation of national heritage identity and cultural national heritage.

Giuseppe Ritella et al.[41] presented the relationship between space, time, and pedagogy. They described how these concepts influence the organisation of educational spaces, classrooms, auditoriums, and school and university campuses. Suggested ways to optimise time in learning processes. The authors emphasised the importance of the influence of the concepts of space and time in the pedagogical process as they contribute to effective human learning. Additional attention can also be given to the study of the time frame of educational programmes and the achievement of learning objectives.

Na Zhang[42] determined that investigating the relationship of space and time to literature and art helps to better understand how these fundamental

---

[38] Mehmet Bilgili, "Approaches to the Philosophy of Space in Geography," *International Journal of Geography and Geographic Education (IGGE)* 41 (2020): 88-102.

[39] Viktor Novozhenov, "Central Asian Rock Art on the Silk Road," *Advances in Science, Technology and Innovation* 1 (2023): 129-137.

[40] Jacek Woźny, "Archeology as a Metaphor in Contemporary Culture," *Qualitative Sociology Review* 17, no. 1 (2021): 28-38.

[41] Giuseppe Ritella, Antti Rajala, and Peter Renshaw, "Using Chronotype to Research the Space-Time Relations of Learning and Education," *Learning, Culture and Social Interaction* 31 (2021): 100381.

[42] Na Zhang, "On Aesthetic Culture and Psychology in Modern Display Art Communication," in *3rd International Workshop on Art, Culture, Literature and Language*, 130-134 (Francis Academic Press, 2019).

concepts permeate the expression of culture and creativity. He noted that the breadth of space can symbolise freedom, while narrowness and limitation can symbolise captivity. Temporal aspects, on the other hand, can express life stages, ageing, momentary events, and eternal cycles. He concluded that the understanding of space and time in literature and art influences the expression of creative ideas, the formation of cultural heritage and the perception of works of art. It is also necessary to consider how the perception of space and time in works of art affects people's emotional responses. These concepts can also be used to construct a plot and predict the course of events and character development.

During the discussion of the study, options were presented on the relationship between space and time and various factors such as history, culture, identity, and modern technology. The discussion revealed the significance and influence of these concepts on human life and emphasised the need for research in this area.

V. Conclusions

The research objective of the problem of space and time in the philosophy of Kazakhstan was to comprehend, analyse and understand the philosophical aspects related to the notions of space and time in the context of Kazakh culture and philosophy. It presents historical aspects influencing the philosophy of Kazakhstan, identifies problems relevant today, and offers perspectives. The study helped to identify the unique features and characteristics of representations of space and time among Kazakh philosophers. The results of this study indicate the presence of multiple concepts related to space and time, as well as their relationship with various factors. These are culture, politics, and history. The following recommendations can be offered: to comprehensively research and analyse the problems of space and time, to actively engage in collecting and preserving unique philosophical texts, and artistic works, and to cooperate interdisciplinarily. It is important to introduce the study of these issues into educational programmes and courses to broaden the educational experience. The organisation of conferences, seminars, publication, and dissemination of results will facilitate the exchange of knowledge and encourage scholars to undertake new research.

The practical significance of the results obtained in the study lies in the enrichment of educational programmes and courses in philosophy. It is possible to integrate the knowledge of space and time of Kazakh philosophy into the world context. The results can become a starting point for the creation of cultural projects, exhibitions, and lectures. They are useful for philosophers, cultural historians, historians, and politicians. They can inspire artists, and

writers and also serve to attract tourists. Further research on the topic will help to expand knowledge about the problem of space and time in Kazakh philosophy, as well as to make discoveries and deepen understanding of philosophical ideas and the cultural heritage of Kazakhstan. This can be research on the relationship of space and time with art, religion, and analyses of ideas about them from ancient times to the present.

Author contribution statement

All authors have contributed equally to the conception and design of the work, the drafting and revising of the manuscript, and the final approval of the version to be published. All images have been compiled by the authors.

References

Auanasova, Alima, and Kamilla Auanassova. "The Struggle for Kazakh Statehood in 1917-1918 through the Prism of the History of Constitutionalism." *Investigaciones Historicas* 44 (2024): 641-661.

Auanasova, Alima, Kamilla Auanassova, and Ganizhamal Kushenova. "Alash Party and Issues of National Statehood of Kazakhstan." *Nuova Rivista Storica* 108, no. 2 (2024): 535-549.

Badejo, Omobola Olufunto. "The Persisting Problem of Persistence: A Call for an Alternative Theory." *Conatus – Journal of Philosophy* 7, no. 1 (2022): 9-31.

Balogun, Babalola Joseph. "How Not to Understand Community: A Critical Engagement with R. Bellah." *Conatus – Journal of Philosophy* 8, no. 1 (2023): 55-76.

Battalov, Askar. "Current Problems of Teaching the History of Kazakhstan in Schools." In *Materials of the International Scientific-Practical Conference "School – Teacher – Innovations in the Modern World*," edited by Zhanbol Zhilbaev, 249-254. Pavlodar Pedagogical University, 2021 [in Kazakh].

Bell, Jacob Andrew. "The Reinstatement and Ontology of Meaning." *Conatus – Journal of Philosophy* 8, no. 1 (2023): 77-86.

Berger, Stefan. "History Making and Ethics – An Integral Relationship?" *History and Theory* 62, no. 1 (2023): 161-173.

Bilgili, Mehmet. "Approaches to the Philosophy of Space in Geography." *International Journal of Geography and Geographic Education (IGGE)* 41 (2020): 88-102 [in Turkish and English].

Bloch, Walter. *Mysteries of Space and Time: Synchronicity and Nonlocality*. Crotona Verlag, 2020 [in German].

Çinkara, Gökhan. "The Geopolitical Context of Nationalism and the Transformation of Political Elites in Turkey: Memory, Identity, and Space." *Akademik Hassasiyetler/The Academic Elegance* 10, no. 21 (2023): 549-581 [in Turkish].

Dalelbekkyzy, Akmaral, and Nurlan Yildiz. "The Influence of the National Code on Historical Consciousness." *KazNU Bulletin. Series: Philological* 184, no. 4 (2021): 156-166 [in Kazakh].

Demenko, Oleksandr. "The Impact of Historical Traditions on the Foreign Policy Priorities of the Republic of Kazakhstan." In *Trends in International Relations and Problems of European Security: A Collection of Scientific Works*, edited by Serhiy Tolstov, 120-137. Institute of World History of the National Academy of Sciences of Ukraine, 2019 [in Ukrainian].

Demirel, Mehmet Volkan. *Time, Aspect and Mood/Modality in Language*. Pegem Academy, 2021 [in Turkish].

Doszhan, Raikhan. "The Problem of Existence in the Heritage of Al-Farabi and Its Continuity with Modern Scientific Knowledge." *Arts Academy. Series: Social and Human Sciences* 3, no. 7 (2023): 50-64 [in Russian].

Duisen, Seytkali, and Adiyet Kayrken. "Foreign Policy of the Republic of Kazakhstan in the Works of Domestic Researchers." *KazNU Bulletin. Series: Historical and Socio-Political Sciences* 72, no. 1 (2022): 164-176 [in Kazakh].

Efremov, Andrei. "The Psychology of Faith and Religious Identity: How Theology Shapes Worldview and Self-Perception." *Pharos Journal of Theology* 106, no. 3 (2025): 1-15.

Gotz, Lyudmila. "Distinction of the Categories 'Synchrony, Diachrony' and 'Chronotope' in the Methodology of Cultural Studies." *Almanac Culture and Contemporaneity* 2 (2021): 45-50 [in Ukrainian].

Ismailov, Alisher, and Lazzat Omarbaeva. "Spiritual Revival – Moving Force, Changing Traditions and World Ideology of the Kazakh." *Scientific Journal "Auezov University"* 4, no. 56 (2020): 188-191 [in Russian].

Jumabekova, Gulnara, Galiya Bazarbayeva, Victor Novozhenov, Elina Altynbekova, Anton Gontscharov, and Aliya Manapova. "'Sun and Steppe – Eternal Entities': A Museum Reconstruction Model with a Mace from the Trans-Tobyl Region (Northern Kazakhstan)." *Archaeology of Kazakhstan* 26, no. 4 (2024): 143-180 [in Bulgarian].

Karagun, Vedat, and Said Aras. "Globalization and the Information Age." *Dicle Akademi Dergidi/Journal of the Dicle Academy* 2, no. 1 (2022): 32-40 [in Turkish].

Kieliszek, Zdzisław, and Ewa Gocłowska. "The Tragedy of Ismena's Fate and Character With the "Theban Trilogy" of Sophocles as the Realization of Aristotle's Catharsis Theory." *Studia Warminskie* 56 (2019): 7-26.

Kieliszek, Zdzisław. "Assessment of the Rationality of Gender Studies from the Perspective of Bocheński's Concept of Philosophical Superstition." *Philosophia* 50, no. 2 (2022): 581-594.

Kulatti, Sanem. *David Hume's Critique on Infinite Divisibility of "Space" and "Time."* MA Thesis, Adnan Menderes University, 2019 [in Turkish].

Kushenova, Ganizhamal, Alima Auanasova, Albina Maxutova, and Ainur Kairullina. "Kazakh History in British Periodicals: Interpretations and Historical Accuracy." *Bylye Gody* 20, no. 1 (2025): 129-137.

Meer, Rudolf. "Alois Riehl's Theory of Space and Time in the Context of Realistic Interpretations of Kant's Transcendental Idealism." *Kant-Studien* 113, no. 3 (2022): 459-486 [in German].

Mutluer, Kenan. "An Inquiry on Space and Time as A Priori Forms of Intuition." *ETHOS: Dialogues in Philosophy and Social Sciences* 13, no. 2 (2020): 203-224 [in Turkish].

Myrzatayeva, Zabira. "Historical Memory." *KazNU Bulletin. Series "Historical and Socio-Political Sciences"* 2, no. 65 (2020) [in Kazakh].

Nedoshovenko, Tamara. "Time and Space as Tools of Cognition of the External World in the Fine Arts." *Kultura i Suchasnist: Almanakh* 1 (2020): 139-145 [in Ukrainian].

Novozhenov, Viktor. "Central Asian Rock Art on the Silk Road." *Advances in Science, Technology and Innovation* 1 (2023): 129-137.

Novozhenov, Viktor. "Vehicles in the Bronze Age Petroglyphs of Kazakhstan: Mobility and Elitism." *Archaeology of Kazakhstan* 24, no. 2 (2024): 70-99.

Ömür, Kızıl, and Cengiz Donmez. "The Model of Synchronic Approach and Scaled Synchronological Charts to Teaching Historical Subjects." *International Journal of New Approaches in Social Studies* 4, no. 2 (2020): 285-308. [in Turkish].

Ozdemir, Ibrahim, Kalmurat Subanbayev, and Ayman Keldinova. "Issues of Continuity of Kazakh Philosophy." *Al-Farabi* 78, no. 2 (2022): 72-85 [in Kazakh].

Pinho, Thiago. "Six Steps Towards an Object-Oriented Social Theory (O.O.S.T)." *Conatus – Journal of Philosophy* 8, no. 1 (2023): 263-283.

Polunin, Oleksiy. "Modelling Explanation in the Space of Multiple Representations of the Flow of Time." *Humanities Studios: Pedagogy, Psychology, Philosophy* 13, no. 1 (2025): 83-97.

Poser, Manfred. *Time and Consciousness: Why Time is an Illusion*. Crotona Verlag, 2020 [in German].

Ritella, Giuseppe, Antti Rajala, and Peter Renshaw. "Using Chronotype to Research the Space-Time Relations of Learning and Education." *Learning, Culture and Social Interaction* 31 (2021): 100381.

Sahakian, William. *History of Philosophy*. Barnes & Noble, 2022.

Sartayeva, Raushan. "Abay's Teaching 'Tolyk Adam' and Modern Tendencies in Solving the Problem of the Whole Man." *Adam Alemi* 89, no. 3 (2021): 76-91 [in Kazakh].

Shaldarbekova, Ajar. "On the Current State of National Identity in Kazakhstan: Historical Aspect." *Asian Journal "Steppe Panorama"* 6, no. 2 (2019): 328-332 [in Russian].

Strand, Michael, and Lyn Spillman. "Cultural Sociology." In *The Cambridge Handbook of Social Theory*, edited by Peter Kivisto, 43-62. Cambridge University Press, 2020.

Temirton, Galia. "The Role of National Traditions in Cultural Integration." *KazNU Bulletin. Series: "Historical and Socio-Political Sciences"* 3, no. 66 (2020): 1-7 [in Russian].

Walton, Jeremy, and Neena Mahadev. "Introduction: Religious Plurality, Interreligious Pluralism, and Spatialities of Religious Difference." *Religion and Society: Advances in Research* 10, no. 1 (2019): 81-91.

Woźny, Jacek. "Archeology as a Metaphor in Contemporary Culture." *Qualitative Sociology Review* 17, no. 1 (2021): 28-38.

Xu, Zhenci, Sophia Chau, Xiuzhi Chen, Jian Zhang, Yingjie Li, Thomas Dietz, Jinyan Wang, Julie Winkler, Fan Fan, Baorong Huang, Shuxin Li, Shaohua Wu, Anna Herzberger, Ying Tang, Dequ Hong, Yunkai Li, and Jianguo Liu. "Assessing Progress Towards Sustainable Development in Space and Time." *Nature* 577, no. 7788 (2020): 74-78.

Yilmaz, Recep. "Perception of Historical Time in the West and Turkish-Islamic Culture." *Journal of Social Sciences of Mus Alparslan University* 10, no. 2 (2022): 899-912 [in Turkish].

Zhang, Na. "On Aesthetic Culture and Psychology in Modern Display Art Communication." In *3rd International Workshop on Art, Culture, Literature and Language*, 130-134. Francis Academic Press, 2019.

discussion

J. Savulescu & P. Giannopoulou · *Conatus* 10, no. 1 (2025): 299-309
doi: https://doi.org/10.12681/cjp.39904

To Be Human is to Be Better: A Discussion with Julian Savulescu

Julian Savulescu
National University of Singapore, Singapore
E-mail address: jsavules@nus.edu.sg
ORCID iD: https://orcid.org/0000-0003-1691-6403

Phaedra Giannopoulou
National and Kapodistrian University Athens, Greece
E-mail address: faidrao2@gmail.com
ORCID iD: https://orcid.org/0000-0002-4689-8720

Abstract

In this paper, Julian Savulescu discusses humanity's trajectory – past, present, and future. As the world undergoes relentless transformation driven by technological advancements, some pressing questions arise: Is it time to provide modern solutions to old problems such as discrimination, inequality, and crime? Should people retain absolute autonomy over their decisions, even in the case that their judgment may falter? What role is Artificial Intelligence going to play in our day-to-day lives, and how far could it go? This dialogue unveils a visionary blueprint for humanity, regarding how much could really be achieved with the help of technology, what are some of the difficult decisions we would have to make, and ultimately what would it look like if we tried to use the tools we have to actually create a society that values justice and equality above individual freedom.

Keywords: *moral enhancement; discrimination; autonomy; artificial intelligence; technological enhancement; freedom; inequality*

Phaedra Giannopoulou: The topic of this year's World Bioethics Day was combating discrimination and stigmatisation. Why do you think discrimination is still such a big issue in 2024?[1] Is it just a societal problem, or do we naturally tend to have an aversion to what we consider different or strange?[2]

---

[1] Uros Prokic, "Contemporary Epistemology of Nationalism: Faltering Foundationalism Contrasted with Holistic Coherentism," *Conatus – Journal of Philosophy* 8, no. 1 (2023): 297-298.

[2] Darija Rupčić Kelam and Ivica Kelam, "Care and Empathy as a Crucial Quality for Social Change," *Conatus – Journal of Philosophy* 7, no. 2 (2022): 168.

If that's the case, will people ever be able to overcome that natural urge on their own without the help of moral enhancement?

Julian Savulescu: The problem of discrimination arises because of human nature and human moral limitations. We are essentially animals that form groups and affiliations; people are nepotistic, xenophobic and distrustful of strangers. We typically, through the course of human history, have existed in groups of 150;[3] we still haven't lost that group orientation towards our group, whether it's our football team, our sex, our race or our nation. Our political and social institutions will just be an expression of our national identity, of our nature and our ideals. While they do have some effect on how those things are expressed, we have never got to the root cause of the disease of discrimination. Despite our best efforts and rhetoric, people are still very concerned about the threat of other groups, and their concerns around immigration in Europe and the United States are expressions of these basic human tendencies.

Therefore, I think that we can try to educate people, we can create laws against discrimination, but we also need to look at the basic underlying psychology and our own psychological limitations that we all share. Potentially, in the future, there may be biological interventions or neurotechnological interventions that can augment education, but the idea that we can just tell people to be better and that "it's wrong to discriminate" is just really not tackling the problem, as you can see from around the world.

Phaedra Giannopoulou: Yes, and is it possible that if this approach was tackling the problem, then it would have to be solved by now, right? Because for so many years we've been learning about the dangers of discrimination, and we've seen throughout history where discrimination has led in the past, and yet the cycle keeps going.

Julian Savulescu: Yeah, another example is that human beings are sort of programmed to identify facial beauty as symmetry, the so-called "golden triangle of the face," that's what plastic surgeons make money from.[4] Even small babies are able to pick out an attractive face from an ugly face. So, while there might be some ideas of beauty that are cultural and temporally dependent, say, body shape to a degree, there are certain characteristics that are very hardwired and have tracked our genetic fitness and our ability to survive and reproduce. So, you can say that people should not be lookist, but

---

[3] Ingmar Persson and Julian Savulescu, "Unfit for the Future? Human Nature, Scientific Progress and the Need for Moral Enhancement," in *Enhancing Human Capacities,* eds. Julian Savulescu, Ruud ter Meulen, and Guy Kahane (Blackwell Publishing, 2011), 487.

[4] Julian Savulescu, "Genetic Interventions and the Ethics of Enhancement of Human Beings," *Gazeta de Antropología* 32, no. 2 (2016): 07.

the reality is that attractive people are more likely to have highly paid jobs, more access to romantic partners and are less likely to be found guilty of crimes. And so, this is not something that can easily be countered simply by education and social institutions, because it's very deeply ingrained.

Phaedra Giannopoulou: On the topic of moral enhancement, you have extensively discussed the thought experiment of the God machine,[5] which poses the question if it would be desirable to prevent people from acting on or even having immoral thoughts using technology. With recent developments in microchip technology, do you think scientists should be looking into whether the God machine could become a real possibility? Would that mean that morality is a more important value than freedom?

Julian Savulescu: Well, just to sort of backtrack, the argument that I proposed going back to 2008 is that, in principle, we could not only use education and social institutions to make people behave more morally, but also improve their moral dispositions and reduce their moral limitations.[6] So, for example, it's possible we could make people less xenophobic or racist, or we could make people more willing to make positive social decisions and small altruistic sacrifices with large benefits to other people. Now there is an obvious difference in moral behavior between men and women. I'm talking about genetic males and females here. There is evidence that these groups have different moral dispositions, and in general, women are more empathetic, more cooperative and so on.[7] So, what we said is that you could make men more like women, and that would be to some degree moral improvement. And one of the objections to this kind of proposition is that it would not actually be a moral enhancement because people wouldn't be free to choose to do the right thing, they would be programmed. Now I think that objection fails for many reasons. If men became more empathetic, they would be more willing to do the right thing, but it wouldn't make them less free. We don't say women are less free because they're more empathetic. If people were to achieve that goal by reading Tolstoy, nobody would object.

But when it comes to moral enhancement, the objection is that now we're restricting people's freedom. So, I have said that in general, moral enhancement wouldn't threaten freedom, but in some extreme cases it might. So, for example, if we were able to deliberately intervene in people's brains and make them stop murdering innocent people, then yes, that would remove their freedom. But still, it might be a worthwhile thing to do if we were

[5] Ingmar Persson and Julian Savulescu, "Moral Enhancement, Freedom and the God Machine," *Monist* 95, no. 3 (2012): 399-421.

[6] Ibid.

[7] Ibid.

talking about, say, the murder of an innocent child or the rape of an innocent child. And my argument was that, if the stakes were high enough, we might prioritize the lives of people over the freedom of people. So, to go back to the question, should we begin to explore this? I think that in a way we've already started to explore this; pedophiles in many countries are offered the possibility of what's called hormonal castration, which is the use of drugs to reduce their libido.[8]

Now, let's just say we were able to put a device that was able to detect whether an adult was about to engage in a sexual relationship with a minor. Let's say it was an ankle bracelet that would be able to detect that and then immobilize the individual until the police could be called or until the child could be protected. That seems to me to be a technology we should embrace, or at the very least, explore and test. It would remove the freedom to abuse children, but it would be isolated to only that particularly obviously immoral behavior. People wouldn't be immobilized for going to a protest about immigration, for example, I don't think we should be controlling that level of behavior. But the sexual abuse of children is something that we should be controlling one way or another; and if we had a God machine, we should consider employing it.

Phaedra Giannopoulou: Yes, and just in general for criminal behavior of that nature. Behavior that is forbidden by law, even now, without any restrictions on our brains and our thoughts. A lot of people feel concerned about our freedom being taken away when they hear about something like the God machine, but they ignore the fact that we are not actually free to murder or to molest children, nor should we want to be, in my opinion.

Julian Savulescu: Yes, I mean, people in jail are not free to abuse children or murder innocent people, so we do employ extreme coercion to prevent crime, because punishment is partly consequentialist as well as retributivist. It's consequentialist to protect people from being harmed in the future. So, this is just a non-biological means of prevention. Now, people's objection is that that's different because we're not intervening in people's thoughts. But in my view, if our intervention is circumscribed and appropriate, then we have to examine whether the value of the intervention is appropriate. We might find that even if we don't have complete freedom of action, it's worth it.

Phaedra Giannopoulou: If we believe that the God machine is desirable, but we don't see it being a possibility shortly, would we also support something that could produce the same results without intervening in the human brain?

---

[8] John McMillan, "The Kindest Cut? Surgical Castration, Sex Offenders and Coercive Offers," *Journal of Medical Ethics* 40, no. 9 (2014): 587-588.

For example, extreme surveillance, making sure that people act ethically in every circumstance.

Julian Savulescu: Yeah, so the God Machine is a philosophical thought experiment meant to test the idea that we must always prioritize freedom. In practice, the problem with God Machines is that they aren't perfect, just like our court system is far from perfect, and many innocent people are jailed or executed. The infringements on people's freedom could be enormous if we started to misapply this sort of technology. So, extremely heavy surveillance is an example where, if it were used just to prevent the abuse or murder of children and the murder of innocent people, and it was perfectly effective at that, then I think we should embrace that. Personally, I enjoy living in Singapore, which has a huge amount of surveillance, resultingly crime is very low, and it doesn't restrict my life. I don't have surveillance within my home, I don't feel that it excessively burdens me. But if that were multiplied to people being under constant surveillance and even for minor misdemeanors like driving through a red light when there was no traffic coming or crossing a road illegally, I think that would be very burdensome. So, while I support the principle, I think in practice, we have to be very cautious about radically enhanced surveillance. Now, there is always a balance of proportionality between infringement of freedom and benefit. In my view, countries like Singapore have that reasonable balance, but it could easily swing the wrong way.

Phaedra Giannopoulou: Is it unethical to vote for political candidates who express and support racist, homophobic, misogynistic or anti-environmentalist views and policies,[9] and if so, does that mean that moral enhancement could also affect the way people make political decisions?

Julian Savulescu: Well, my memory is that Adolf Hitler was elected, so yes, it's true that you play a part in responsibility if you vote in a leader that causes great harm, and you facilitated those crimes, in a sense, you are morally responsible. However, in modern day, political leaders have a mixture of virtues and vices; they're not uniformly evil. They may have undesirable traits, but then they also have policies that are legitimately attractive to their electorate. That's something that I think is a function of democracy; people are not perfect, and our leaders are not perfect. People ought to be free to vote for imperfect leaders. When you get all the way to Hitler, we can say that it clearly is wrong, but to extend that argument to every vote for a political leader who has some unethical views, I think is too extreme. Now, it's true that moral enhancement will affect people's political choices because it

[9] German Bula Caraballo et al., "Authoritarian Leaders as Successful Psychopaths: Towards an Understanding of the Role of Emotions in Political Decision-Making," *Conatus – Journal of Philosophy* 9, no. 2 (2024): 54.

would affect people's overall assessment of our leaders and their potential actions. I think that's a good thing, and that's what we would hope moral education would do. A part of the problem of discrimination and these kinds of vices that many international leaders display is that there's a lack of education, and in particular moral education, of the public that puts them in those positions. So, it goes back to the most basic level, we need the capacities to be able to think and in particular to morally deliberate, and we need the education to mature those capacities, both of which are important. Right now, we live in a world that is just the product of the limitations and the natural distribution of moral talents and biases and the imperfections of our educational system.

Phaedra Giannopoulou: You have claimed that for someone to make an autonomous decision, they should be fully informed and acting within reason.[10] However, different cultures often deem different things as reasonable. Would it be possible to create a universal ethical code without disregarding people's autonomy and cultural differences?

Julian Savulescu: Yeah, that's a very good question. Often people argue that ethics is relative, it's relative to culture or time or groups or individuals. I think that's a mistake because then there would be nothing to criticize about the Nazis, they just had different values to us. So, the whole movement to universal human rights has been a movement away from ethical relativism, but as you correctly point out, cultures differ in terms of their values and their reasonable values. One thing is that there isn't a very precise cardinal ordering of values that we can use to create some sort of list of human rights, how to rank them and how to apply them. How much weight you give to freedom versus security differs from China to America. It's not that one country has an answer to that. We can all agree that freedom, security, health, well-being and autonomy are all important values; but for how you instantiate those, there can be reasonable disagreement. Secondly, the circumstances of each country differ. For example, some countries are richer than others, and they can afford to have more freedom for their citizens because they can financially support those choices.

So, there will be universal values that are important to all human beings, but how they're balanced and how they manifest themselves in different cultures will vary. What we need is not a kind of 'one size fits all' set of values or rights, but a framework where those values or rights can be interpreted in a reasonable way by different countries or cultures according to the circumstances or the people. *Brothers Karamazov* is a famous Russian novel by

[10] Julian Savulescu and Evangelos D. Protopapadakis, "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu," *Conatus – Journal of Philosophy* 4, no. 1 (2019): 129.

Fyodor Dostoevsky, and there's a famous conversation between one of the brothers, Ivan, and the Grand Inquisitor, who's sort of a manifestation of the devil. The Grand Inquisitor says to Ivan, "I can make humanity very happy; it will just require that they give themselves up to me."[11] So essentially, they sacrifice their freedom for happiness. Now, freedom and happiness are both important, and just having one without the other is probably worthless. So, the question is how to balance those. And that's a question where we can defer to the autonomy of individuals, but also the autonomy of nation states to find a reasonable balance. And again, that's not to say that anything goes, but it is to say that ethics is not just black and white; it's black, white and grey, and there's a grey area where different countries will have different practices.

Phaedra Giannopoulou: Regarding the topic of physical enhancement, do you think technological advancements in that area would be helpful to combat discrimination, or would they create an environment in which you have to undergo some kind of physical enhancement to not face discrimination? For example, if a large part of society decided to use technology to make themselves faster, anyone who didn't want to do that and kept their natural speed would be at a disadvantage.

Julian Savulescu: Well, when it comes to cognitive enhancements or moral enhancements or physical enhancements, it's true that one way to correct disadvantage is to modify someone's biology. Another way is to improve the situation of that person. People who are disability activists and adopt a social constructivist model of disability say that all the disadvantages associated with disabilities exist because society is arranged in a certain way. Now I think that's too strong. I think that some of the disadvantages are socially constructed, but some are also biological, therefore, both social interventions and biological interventions could be used to provide advantages to that individual. So, you can either provide extremely good wheelchairs and wheelchair ramps and elevators and other means by which people with paraplegia can mobilize effectively, I think that's a really good point, or you can cure their paraplegia. Both of those will reduce the disadvantage and, subsequently, also the discrimination against those individuals.

But it's true also that this could create pressure to utilize whatever enhancements there are in order to maximize productivity or competitiveness. So, there is an arms race of enhancement. I think that what we need to be careful of is that those enhancements don't have downsides. When it comes to taking drugs that would improve physical performance, for example, I'm 60, and as you get older, the level of testosterone reduces. So, one physical

[11] Roger L. Cox, "Dostoevsky's Grand Inquisitor," *CrossCurrents* 17, no. 4 (1967): 431.

enhancement to maintain muscle mass is to take testosterone replacements, it is basically an enhancement for the elderly. Now, I don't take those because I'm worried about it increasing my risk of prostate cancer, but if it didn't have a risk of prostate cancer, I would. That indeed would put pressure on a lot of people to maintain their physical abilities by taking physical enhancements like testosterone. So, I think that the critical question when it comes to the arms race or the pressure that people would experience for enhancement is to ensure the enhancements are reasonably safe and don't come with adverse effects.

Phaedra Giannopoulou: If we are approaching human enhancement from the standpoint of combating discrimination, wouldn't that mean that human enhancement should only be considered ethical if it is widely available to everyone, despite their economic status? Because if it was to only become available to people who belong in higher economic classes, it would produce more discrimination rather than eliminating it.

Julian Savulescu: So, yes, one of the problems of all enhancements, not just biological enhancements, but technological enhancements, like computers or AI, is that typically they are available in capitalist societies according to the market and people's ability to afford them. So, the rich get the best computers, they get the best healthcare, and they might get the best biological enhancements. And that increases inequality, the rich get richer, the rich live longer, the rich get healthier, the rich get more and get happier. That is the way of the world in general. However, it's not determined, you could make enhancements available like we've already made enhancements available; general education is an enhancement, but we think it's so important that it's provided to everyone, and there's a basic level of education that everyone can access. My view is, if the enhancement is important enough, then we should make it available to everyone, or you could even use it to reduce inequality by only making it available to the people who are worst off. So, for example, you might make cognitive enhancements freely available or even only available to people with low to normal IQ, between say 70 and 85, who struggle to find jobs and to be productive in a technologically advanced society, which would reduce inequality. So, what impact enhancement has on inequality depends on whether it's driven by the market, driven by public funding, or targeted to correct inequality. That, of course, is up to us, given our moral limitations we probably will just make it available on the market, so I think it's likely that it will increase inequality, but it really is our choice.

Phaedra Giannopoulou: We're experiencing a meteoric evolution in AI technology. Since artificial intelligence is inherently more reasonable and more informed than a human being, does it fit the criteria that are required to make

an autonomous decision?[12] Would it be beneficial to use this technology to make decisions for humans in difficult situations?[13]

Julian Savulescu: Well, we will definitely use artificial intelligence to help us make difficult decisions in time-critical situations. The most obvious example is the programming of driverless cars, or even the programming of regular cars. People are often not able to make decisions quickly enough. Now, if we were to program a driverless vehicle to swerve off a road when a tree falls over, that could potentially kill an innocent pedestrian. In cases like these, the technology will be able to assess the situation within fractions of a second, and it will be able to make a decision. The way in which those decisions will be made will be by pre-programming our values into the AI. So, it's essentially like a Ulysses contract, this is a pre-commitment contract where you program in your values beforehand and at the time when a decision has to be made, those values then express themselves. So that will be a straightforward way in which AI will be used, in the use of technology like cars or possibly even in emergency medicine when decisions need to be made as fast as possible.

When it comes to more fundamental decisions, for example, deciding whether to have an operation or not on some area of your brain, I think side effects of AI technology could be very helpful. We've even explored the use of ethical avatars; large language models potentially trained on our own work – in my case my own academic papers – but it could be somebody's writings or their blogs to reflect that individual and also access to the body of human knowledge through the conventional large language model. So, I'll be able to talk to an enhanced version of myself or I may be able to talk to a large language model trained on Aristotle or even a large language model of Jesus, that kind of dialogue will be very enabling for making moral decisions. But I think where we can, we should make our own moral decisions because essentially, we have to take responsibility for our actions and our lives, not a machine. In situations like this there is a risk of machine paternalism, where the machine decides what's best for you. I think that would be deeply dehumanizing. What it is to be a human being, not an animal, is to make your own decisions for yourself. And AI can be used through large language models, even personalized large language models to enhance your deliberation, but it shouldn't be used to replace it. I think that would be something that would undermine our human dignity and essentially our humanity.

---

[12] Michael Anderson et al., "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson," *Conatus – Journal of Philosophy* 6, no. 1 (2021): 192.

[13] As par excellence are decisions during wartime. See Ioanna K. Lekea et al., "Exploring Enhanced Military Ethics and Legal Compliance through Automated Insights: An Experiment on Military Decision-making in Extremis," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 366; also, Nigel Biggar, "An Ethic of Military Uses of Artificial Intelligence: Sustaining Virtue, Granting Autonomy, and Calibrating Risk," *Conatus – Journal of Philosophy* 8, no. 2 (2023): 66-76.

Phaedra Giannopoulou: What do you think should be the goal of human enhancement? Should it just be a way to eliminate pain and maximize pleasure in a utilitarian sense, or could it ultimately contribute to creating lasting equality and social justice?

Julian Savulescu: To be human is to be better, we are always looking at ways of enhancing ourselves and that's a part of human nature. The deep philosophical question is what is human enhancement? What constitutes better humans? Is it humans who have better lives, more well-being? And then what is well-being? Is it just happiness in the absence of pain or is there some sort of Aristotelian account of human flourishing that we should be aiming at? Rather than just making humans who will have better lives, more well-being, should we be aiming for morally better human beings? What is that and what is the best way to do that? A challenge for ethics and human enhancement is to try to understand what is a good life and what does morality require. What I worry about is people applying very simplistic views both of well-being, for example economics just equates well-being with preference satisfaction, and also of morality, for example the current fad of wokeism that dominates the view of what a morally good outcome is.

So, I think that we need to be looking at sciences – such as psychology and neuroscience – that underpin our choices, our behavior, our talents and our abilities to achieve well-being for ourselves and to behave morally. But we also need a philosophical revolution that tries to outline what the reasonable conceptions of the good life are for human beings and for a morally better society, and we need that whether or not we're discussing bio-enhancements because our society in many cases appears to be on the verge of collapse. Collapse because of relativism, collapse because of our basic moral limitations and collapse probably because of postmodernism. You're in Greece which is short of the home and origin of much Western philosophical thought, in many ways we need to return to those origins of trying to understand what the great Greek philosophers like Plato and Aristotle were deliberating about and those issues are more urgent now than ever.

Author contribution statement

Phaedra Giannopoulou conceived and designed the study. Both authors contributed to the writing and critical revision of the manuscript to an extent clearly reflected in the content. Both authors approved the final version for submission.

References

Anderson, Michael, Susan Leigh Anderson, Alkis Gounaris, and George Kosteletos. "Towards Moral Machines: A Discussion with Michael Anderson and Susan Leigh Anderson." *Conatus – Journal of Philosophy* 6, no. 1 (2021): 177-202.

Biggar, Nigel. "An Ethic of Military Uses of Artificial Intelligence: Sustaining Virtue, Granting Autonomy, and Calibrating Risk." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 66-76.

Caraballo, German Bula, Maria Clara Garavito, and Sebastián Alejandro González. "Authoritarian Leaders as Successful Psychopaths: Towards an Understanding of the Role of Emotions in Political Decision-making." *Conatus – Journal of Philosophy* 9, no. 2 (2024): 45-74.

Cox, Roger L. "Dostoevsky's Grand Inquisitor." *CrossCurrents* 17, no. 4 (1967): 427-444.

Lekea, Ioanna K., George K. Lekeas, and Pavlos Topalnakos. "Exploring Enhanced Military Ethics and Legal Compliance through Automated Insights: An Experiment on Military Decision-making in Extremis." *Conatus – Journal of Philosophy* 8, no. 2 (2023): 345-372.

McMillan, John. "The Kindest Cut? Surgical Castration, Sex Offenders and Coercive Offers." *Journal of Medical Ethics* 40, no. 9 (2014): 583-590.

Persson, Ingmar, and Julian Savulescu. "Moral Enhancement, Freedom and the God Machine." *Monist* 95, no. 3 (2012): 399-421.

Persson, Ingmar, and Julian Savulescu. "Unfit for the Future? Human Nature, Scientific Progress, and the Need for Moral Enhancement." In *Enhancing Human Capacities,* edited by Julian Savulescu, Ruud ter Meulen, and Guy Kahane, 486-500. Blackwell Publishing, 2011.

Prokic, Uros. "Contemporary Epistemology of Nationalism: Faltering Foundationalism Contrasted with Holistic Coherentism." *Conatus – Journal of Philosophy* 8, no. 1 (2023): 285-302.

Rupčić Kelam, Darija, and Ivica Kelam. "Care and Empathy as a Crucial Quality for Social Change." *Conatus – Journal of Philosophy* 7, no. 2 (2022): 157-172.

Savulescu, Julian, and Evangelos D. Protopapadakis. "'Ethical Minefields' and the Voice of Common Sense: A Discussion with Julian Savulescu." *Conatus – Journal of Philosophy* 4, no. 1 (2019): 125-133.

Savulescu, Julian. "Genetic Interventions and the Ethics of Enhancement of Human Beings." *Gazeta de Antropología* 32, no. 2 (2016): 07.

CONATUS 10, NO. 1 WAS DESIGNED BY ACHILLEAS KLEISOURAS, TYPESETTED IN PF AND UB TYPEFACE, PRODUCED BY THE NKUA APPLIED PHILOSOPHY RESEARCH LABORATORY AND PRINTED IN 120 GR. ACID FREE PAPER. THE ELECTRONIC VERSION OF THIS ISSUE IS HOSTED BY THE NATIONAL DOCUMENTATION CENTER.

www.conatus.philosophy.uoa.gr

p-ISSN 2653-9373
e-ISSN 2459-3842