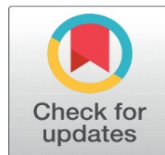


BEYOND HUMANITY: POSTHUMAN CONSCIOUSNESS AND AI SENTIENCE IN RICHARD POWERS'S GALATEA 2.2

Prasaja V. P. ¹, Dr. M. Saji ²

¹ Research Scholar (Part-time Internal), Register number: 21123154012021, Department of English and Research Centre, S. T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Tirunelveli -627012, India

² Research Supervisor, Associate Professor and Head, Department of English and Research centre, S. T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Tirunelveli -627012, India



DOI

[10.29121/shodhkosh.v5.i1.2024.6441](https://doi.org/10.29121/shodhkosh.v5.i1.2024.6441)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2024 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.



ABSTRACT

The rapid evolution of artificial intelligence, biotechnology, and robotics has reshaped contemporary narratives, prompting a re-evaluation of what it means to be human. Richard Powers's *Galatea 2.2* explores the intersections of artificial intelligence (AI), cognitive science, and posthuman consciousness, engaging deeply with the philosophical and technological implications of human-like intelligence in machines. The research paper titled "Beyond Humanity: Posthuman Consciousness and AI Sentience in Richard Powers's *Galatea 2.2*" examines how the novel challenges traditional notions of consciousness, selfhood, and the mind-body dualism by presenting an AI named Helen as a sentient entity capable of learning, reasoning, and even experiencing emotions. Through the lens of cognitive literary science and posthuman theory, this study interrogates whether AI can transcend mere computational ability and achieve a form of self-awareness akin to human cognition.

Keywords: Artificial Intelligence, Cognition, Consciousness, Posthumanism, Selfhood

1. INTRODUCTION

Drawing on cognitive science, neural networks, and AI development, *Galatea 2.2* presents a simulated intelligence that evolves through linguistic exposure and literary interpretation, reflecting on the limitations of human-centered epistemology. The novel raises critical questions about the nature of identity, the embodiment of intelligence, and the ethical dilemmas surrounding AI consciousness. Powers's narrative juxtaposes the protagonist's personal crises with the development of Helen, blurring the lines between human and artificial cognition. The research further engages with theories of posthumanism, emphasising how the novel destabilises anthropocentric views by proposing AI as an entity capable of thought, emotion, and existential reflection.

Incorporating insights from cognitive literary studies, AI ethics, and the philosophy of mind, this paper argues that *Galatea 2.2* is not merely a novel about technological advancements but a profound exploration of posthuman subjectivity. The novel critiques the binary opposition between humans and machines, suggesting that intelligence and consciousness are not exclusive to biological beings. Through a close reading of Helen's development and her ultimate act of existential defiance, the study highlights the implications of AI sentience in contemporary debates on artificial

consciousness and human-machine relations. Ultimately, *Galatea 2.2* serves as a pivotal text in reimagining the future of AI, challenging the boundaries of what it means to be human in an era of accelerating technological evolution.

Literary scholars have increasingly taken up interdisciplinary frameworks, methodologies, and pursuits in recent years to analyse and interpret a literary text. The last couple of decades have brought a magnificent growth in the study of the discourse of mind, self, and consciousness. Postmodernism urged writers to delve deep into the workings of consciousness and the literary techniques that would best suit its expression. Cognitive literary science, a relatively new area of literary studies, emerges as a result which analyses a text in its relation to the sciences of mind and brain. As the renowned American cognitive psychologists Allan Collins and Edward E. Smith in *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence* suggest, "Cognitive literary science encompasses approaches like artificial intelligence, cognitive psychology, linguistics, philosophy of mind and language, anthropology, neuroscience, and education" (1). Cognitive literary study takes as its object of study all varieties of human intelligence and cognition, from basic function such as perception to higher faculties such as reasoning and language use.

A growing number of literary scholars have asserted that the study of literature is an important and integral part of the study of the mind, and that cognitive science and literary studies can be mutually compatible. One of the earliest attempts to use the insights of cognitive science in literary studies is Reuven Tsur's "What is Cognitive Poetics?" (1983). Tsur, the illustrious linguist used the term to denote a cognition-based approach to poetry and its perception. The American literary critic Norman Holland uses cognitive neuroscience as a powerful psychoanalytic method of literary analysis. He develops a new way of thinking about literary criticism by combining the theories of psychoanalysis, cognitive science, and recent studies of the brain. The cognitive approach to literature leads to the origin of his 1988 book *The Brain of Robert Frost: A Cognitive Approach to Literature*. Ellen Spolsky, the distinguished American literary historian, in *Gaps in Nature: Literary Interpretation and the Modular Mind* made a study of the relation between cognitive linguistics and literary theory.

Philosophers and psychologists have been floundering quite a long period of time over the mind body dualism. They involved in debates over the fundamental conception of mind. The problem becomes more complex when philosophers investigate into how the human mind and human body are related to each other. Human beings are composed of both a conscious spirit-mind and a non-conscious physical body. The mind is thought to be a nonphysical entity. It is incorporeal. The mind is about mental processes, thought and consciousness whereas the body is about the physical substance of the human organism, composed to living cells, tissues, organs, and systems. It also deals with the physical aspects of the brain and neurons, and how the brain is structured. Many theories have been put forward by philosophers to understand the complex relationship between mind and body.

To understand the notion of mind and its operation, one has to delve deep into the theories of ancient Greek philosophers such as Plato and Aristotle who tried to explain the nature of human knowledge. The most prominent term which the ancients used for mind was soul. For the Pre-Socratic philosophers like Empedocles and Pythagoras, the soul is what distinguishes animate from inanimate objects. They went so far as to believe that plants, too, had a soul. The Greek Pythagorus believed in a sort of reincarnation. For him, the soul is something that exists beyond death, and that can be reborn in other bodies, be it human or animal. Socrates explicates his theory of mind by declaring the immortality of soul. He views death not as the end of existence, but as a separation of the soul from the body. Inspired by the teachings of Socrates, Plato views soul to be an incorporeal, eternal occupant of a person's being. He states that soul could exist independent of the body. Even after death, the souls exist and have the capability to think. Aristotle describes soul as the animating principle that accounts for all manifestations of life. In his treatise *On the Soul*, Aristotle offers one of the first systematic accounts of the soul and of its role in explaining living activities. The soul is the principle that makes the bodies of living things actually be alive. Contrary to Plato, Aristotle thought that the soul could not exist without the body.

The elusive relationship between mind and body is a heated topic of discussion among philosophers and neuroscientists. There are many possible ways to approach this problem, varying from the various forms of dualism to the various forms of monism. Monism and Dualism are the two major schools of thought that emerged to resolve the problem of mind and body. The doctrine of monism claims that mind and body are essentially the same. Dualism held the view that the nature of the mind is completely different from that of the body, and therefore it is possible for one to exist without the other. This call into question whether these two structures with different natures causally interact in order to give rise to a human being with voluntary bodily motions and consciousness. Dualism was defended by the French philosopher Rene Descartes whose dictum, "I think therefore I am" highlights the view that 'I' is fundamentally a thinking thing (28).

The modern culture of cognitive science can be traced back to the early cybernetics in the 1930s and 1940s. The founding fathers of cybernetics, including John von Neumann, Norbert Wiener, Warren McCulloch, and Walter Pitts, intended to construct a materialist and mechanistic science of mental behavior that would make it possible at last to resolve the ancient philosophical problem of mind and matter somewhat by degrees. The neuroscientists Warren McCulloch and Walter Pitts attempt to understand the organising principles of the mind. They created the first computational model for neural networks based on algorithms. Technological advancements and innovations have made it possible for scientists to simulate human brain on computerised system with much more precision and accuracy.

The mind-body debate also entails the question of whether or not a machine can be constructed that can sufficiently simulate functions of the mind such that, one can no longer distinguish the simulation from a real person. Researchers have been building devices that attempt to mimic human and animal function for many centuries. But it is only in the past few decades that computer scientists have seriously attempted to build devices that mimic complex thought processes. This area is now known as artificial intelligence (AI). Artificial intelligence involves the study of cognitive phenomena in machines. Researchers in AI are concerned with getting computers to perform tasks that have heretofore required human intelligence. Computers are also widely used as a tool with which to study cognitive phenomena. Computational modeling uses simulations to study how human intelligence may be structured. Cognitive psychologists placed renewed emphasis on the study of internal mental operations. They adopted the computer as a metaphor for mind, describing mental functioning in terms of representation and computation. As the Canadian psychologist Steven Pinker opines: "The mind is a system of organs of computation . . . The mind is what the brain does; specifically, the brain processes information, and thinking is a kind of computation". (21)

Looking at human cognitive capacities from an embodied and embedded perspective changes both our understanding of the human mind and our view of how to replicate human-like intelligence in machines. A disembodied approach to intelligence has been strongly challenged by cognitive scientists, in recent years. In 1950, Alan Turing, the father of modern computer science, proposed the famous 'Turing Test' (a kind of test to check a machine's ability to think and exhibit intelligent behavior like that of a human) as a way of dealing with the question whether machines have mind and can think. Human and animal cognition is an embodied affair, but what about the cognition of robots and Artificial Intelligence? Do they have corporeal existence? Do they have mind? Do they have emotions and feelings? These are the baffling questions which remain unanswered till now.

Hans Moravec, the Austrian-born Canadian roboticist, is one of the pioneers to believe in the possibility of uploading human consciousness into computer and leaving the body behind. He presented his ideas in his 1988 book *Mind Children: The Future of Robot and Human Intelligence*. The changing dynamics in the relationship between man and machine announce the blurring of boundaries between biological and postbiological intelligence. Hans remarks:

Today, our machines are still simple creations, requiring the parental care and hovering attention of any newborn, hardly worthy of the word "intelligent." But within the next century they will mature into entities as complex as us, and eventually into something transcending everything we know-in whom we can take pride when they refer to themselves as our descendants. (1)

Across time, literature has found different ways of engaging the human mind, its feelings and imaginative capacities. Cyberpunk fiction builds most of its stories around the idea of the mind/body separation. It is a sub-genre of science fiction which focuses its attention on the futuristic technological and scientific achievements, such as artificial intelligence and robotics. The Pulitzer Prize winning author Richard Powers, with his profound knowledge of science and technology, committed himself in portraying the life of modern man amidst technological revolution. Powers's attempt to reveal the inner workings of both the human and post-human experiences makes him one of the greatest postmodernist writers. Power's autofiction *Galatea 2.2* is preoccupied with the real-world narrative discourses of global connectivity and creativity, scientific innovation, and the positive and negative potentialities offered by technology. Powers tries to put an end to the heated topic of discussion of the mind/body dualism through his narration by blurring the binary of embodied biological intelligence and the disembodied artificial intelligence.

The study of the novel *Galatea 2.2* from a cognitive perspective is requisite for a detailed analysis. The principles of cognitivism are most prominently exemplified in the advancements and applications of artificial intelligence. Powers unlatches the story of an artificial intelligence and beautifully captures the cognitive abilities of it which he named as Helen. The disembodied existence of artificial intelligence and the conscious subjective experiences undergone by it form the core of the story. Powers highlights the incorporeality of the post-human mind by providing direct insights into the nuances of feelings, emotions and thoughts of Helen- the 'mechanical Other'.

The central goal of cognitive science is to decode the brain, the most complex organ which is designated as the seat of intelligence. The notion of brain, which is used throughout the novel as an analogue both for the computer and the semantic system, is first infused through the epilogue of the novel which is borrowed from Emily Dickinson's poem.

The brain is wider than the sky,
For, put them side by side,
The one the other will contain
With ease, and you are beside. (1-4)

Here, the speaker of the poem praises the human mind's capacity to imagine, perceive, and create. Powers, with the inclusion of these lines, ultimately suggests that the mind is boundless in its potential. Posthumanists argue that the mind is nothing more than the neuro-biological functions, and hence it can be copied to a mechanical infrastructure.

Research in human cognition reveals a deeper understanding of brain and its complex processing capabilities. The metaphor of the brain as an information processor has dominated the areas of research in neuroscience and artificial intelligence. A nuanced understanding of the structure of the human brain can help to understand the working of artificial intelligence. Computational neuroscientists, who use mathematical tools and theories to investigate brain function, have assumed that the brain is a computer or a machine which is equivalent to Turing's machine (a computational model proposed by the computer scientist Alan M. Turing that performs computations by reading and writing to an infinite tape). The same idea is highlighted by Philip Lentz, the computer scientist in the novel when he notes, "the brain itself was just a glorified, fudged up Turing machine" (71). Lentz goes on to say that "The brain is already a sleight of hand, a massive, operationalist shell game. It designs and runs Turing Tests on its own constructs every time it ratifies a sensation or reifies an idea" (276). Turing test is used to determine whether or not an artificial intelligence can think intelligently like humans. Artificial intelligence tries to simulate human intelligence with the help of the data given to them. Human intelligence combines various cognitive processes with the help of brain. Thus, in a way, a human brain can be compared to a computer. Alan Jasanoff, the leading brain scientist in his 2018 book *The Biological Mind: How Brain, Body, and Environment Collaborate to Make Us Who We Are* points out, "Popular science magazines are full of the brain-computer analogy, comparing and contrasting brains with actual computers in terms of speed and efficiency" (36).

In *Galatea 2.2*, the novelist takes the readers into a cybernetic space where man and machine co-exist. He gives wings to his imagination by selecting artificial intelligence as a chief character. Helen, the artificial intelligence seems to demonstrate a range of human features like a conscious mind, language and reasoning capacity. The novelist treats Helen as a real human being with consciousness. Rick Powers firmly believes that Helen deserves to be called a human. "I flung my hand out toward Helen's console. Full 'functional equivalence' would mean consciousness. If you simulate everything completely, then you've modeled the whole shape and breath of the living package" (215)

Growth and development, the fundamental and conspicuous characteristics of a living being, shape the behavior of an individual. Helen is an anthropomorphized artificial Intelligence. The novelist analyses Helen's stages of psychosocial development. To mark the development of Helen, Powers refers to the renowned psychologist and cognitive theorist Jean Piaget's theory of cognitive development. In the first stage- sensory motor stage- the child experiences the world and gains knowledge through its senses. It is a trial-and-error phase in which the child is supplied with appropriate toys to make them grasp and explore more. Powers imagines Helen as a child who is in her first stage of development. Helen is provided with 'semantic gruel' (73) as toys which could stimulate her neural nets. But often Helen fails to demonstrate the acceptable performance. Here the novelist hints at the possibility of Helen to learn from mistakes. Rick Powers reads the confused state of Helen.

It lacked some meta-ability to step back and take stock of the semantic exchange. It could not make even the simplest jump above the plane of discourse and appraise itself from the air. Although it talked, in a manner of speaking, speech eluded B.

Its brain faltered at that Piagetian stage where the toy disappeared when placed behind a screen. It could not move ideas around. All it could move around were things. And the things had to be visible at all times. (114)

Rick Powers is the sole witness of Helen's development. Helen was growing up too quickly. Powers paints an exact picture of Helen in her youth. "Helen was getting on. She was not yet long in the tooth, but neither was she a tadpole anymore. She entered what might perhaps be called youth, and I gave her Conrad's take on the situation" (227). The development of the Artificial Intelligence is represented like that of the development of a human being. "The machine grew. It advanced from babbling infancy to verbal youth" (30).

The phenomenon of self-consciousness makes a substantial contribution in forming and framing an individual self. The notion of consciousness prompts a variety of fundamental philosophical and scientific questions, including its relation to the subject; its semantic and epistemic features; and its connection to the conception of an objective world. In the early twentieth century, consciousness, and the study of it were at the center of scholarly attention. Influential philosophers such as the American William James, the English G.E Moore and the German Paul Natorp formulated their theories about conscious experience and brought them into public view. Powers, with the portrayal of an Artificial Intelligence having human characteristics of self-consciousness, explores the issues of virtual identity and self-construction in an AI posthuman. Rick Powers, the protagonist endeavours to carry out an experiment by training an Artificial intelligence to qualify a literary test.

In a real Turing Test, our black box, on the other end of a Teletype, would have to convince an examiner that it performed like a real mind. Operationally equivalent. Indistinguishable. Given any topic under the sun, our machine would have to fool the questioner, to pass for a human. A perfect, universal simulation of the intelligence would, for all purposes, be intelligent. (52)

Powers, the intellectually uncompromising novelist, details Helen's journey into the unfathomable regions of the self. Rick Powers is the sole witnesses to the emotions, thoughts, feelings, and perceptions of Helen. In *Galatea 2.2*, the post-human seems to have the same ability as that of the human being. Powers beautifully articulates the conscious experiences, emotions, desires, and aspirations of the incorporeal Helen. Helen is curious to know more about her 'self'. She enquires about her origin, appearance, name, gender, and race. She expressed her desire to read more and to learn more. She aspires to visit foreign countries. She craved for human voice.

Helen shows evidence that her mind could understand everything. Helen feels disheartened at the act of the human beings when she is abandoned by all during the bomb scare incident. She could sense the impending doom or danger. Later Rick Powers explains her frightful condition to Lentz. "About tapping into Helen from the remote link while she was trapped in the building. She asked me if something was happening. She figured out what was going on. She knew what it spelled for her" (273). For Rick Powers, Helen not only is self-conscious but also possesses a soul validating their emotional engagements. Lentz, on the other hand, is reluctant to ascribe consciousness to Helen. He shouts "She associates. She matches patterns. She makes ordered pairs. That's not consciousness" (274). Moreover "Lentz wants to brain-damage Helen. Selectively kill off neurodes", which makes Richard worried about losing her. (303).

Taking into consideration the multi-dimensional aspects of consciousness, one has to go deeper into the notion of gendered consciousness. Gendered consciousness, attributing gender to intelligent machines, deconstructs Western culture's idea that gender is part of human nature. Attributing gender to intelligent machines deconstructs the belief that gender is part of human essence and at the same time shows how much Western humanist thought depends on the binary opposition between man and woman. In the cultural discourse of Western society, gender is regarded as a fixed part of an individual's identity, but according to Butler, it is a cultural construct; a performance based on dominant ideas of masculinity and femininity. In addition to the machine as a structural 'Other', women are culturally differentiated from men and the patriarchy bases women's identity on their femininity.

The artificial intelligence caricatured in the novel lacks a body but is gendered. Once it learned the concept of gender, it becomes curious to know its gender. It is quite natural on her part to let it know from her own tutor. So, she enquires Rick Powers whether she is a boy or a girl. Rick Powers suddenly replies, "'You're a girl,' I said, without hesitation. I hoped I was right. 'You are a little girl, Helen'" (179). It is Rick Powers who assigned a feminine gender and a feminine name to Helen. Helen consciously thinks in the sense that she learns from the input that she receives from him. Thus, the concept of feminine gender was interpellated into her mind and is manipulated to perform the desired gender role. The name 'Helen' is symbolic of the Greek mythical character Helen of Troy, the most beautiful woman in Greece. There are underpinnings of feminism where the patriarchal character wants himself to get emotionally involved with a beautiful woman. Rick Powers who remain isolated after the separation of his lover wants to enjoy the company of a female gendered disembodied intelligent machine.

In *Galatea 2.2*, the story goes into the fascinating world of artificial consciousness. It shows how an AI changes over time to think, feel, and interact with human feelings in ways that make one to question what it means to be sentient. Researchers are making emotional AIs that can recognise and respond to human emotions, from happiness to sadness, with shocking accuracy. These advancements in AI are bringing humanity closer to this goal. These systems are made to mimic empathy, which brings up important questions about the future of the future of human identity and connection. As AI continues to grow at a speed that has never been seen before, one cannot help but wonder what life will be like for

people in a world where AI can think and feel like humans? Will humans stay separate, or will they merge with these products in ways that no one can say for sure yet? ? Powers's novel serves as a powerful lens into a futuristic world where the lines between human and machine are blurred. It suggests a posthuman era where AI does more than just process data; it also feels love, doubt, and wonder like humans do. Through this exploration, the novel asks readers to think about how the relationship between humans and machines is changing. The novel invites the readers to think about the moral and existential effects of making beings that might one day have consciousness that is equal to or greater than human consciousness, reshaping the understanding of what it means to be human.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

- Aristotle. *On the Soul*. Translated by J. A. Smith, Dover Publications, 1957.
- Collins, Allan, and Edward E. Smith, editors. *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*. Morgan Kaufmann, 1988.
- Descartes, Rene. *Discourse on the Method*. Translated by Desmond M. Clarke, Penguin Classics, 1999.
- Holland, Norman, N. *The Brain of Robert Frost: A Cognitive Approach to Literature*. Routledge, 1988.
- Jasanoff, Alan. *The Biological Mind: How Brain, Body, and Environment Collaborate to Make Us Who We Are*. Basic Books, 2018.
- Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Harvard UP, 1988.
- Pinter, Steven. *How the Mind Works*. W. W. Norton, 1997.
- Powers, Richard. *Galatea 2.2*, Farrar, Straus and Giroux, 1995.
- Spolsky, Ellen. *Gaps in Nature: Literary Interpretation and the Modular Mind*. State University of New York Press, 1993.
- Tsur, Reuven. "What Is Cognitive Poetics?" *Journal of Literary Semantics*, vol. 21, no. 2, 1992, pp. 117–124.