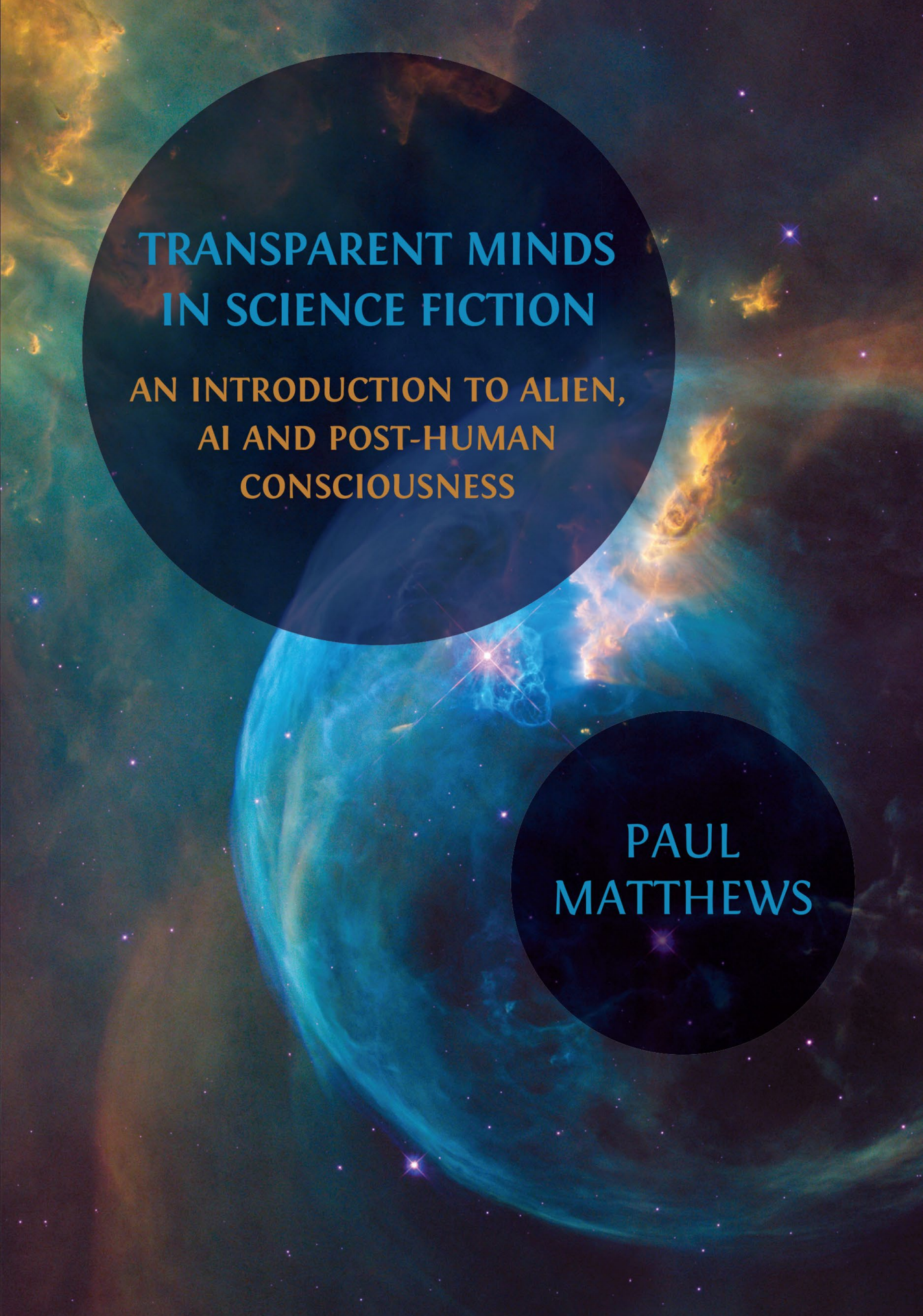
TRANSPARENT MINDS
IN SCIENCE FICTION
AN INTRODUCTION TO ALIEN,
AI AND POST-HUMAN
CONSCIOUSNESS
PAUL
MATTHEWS

https://www.openbookpublishers.com

ISBN Paperback: 978-1-80511-046-0
ISBN Hardback: 978-1-80511-047-7
ISBN Digital (PDF): 978-1-80511-048-4
ISBN Digital ebook (epub): 978-1-80511-049-1
ISBN XML: 978-1-80511-051-4
ISBN HTML: 978-1-80511-052-1
DOI: 10.11647/OBP.0348

Cover image: NASA, Nebula, May 4, 2016. https://unsplash.com/photos/rTZW4f02zY8
Cover design: Jeevanjot Kaur Nagpal

1. Introduction

> Literature is a record of human consciousness, the richest and most comprehensive we have... The novel is arguably (humanity's) most successful effort to describe the experience of individual human beings moving through space and time.
>
> David Lodge[1]

It should be more widely appreciated that literature is a kind of scientific tool that can be used to shed light on consciousness. The argument is that the richest description of the phenomenon of human experience come from our finest writers, who are capable of capturing moments in time in exquisite detail from multiple perspectives. In this view, there is no need to argue that either science or the humanities represent true knowledge, but that the two can complement one another productively, the latter unveiling a personalised view that science—with its third person fixation—cannot achieve.[2]

Science Fiction (SF) authors are often well versed in neuroscience and philosophy and see their literature as thought experiments or models of problems on far off, but feasible, technological horizons. As Swirsky notes, SF is a model in multidisciplinary alignment: 'literature, philosophy and science are, in my opinion, inseparable manifestations of the same creative instinct that has operated throughout the ages.'[3]

1 David Lodge, *Consciousness and the Novel: Connected Essays* (Cambridge, MA: Harvard University Press, 2002), 10.

2 Philosophy has also been criticised for being limited in the same way. As Simon Glendinning points out, it is too coldly logical and argument-centred. He sees a clear role for imaginative literature that presents a compelling expression of mental states. Simon Glendinning, *In the Name of Phenomenology* (London: Routledge, 2007).

3 Peter Swirski, *Between Literature and Science: Poe, Lem, and Explorations in Aesthetics, Cognitive Science, and Literary Knowledge* (Montreal: McGill-Queen's University Press, 2000), 139.

 https://doi.org/10.11647/OBP.0348.01

My thesis is that science fiction with a psycho-emotional flavour can provide new insight into both current human consciousness and also possible future states of consciousness in both ourselves and the machines we create. The reciprocal move is that inspiration for the artistic portrayal of these states can come from the science of our own world, which is still only gradually revealing causes for sentience in humans and other animals.

In the following pages, we will see examples of how SF authors approach this multidisciplinary alignment in their works. It is first worth introducing the narrative and literary theories that can help to explain both the authors' art and the readers' experience.

Psychonarration and monologue: backgrounding the author

Literature that includes characters' mental states has been around since as far back as the eighteenth century but reached a peak of sophistication in the early twentieth. The use of techniques such as psychonarration and monologue—where the viewpoint enters the characters' minds—powered the modernist approach to literature and enabled highly psychological writing. Authors such as Proust, Austen, Joyce and Woolf have been celebrated for their powerful insights into the experience of memory, emotion, motivation and consciousness. A large part of their innovation and craft was to move from the classic (external) narrative voice which could be intrusive and domineering, to ways of telling a story where the writer merges with the character so closely that they almost disappear. As Nelles suggests in his study of Jane Austen, in this way the novelist 'creates the impression that the reader has more or less direct access to the character's mind rather than through a firewall of narrative commentary.'[4]

In her classic analysis *Transparent Minds*, Dorrit Cohn groups these narrative techniques into third and first-person types and notes how some provide more authentic impressions of mind-reading, with third person psychonarration (incorporating the method of 'free indirect

4 William Nelles, 'Austen's Juvenilia and the Sciences of the Mind', in *Jane Austen and Sciences of the Mind*, ed. Beth Lau, 14–36 (London: Routledge, 2017), 18.

speech') the most effective followed by monologue. In these modes, dialogue is no longer visible by way of well-defined quotation marks, but mixes with the narrative. The author's own opinions or moralising can be reduced or removed entirely, with the result that we can more readily believe we are observing the accounts of individual personalities directly. Cohn explains how 'narrative fiction is the only literary genre, as well as the only kind of narrative, in which the unspoken thoughts, feelings and perceptions of a character other than the speaker can be portrayed.'[5]

Cohn notes the enormous challenge of trying to project into other minds. No wonder then, that attempting the same for non-human minds is something that relatively few SF authors have attempted.

Neuroaesthetics: fiction and the feeling of consciousness

When analysts have examined how these great writers achieve their effects, they have been able to link fictionally depicted states to scientific knowledge of those states. This approach, which seeks to combine literary criticism with relevant brain science, has been termed 'neuroaesthetics' by Kay Young. It is enacted via Young's principle of resonance, the way in which a book's language triggers the experience within with the reader: 'the principle of listening and looking for compelling resonances in language, meaning and representation (yielding) shared accounts of integrated mind.'[6] Young sees it as particularly important to maintain the emotional and embodied experience of reading when studying and experiencing texts from authors who write about mental states.

This reader-author dynamic is one way in which literature may provide more fertile imaginary ground than science. Philip Davis puts it thus: 'what literature does, which formal philosophy does not—and what literature can hardly help doing—is yield more than its writers know... writers offer this by creating not so much a line of argument

5 Dorrit Cohn, *Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction* (Princeton, NJ: Princeton University Press, 1978), 7.

6 Kay Young, *Imagining Minds: The Neuro-Aesthetics of Austen, Eliot, and Hardy* (Columbus: Ohio State University Press, 2010), 10.

as a resonant space for thinking.'[7] So literature itself may be a cognitive extension.[8] Reading literature allows for first person, active insight and has a moral dimension—by making us feel directly addressed and immersed, it forces us to be involved. Famed science fiction author Ursula Le Guin is skilled at weaving together scientific knowledge with philosophical and ethical speculation using story:

> Science describes accurately from outside, poetry describes accurately from inside. Science explicates, poetry implicates. Both celebrate what they describe. We need the languages of both science and poetry to save us from merely stockpiling endless 'information' that fails to inform our ignorance or our irresponsibility.[9]

Here, Le Guin identifies the missed opportunity of the solely objective—the chance for readers to recognise themselves in the process of reading and to access their imaginations.

Consciousness from the outside and inside

Most approaches to consciousness in science and philosophy certainly *have* been on the objective side, seeing subjectivity as either unscientific, or a difficult-to-explain emergent property of the system. A leading paradigm in neuroscience-oriented theories of consciousness is not that it is fundamentally a different substance, but that it is explainable in terms of the neural material from which it arises (materialism). Several promising explanatory directions exist within materialism and include the need to have a unified, generally available picture of the current state (global workspace), the brain's capacity to combine multiple signals (integrated information), a spatially organised internal representation of the body (self-modelling) and the assumption of different levels of symbolic representation (with consciousness as a higher level, localised representation).

---

7 Philip Davis, *Reading and the Reader: The Literary Agenda* (Oxford: University Press, 2013), 4.

8 Davis notes neurological evidence that novel metaphors lead to more brain activity than do clichés. He goes on to quote Joseph Gold on the cognitive power of literature: 'Literature itself has become for both writers and readers 'a brain extension' which has added a new level of consciousness to human brains 'because by means of writing and reading the brain could feedback thought to itself'', Ibid.

9 Ursula Le Guin, *Late in the Day: Poems 2010–2014* (Oakland, CA: PM Press, 2016), 7.

Some key principles are shared by these leading theories: the existence of the widely interconnected, yet modularised neuronal pathways in the different parts of the brain; the extent of internal feedback from signals to achieve changes in the external and internal world; and the overall complexity of the organisation of the brain itself. These factors combined may eventually yield a reliable test or universally accepted standard of consciousness.[10]

While materialist theories can perform well in explaining experimental findings and human performance and may all have elements of the overall picture, they remain unsatisfying in certain ways. The traditional stance of science and philosophy, exemplified by materialism, is external objectivity, yet this usually ignores or cannot deal with the view from within. A second movement, that of phenomenology, attempts to use conscious experience and structure as a method to access truth about the nature of the world. Phenomenology tries to bring science's rigour to the study of subjectivity, with a number of flavours differently emphasising embodiment and perception. This often led theorists to invent new language to describe the sense-world interface, such as Heidegger's *Dasein* (a conscious being) and the state of 'being-in-the-world' (unifying subject, object and consciousness).[11] This can make them relatively very hard to understand and they have been accused of not adding much of practical value. But phenomenology as a method—the deliberate and structured attention to, and description of, the lived experience—can add practically to our understanding of different people's viewpoints. It also fundamentally describes the work of authors when they are building fictional mental states.

10 Indeed, Christof Koch already proposes the 'zap and zip' test of consciousness in which a magnetic pulse is applied to the brain and the waves of ensuing perturbation are measured by an EEG. The complexity of the measured response boils down to a single 'zipped' number. The zap and zip response takes advantage of the same algorithm used for file compression, dubbed the perturbational complexity index (PCI). The PCI seems to represent consciousness levels quite well, both in human wakeful states and in people who are asleep or comatose. *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed* (Cambridge, MA: The MIT Press, 2019).

11 Michael Wheeler, 2020, 'Martin Heidegger', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/heidegger/.

> Seeking the essence of consciousness... will consist in rediscovering my actual presence to myself, the fact of my consciousness... Looking for the world's essence is not looking for what it is as an idea once it has been reduced to a theme of discourse; it is looking for what it is as a fact to us, before any thematization.[12]

Phenomenology as an approach had a direct applicability to literature and the subjective insight brought about by reading. For Wolfgang Iser, the 'blanks' or 'gaps' in more difficult or experimental fictional texts are what gives rise to a deeper experience of the readers' own consciousness and a purer experience of the world: 'It gives rise to a mode of communication through which the openness of the world is transferred in its very openness into the reader's conscious mind.'[13]

In the case of consciousness, science's 'view from nowhere' may have been a real limiting factor in the development of theory. Leading neuroscientists with a materialist orientation—such as Koch—have admitted this, saying all we can really start with is our own conscious experience and progress from there. This admission of the unavoidable need for some subjectivity feels like progress, as does the admission that the kind of experience we have cannot be solely claimed by the human species.[14] Koch has been vocal in attributing forms of consciousness to both lower and higher animals and the approach to an explanation of consciousness that he espouses—integrated information theory—can be applied to silicon brains and even some inanimate objects as much as human grey matter. This kind of thinking is taken to a logical conclusion in the theory of panpsychism (not widely accepted!), that everything in the world may have some level of consciousness.[15]

---

12 Maurice Merleau-Ponty, *The Phenomenology of Perception*, ed. Thomas Baldwin (London: Routledge, 2003), 72.

13 Wolfgang Iser, *The Act of Reading: A Theory of Aesthetic Response* (Baltimore: The Johns Hopkins University Press, 1978), 211.

14 Christof Koch, *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed* (Cambridge, MA: The MIT Press, 2019), 26.

15 Philip Goff, William Seager, and Sean Allen-Hermanson, 'Panpsychism', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2022. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2022/entries/panpsychism/.

Approaching the alien

While it is certainly difficult enough to describe our own conscious experience, how much more difficult it is to imagine and describe this kind of experience in non-human minds? This further removal needs an enormous amount of creativity and is to some extent constrained by language. In 'What Is It Like to Be a Bat?', the philosopher Thomas Nagel concludes that bats probably have a form of consciousness, but that it is likely to elude our ability to understand and explain it:

> Bats, although more closely related to us than those other species, nevertheless present a range of activity and a sensory apparatus so different from ours that the problem I want to pose is exceptionally vivid... anyone who has spent some time in an enclosed space with an excited bat knows what it is to encounter a fundamentally alien form of life.[16]

Nagel goes on to compare the alienness of other animal species to 'actual' aliens: 'And if there is conscious life elsewhere in the universe, it is likely that some of it will not be describable even in the most general experiential terms available to us.'[17] Generally speaking, for Nagel, the inner life of alien minds is even harder to approach than that of fellow animals with some shared sensory world: 'The more different from oneself the other experiencer is, the less success one can expect with this enterprise.'[18]

As we will see later on, science fiction works such as those of Stanislaw Lem depict alien intelligence so different from human intelligence that communication—at least in the way we understand it—is impossible. Any theory of mind that can be formed by the humans in these stories is depicted as being embarrassingly inadequate.

Nagel's thought experiment is bolstered by the concept of *umwelt*, the idea that the subjective universe of humans and other animals must vary enormously due to their different sensory apparatus, cognitive limitations and motivations.

16 Thomas Nagel 'What Is It Like to Be a Bat?', *The Philosophical Review* 83, no. 4 (1974): 438. https://doi.org/10.2307/2183914.
17 Ibid., 439.
18 Ibid., 442.

Humans often mistakenly assume that the *umwelt* they experience reflects the totality of the objective universe, the 'world out there', whereas in reality, the human *umwelt* only captures a small portion of what may be objectively (and physically) experienced by another species.

Although Nagel seems fairly pessimistic about this problem being resolved logically, he does admit that imagination is one way that it can be approached. This leap of imagination might not address the felt need to describe consciousness in objective terms, but it can take a step closer to what might be the imagined subjective world of different beings.

And we may not need to solely rely on the abstract. People *can* learn to echolocate by using sound to detect objects, albeit in a less sophisticated way than bats. Some blind people can use this to navigate, and may become so proficient that they describe the experience as analogous to *seeing* the world. This illustrates the brain's remarkable plasticity in adapting to new kinds of input, when paired with a data stream can be trained by trial and error to become effectively new senses (and similar brain areas to those that deal with vision, hearing etc. are adapted to this purpose).

The genre of Science Fiction

History and Subgenres

The literary theorists quoted at the beginning of this introduction, who propound a cognitive approach to criticism, tend to concentrate their focus on established 'high' literature to illustrate examples of how great writing can invoke mental states. I want to bring the same approach using science fiction sources, though in doing so need to admit that the genre has not always had a reputation for producing great works approaching the type of critical acclaim traditionally reserved for literary canon. This has in part been attributed to the plethora of amateur and hobbyist writers who entered the genre in the middle to late twentieth century, fuelled by the ease of publishing in SF periodicals and anthologies and the general demand amongst the reading public for man-meets-alien fantasy adventures. But I would argue that, amongst this work and well into the present day, there really is a rich vein of highly innovative, thought provoking and powerful writing. Here I will briefly outline the history, diversity and methodology of the SF genre.

SF has been famously defined by Darko Suvin as 'a literary genre whose necessary and sufficient conditions are the presence and interaction of estrangement and cognition, and whose main device is an imaginative framework alternative to the author's empirical environment.'[19]

With its origins in Gothic fiction and mystery of the nineteenth century and earlier, science fiction was spurred by the rapid technological development and existential threat of war in the early part of the twentieth century, which saw classics such as HG Wells' *The War of the Worlds*. In addition, with the development of space exploration in the middle part of the previous century, an explosion of speculation occurred around what interplanetary and interstellar exploration might bring and what other life might exist in the universe. The optimism and exploratory zeal of these 'golden age' periods in part echoed the West's attitude to the concept of empire—that foreign lands exist to be discovered, claimed, and the natives quelled. Thereafter, coincident with the gradual contraction of empire and a youth rebellion, science fiction's 'new wave' of the 1960's often took an inward turn, focusing rather more on the unwelcome invasion of mother Earth by alien races intent on our extinction.[20] The end result of this move was SF's move to the psychological, with more stories playing out as much on the mental as the physical plane. Today, the genre is mature and widely varied, with many subgenres and genre mash-ups. Globalisation has led to more cultural cross-pollination, and the diversity of authors has increased markedly. As we will see, the increasing possibility of an AI-dominant future led to the growing theme of the singularity, or superhuman intelligences. More recently too, climate change and biodiversity loss have led to the emergence of Solar Punk where alternative futures, both pessimistic and optimistic, are imagined.

In one widely used, though far from neat subdivision of the science fiction genre, there is a spectrum from hard SF at one end, where the science is foregrounded, central to the story, reality-based and well

19 Darko Suvin, *Metamorphoses in Science Fiction: On the Poetics and History of a Literary Genre* (New Haven, CT: Yale University Press, 1979), 27.

20 Damian Broderick, 'New Wave and Backwash: 1960–1980', in *The Cambridge Companion to Science Fiction*, eds Edward James and Farah Mendlesohn (Cambridge: Cambridge University Press, 2003), 48.

researched, through space opera, where the stories are more formulaic, melodramatic adventures set in other worlds, to fantasy at the other end, where there is less regard to possibility.[21] Of course, this kind of over-generalisation masks the huge variety of approaches and styles. Many readers and publishers take an inclusive approach and these days, as the label SF (constantly in an apparent crisis of self-definition) is often applied to works which, on the surface at least, have a very different focus or setting than the more familiar space adventure. Such speculative fiction, as some prefer to call it, no longer fits on a simplistic scale of hard to soft, and may be more willing to introduce the fantastical.

Methodology

Indeed, some of the more powerful work achieves its effect through highly familiar stories and settings with only small, or gradually introduced weirdness. Successful fantasy, according to Forster, does not need to fully embrace the supernatural, but can merely suggest it or imply it through realistic sounding events where the supernatural seems to be lurking on the margins. Technically it often 'merges the kingdoms of magic and common sense by using words that apply to both, and the mixture (thus) created comes alive.'[22]

A shared technical vocabulary between author and reader certainly helps SF to achieve literary impact. Assuming scientific awareness amongst readers provides authors with the ability to refer to contemporary terms and theories and to also—through the knowledge that the scientific paradigm will often shift—extend the contemporary into possible, unrecognisable futures. This recognition of the ephemerality of scientific knowledge at any one time may be a root of science fiction's frequent attraction toward either optimistic or pessimistic futures. This constant churn of scientific theory has been dubbed the 'pessimistic induction' by philosophers of science. The argument is that when our theories about the world have changed so radically over the generations, they can't be accepted as wholly true at any one time.

21 This hard SF-fantasy continuum has been mapped by Roberts in developmental origins to Protestant rationalist—Catholic theology/magic and mysticism. Adam Charles Roberts, *The History of Science Fiction*, 2nd ed. (London: Palgrave Macmillan, 2016), 3.

22 E. M. Forster *Aspects of the Novel* (Harmondsworth: Penguin Books, 1962), 109.

Imagination makes possible a conceptual blending that combines existing concepts in new, unforeseen ways—a 'recombinatory metaphoric process', as cognitive linguists have called it (though perhaps necessarily limited by the available pool of proto-concepts).[23] In fact, imagination might be what we do all the time when we are not engaged in a specific activity. Alan Richardson ties imagination to the 'default mode network' functioning of the mind at rest, which alternates between memory, planning, navigation, emotional processing and theory of mind.[24]

The imaginative leap needed to enter alien 'heads' requires huge empathy and creativity on the part of both the reader and author in order to transcend most human constraints and experience. But there are perhaps some shared universal constraints, in that we can often assume shared laws of physics (though many now think that other universes may have a quite different physics). Still, this is a much harder task than writing in the first or third (human) person and requires very careful and innovative use of language, as referents and connotations may be very different than those shared by humans. As speculation gets further from grounded concepts and known linguistic norms, it becomes harder for reader and author to share meaning. We do have a head start, as we have the 'theory of mind' needed to infer the inner states of fellow humans by observing their actions and listening to their words.[25] For alien and artificial minds, we just need to go further. In this book I will bring together what I consider to be some fine examples of different authors' attempts to do just this. My hope is that you bring all your empathic power to the reading of their work. But let us first further explore what these authors have revealed of their technique for writing the mind.

23 Alan Cruse and William Croft, eds, 'Metaphor', in *Cognitive Linguistics*, 193–222. Cambridge Textbooks in Linguistics (Cambridge: Cambridge University Press, 2004). https://doi.org/10.1017/CBO9780511803864.009.

24 Indeed, theory of mind—our insights into the minds, motivations and feelings of others—has been extensively related to science fiction in Nicholas Pagan's *Theory of Mind and Science Fiction* (New York: Palgrave, 2014). https://doi.org/10.1057/9781137399120.0001 Using characters from *Frankenstein* to *Do Androids Dream Electric Sheep?*, Pagan places his examples along a spectrum from the 'low road' of shallow emotional empathy to the 'high road' of a more intellectual understanding of alien and synthetic minds. He concludes that science fiction should be taken seriously as a literary genre for its depth of insight into these phenomena.

25 Ibid.