Self-Reflective Fiction and 4E Cognition : An Enactive Approach to Literary Artifice

Polvinen, Merja

2022-12-30

Polvinen, M 2022, Self-Reflective Fiction and 4E Cognition : An Enactive Approach to Literary Artifice. Routledge Research in Cognitive Humanities, Routledge, Abingdon.

http://hdl.handle.net/10138/578772

1. Introduction: An Enactive Approach to Self-Reflective Fiction

Merja Polvinen

https://orcid.org/0000-0002-4976-8104

Abstract
This opening chapter to the volume offers an overview of the theoretical approach put forward in *Self-Reflective Fiction and 4E Cognition*. It takes a look at the history and current state of studies on literary self-reflection, and presents the theories of enactive cognition that form the core of the analyses presented in the volume. Enactive cognition affords a way of explaining just how readers's embodied and emotionally engaged experiences may arise from the artefactual nature of narrative fictions. The introduction presents the main argument of the volume: that readers encounter fictions as artefacts rather than as an illusion of reality, and that while reading fictions relies on fundamental cognitive abilities, those abilities are engaged by environments whose qualities do not completely match those of physical reality. The result is an experience of being fully cognitively and emotionally engaged with both the fictional events and with their artifice. This doubled perspective, it is argued, can be fruitfully examined through enactive cognition. The introduction also gives an outline of the chapters that follow.

In *Reader, Come Home* (2018) Maryanne Wolf makes an impassioned plea for the value of reading as a human activity. In particular, Wolf wants to talk about the deep reading of long texts, as opposed to the skimming of briefer texts, which is what most of us now do online. Both of these forms of reading are specific cognitive skills that make use of a 'reading-brain circuit' formed between various networks in our brains as we learn to read. Wolf's main point is that the dynamics of this circuit remould themselves in response to the kinds of texts we read and the ways in which we read them. Thus, both of the two reading styles can be learned – and unlearned. Building on her own experience of no longer being able to enjoy the complex novels she read voraciously as a student, and then reflecting on this fact on the basis of the neuroscience of language and reading, Wolf comes to the conclusion that the deep reading of long and multi-layered texts is a skill that she used to have, but which she has let lapse. In its stead, she has perfected the hectic pace and 'spasmodic quality of attention' that serves us so well during online reading. Thus, her inablity to enjoy the novels she used to love was not the result of simple lack of time and concentration, but of trying to impose the rapid rhythms of her lately acquired online reading skills on texts whose own rhythms are much slower and more 'dense' (M. Wolf 2018, 96–102).

We are still in the early stages of finding neurological explanations for how and why different texts call for different kinds of reading actions, and what might be the consequences of such actions to the development of further cognitive capacities, but Wolf's description of her individual reading experience raises issues that I want to spin further in this book. The first of them is the conceptualisation of reading as an engagement with a textual environment that offers us particular experiential and conceptual rewards, but which does so only if we are able to interact with it in a way that activates those rewards. The second has to do with the particular textual mode Wolf is interested in: long, dense or difficult texts that seem to be asking their readers to do something that does not come naturally. The moments of expansion or insight we experience when reading is at its best depend, according to Wolf (2018, 64–68), on both what we read and how we read it, and our experiences then feed back into our later

encounters with other texts and with the world. *Self-Reflective Fiction and 4E Cognition: An Enactive Approach to Artifice* is about the ways in which literary texts are able to afford their readers the immersive, emotional, reflective, and even transcendent experiences that Wolf describes.

During the past few decades, many literary scholars have, like Wolf, begun to examine age-old questions of representation with the help of the cognitive sciences. These tools include new theoretical conceptualisations of, for example, literary empathy and imagined worlds that draw on studies of social cognition or mental simulations. Or they may involve more general theories of information processing during reading which draw on studies of sensory perception of the physical world around us. Many scholars have also adopted empirical methods from psychology, where readers' experiences are analysed through questionnaires or eye and brain scans.

The role of storytelling in human life is a complex issue that benefits from such rich interdisciplinary cooperation between the sciences and the humanities. The cognitive sciences can offer literary studies both experimental results and philosophical insights into the reading mind, and many cognitive scientists, for their part, have drawn on literary works as rich repositories of human mental life. However, more work remains to be done, especially in understanding how the cognitive environment of a literary work functions differently than many of the real-life situations we encounter. Many of the explanatory models currently used in the cognitive sciences do not always do justice to the complexities of literary imagining, and if transferred into literary studies without critique, they run the risk of producing an overly simplified picture of what literature is and what it does to its readers. Thus, it is important for humanities scholars engaging the cognitive sciences to continue to develop concepts and methods that take into account the history and diversity of literary forms, and to conduct detailed analyses of how complex human experience is represented in individual literary texts.

This book builds on the wealth of such work done within cognitive literary studies. My starting point is to examine literary works not as simulations of human life and experiences, but as crafted artworks. From this perspective, cognitive literary studies need to be able to explain not only the simulatory qualities of fictions – the reasons why it is so easy for us to speak of them as 'worlds' and of literary characters as if they were real individuals – but we also need to specify the cognitive effects of their artificial qualities, such as craftedness and fictionality. Thus, the central claim of *Self-Reflective Fiction and 4E Cognition* is that the models of cognitive science, to be most useful for the analysis of literary imagining, need to be able to discuss not just those literary phenomena that are deemed the most realistic (and most often cited as the ordinary elements of reading fiction), but also the less immediately life-like cases. For this purpose, I focus on the analysis of self-reflective fictions, where the issue of form, fictionality and artefactuality is encountered head-on.

Literary self-reflection is in this volume defined as those elements in a narrative that emphasize its fictionality, artifice, and constructedness. This mode is taken to include both traditional metafiction and other forms of less overt self-awareness, where the artifice of the narrative is underlined. As I will discuss in more detail below, the theories of metafiction of the 1980s and 1990s, including work by Patricia Waugh, Linda Hutcheon, and Brian McHale, took the logical and ontological paradoxes of postmodernism as their starting point, and presented metafiction as a way of exploring the constructedness of both textual meaning and reality itself. In these theories the play on artifice was also presented as overriding the immersive effects of the story. After the heyday of postmodernism, metafiction has attracted

much less attention, and views of its effects in postmodern works solidified into unexamined assumptions in later criticism. More recent work on self-reflective fiction has begun to unpick some of those assumptions, including, for example, the examination of particular narrative techniques by Ansgar Nünning, Dorrit Cohn, and James Phelan, as well as Irmtraud Huber's study of the authorial attitudes that extend beyond irony in later metafictive works. However, only a few scholars have taken a particularly cognitive perspective to the issue.

Within cognitive approaches to literature, the experience of fiction is often presented as an immersion in or a transport to a fictional world. Furthermore, and following the 1980s theories on metafiction, this experience of immersion is generally understood to be resisted, interrupted or cancelled out by moments of self-reflection that occur when readers are explicitly made aware of the fictionality of what they are reading. Finally, the immersion is most often seen in terms of specifically emotional engagement, whereas the self-reflective moments are taken as breaks in the emotional connection, caused by the fading of the aesthetic illusion and the arrival of an explicit awareness that the people and events we have so passionately cared about are, in fact, made up. Cognitive literary studies and cognitive aesthetics have also worked on defining emotions that can adhere to the artefact itself, but often this separate category of aesthetic emotional engagement connected with difficult, self-reflective texts is specifically contrasted with the immersive texts that sweep readers into an illusion of engaging with the events in a detailed world. Thus the relationship between the engagement with story events on the one hand and with the formal qualities of the artefact on the other continue to be thought of as a zero-sum game.

My starting point in offering an alternative to these views of literary artifice is to focus on the rhetorical options available to metafictional texts that previous cognitive theories seem to overlook on the basis of their understanding of what is central to fictionality. Many psychological and neurophysiological studies of literature encounter conceptual problems that arise from the idea of verisimilitude and illusion being central to fiction. Consequently they have not been fully successful either in addressing the way in which engaging with a work of fiction – however life-like – is always a coupling of mind with a crafted construct, or in analysing self-reflective works whose rhetorical structure is not, after all, built with the purpose of dismantling illusions. Thus, I want to turn the focus on how readers' minds are activated not only by life-like environments or believable characters in a text, but also by the fact of their constructedness and fictionality. I hope to show that the presence of artificiality, front-loaded in the self-reflective fictions I study here, is also central to the experience of fiction in general, and that continued development of theories that fully accommodate the cognitive exceptionality of narrative fiction is needed in the field of cognitive literary studies.

Throughout this volume, issues of literary theory are discussed in dialogue with research into the cognitive processes involved, and the claims are put to the test in readings of works of contemporary English literature. As I want to set cognitive literary theories in dialogue with works where conventional forms of representation are self-reflectively turned back upon themselves in different ways, my case studies include both more traditional works of metafiction and works of speculative fiction – a genre in which self-reflection on the structures of worldbuilding and storytelling is common, but where the function of that self-reflection is less overtly metafictional. Thus the writers included here span both mainstream metafiction and the speculative genres of science fiction and fantasy: John Barth (b. 1930), A. S. Byatt (b. 1936), Dave Eggers (b. 1970) and Ali Smith (b. 1962) are studied alongside Ted Chiang (b. 1967), China Miéville (b. 1972), Christopher Priest (b. 1943) and Catherynne M. Valente (b. 1979). All these authors, I argue, have shaped literary artefacts that assume

readers to be able to maintain imaginative engagement and skilful awareness at the same time, and that they reveal how a fundamental doubleness can reside at the root of our experience of fiction.

Within the multitude of approaches that make up the current cognitive sciences, I draw specifically on insights offered by the 4E or second-generation approaches to cognition, which take the mind to be embodied, embedded, enactive and extended, and which describe cognition as a dynamic process that involves complex feedback between an embodied being and its environment. Scholars of the 4E approaches such as Shaun Gallagher, Alva Noë, Evan Thompson, and Dan Zahavi draw not only on neuropsychology but also on phenomenology to form a philosophy of mind that is, in my view, more compatible with the experience of literature than the first-generation cognitive sciences, which drew from computational models and analytical philosophy.

On the basis of the combination of the 4E-cognitive and literary-theoretical perspectives, I present the experience of narrative fiction as a form of enactive perception where fictionality, instead of being a quality that fades to invisibility in order to be effective, is rather the perspective that makes the perception and comprehension of fiction as fiction possible. Furthermore, our experience of fictionality need not clash with our perception of and engagement with the fictional world and its characters but is conjoined with it. The enactive actualizing of the cognitive process presented by a fictional narrative need not be understood as shuttling between a re-enactment of a character's experience (the verisimilar content of the fictional representation), and the artificiality of the discourse patterning of the narrative. Instead it is an engagement with an abstract cognitive environment that can encompass everything from deep emotional connection with avowedly imaginary people with individual linguistic details and to the literary work's communicative status as a work of fiction.

Within cognitive literary studies, scholars such as Marco Caracciolo, Terence Cave, and Karin Kukkonen have taken up the second-generation cognitive approaches and are pushing the field in a more embodied and enactive direction. However, even in that discussion the artificiality of fictional narratives warrants more attention. One useful way of thinking about why artificiality changes some of the basic premises of cognitive literary studies is to engage with the concept of affordance: the action potentials offered by an environment to a particular cogniser. The concept was developed as part of the ecological psychology of James J. Gibson in the 1970s, and it has been adopted into literary studies for example by Cave and Caroline Levine, to help us think about the forms of literary works as action potentials offered to readers' imaginations. A simple example of an object's affordances can be found in a chair: it affords sitting (to human-shaped and adult-sized creatures), or it affords changing our embodied action potentials in a physical environment (if I step on it to reach a high shelf). It does not, however, afford eating it or listening to its story. If fictional narratives are thought of as objects with action potentials in this sense, what cognitive action do they afford? How do we recognise those affordances, and how do we as readers engage the cognitive environment afforded by the text? How might that action change our repertoire of available cognitive affordances beyond the text? What follows to our idea of what a literary work is if we think about what it affords, rather than about what it represents? In order to answer these questions my work will take up the 4E formulations of the dynamic interface between texts and their readers, and show how textual self-reflection can be looked at from a cognitive perspective without setting reflective action in opposition with the more commonly studied, naturalising cognitive-emotional effects of fiction.

Throughout the book, my aim is to show how crucially our understanding of the human imagination depends on our literary conceptualisations of artifice in fiction. This, of course was also a large part of the message of the 1960s and 70s metafiction, and John Barth suggests as much in *Lost in the Funhouse* (1968/1972): 'You tell me it's self-defeating to talk about it instead of just up and doing it; but to acknowledge what I'm doing while I'm doing it is exactly the point'. Slyly yet earnestly Barth's autofictional voice underlines the role of human action in the construction of art, and declares the value of acknowledging the fact that literary works are not only slices of life for us to experience, but also objects that have been artfully crafted. In the following pages of this Introduction I lay out the theoretical background concerning artificiality in more detail, and introduce the cognitive framework on which my thinking relies. I also give a brief orientation to the concept of literary self-reflection, and to the reasons why I have chosen to couple my argument on self-reflection with not only mainstream literary fictions but also with works of speculative fiction.

Self-Reflective Fictions

As long as there has been literary theory, there has been discussion about the artifice inherent in fictional representation, as well as literary works that make that artificiality explicit in their own rhetorical structure. Self-reflection comes in many forms, some more unambiguous than others. It ranges from metanarrative commentary by voices like Henry Fielding's or Anthony Trollope's narrators, who address their readers with remarks about their own storytelling, to the most radical metafictions, which flaunt not only the rhetorical construction of their narratives, but also the inventedness of the events depicted as well as the constructedness of human reality as a whole (Fludernik 2003; Neumann and Nünning 2012/2019). Metafiction – literary self-reflection at its most extreme – is fiction about fiction; that is, the kind of fiction that refers to its own fictionality, and as such it is often seen as a mode of writing designed to break the illusion of reality created by storytelling. It is most commonly associated with postmodern writing, and its central theories were formulated in connection with postmodern theory in the 1980s (Hutcheon 1980/2013, 1988; McHale 1987; Waugh 1984/1985; see also the texts collected in Currie 1995a). However, despite the fact that the term was not coined until the 1970s, and that postmodernism's rebellion against literary realism and the societal structures of the time made metafiction a particularly attractive mode for the writers of that period, metafiction is part of a larger family of self-reflective modes which have been part of fiction through the ages, and its repertoire and effects are not limited to those emphasised by postmodernism (Huber 2014). In this volume, therefore, I adopt the term 'self-reflection' as an umbrella term for a wide variety of texts and effects that are connected through their interest in artificiality and their willingness to be explicit about their own status as artefacts.[1]

The favourite early examples of textual self-reflection include Miguel de Cervantes's *Don Quixote* (1605–1615/1993), in which the first part of the two-part story is discussed by characters in the second part, and Laurence Sterne's *The Life and Opinions of Tristram Shandy, Gentleman* (1759–1767/1996), in which the narrative voice explicitly muses on the nature of storytelling, and the blank or blacked out pages draw attention to the book as a material object. Many theorists of literary self-reflection have drawn on Sterne's novel, including Viktor Shklovsky (1921/2012), who referred to its metafictive elements as the laying bare of its own devices, and Wayne Booth (1961/1983), who focused on the reflective relationship between the novel's narrator and its author. All these aspects of metafictive texts have thus been present in literature and in the theoretical discussion even before the postmodernist explosion.

The term ‘metafiction’ itself was first coined first by William Gass (1970/1971) and then brought into literary scholarship by Robert Scholes (1970/1995, 1979). Gass thought up the term as a way to draw attention the forms of self-reflective fictions; to describe authors who do not merely aim to write ‘about writers who are writing about what they are writing’ but rather aim to make the ‘forms of fiction serve as the material upon which other forms can be imposed’ (1970/1970, 25). Scholes went further along the same path by studying how metafiction incorporated the level of explicit literary criticism within its subject matter. Both of these interests – in forms as a metadiscourse about other forms and the consequent interest in metafiction as a literary critical discourse – have remained central to the study of metafiction, and they are clearly present in the works of those ‘fabulists’ (Scholes 1979) and postmodernists who knowingly blended their own theoretical knowingness into their stories – for example John Barth, Donald Barthelme, and John Fowles. These authors were part of a surge of interest towards ontological questions concerning language, reference, and fictionality, and their combination of self-aware irony and questioning of foundational concepts such as history, identity and meaning became the hallmarks of postmodern writing. While some critics, for example Gerald Graff, saw metafiction in a negative light and suggested that its central aim was to entirely sever the link between art and life, others, for example Robert Alter, saw metafiction to engage in a dialectic between the two (Hutcheon 1980/2013, 3).

This focus on questions of language and existence continued in the theories of metafiction of the 1980s (Hutcheon 1980/2013, 1988; Waugh 1984/1985; McHale 1987) which drew largely on the background of logical and ontological paradoxes. Patricia Waugh’s seminal definition of metafiction draws attention to the way the mode aims to dismantle the view from nowhere and use representation to look at representation:

> *Metafiction* is a term given to fictional writing which self-consciously and systematically draws attention to its status as an artefact in order to pose questions about the relationship between fiction and reality. In providing a critique of their own methods of construction, such writings not only examine the fundamental structures of narrative fiction, but they also explore the possible fictionality of the world outside the literary fictional text. (1984/1985, 2; emphasis original)

Linda Hutcheon also points out in *Narcissistic Narrative* (1980/2013) the demands that metafiction puts on its readers to actively construct yet simultaneously question the fiction. Literary self-reflection, she argues,

> attempts representation while discarding the myth of representation. It tries to transcend its own textual limitations while never forgetting that this is impossible. It ‘makes a primary virtue of honesty and yet proves its virtue by means of cunning tricks’. Such fiction is narcissistic in that it encourages an active personal response to itself *and* creates a space for that response within itself (1980, 141; citing David Caute).

Later, in *Poetics of Postmodernism* (1988) Hutcheon introduced the famous concept of ‘historiographic metafiction’ – a form of writing that denies both the naïve referentiality of realism and the solipsism that would completely separate fiction from the referential mode of history. Thus historiographic metafiction positions itself not just between fiction and literary criticism, as the theories of the 1970s metafiction did, but between fiction and historiography (1988, 105–123). Brian McHale’s (1987) discussion of the forms of postmodern fiction, for its part, focuses on ontological phenomena in which the hierarchy of narrative levels is

violated and thus foregrounded, making the existence of worlds – fictional or real – an explicit topic to be discussed. All these theories naturally drew on the concerns of the then contemporary postmodernism, and consequently the theoretical focus is on the constructedness of both text and reality. Similarly, because of the centrality of irony and play in postmodern metafiction, this work either implicitly or explicitly presents metafiction as undermining the immersive effects of the text and emotional engagement with the events represented.

Contemporary scholarly interest in metafiction builds directly on these theories. Many of the latest publications take the form of handbook introductions (e.g. O´Donnell 2005/2007; Berry 2012; del Pont 2018) or analyses of specific texts by later postmodernists such as Salman Rushdie, Paul Auster and Ian McEwan, or those of the later generation represented by David Foster Wallace, Jennifer Egan and David Mitchell.[2] New, more explicitly theoretical explorations can also be found in the narratological examination and classification of the metaleptic moves that occur between the levels of fictional discourse, and that break the borders between nested worlds of story, discourse and the readers' reality in various ways (e.g. W. Wolf 2004, Nünning 2004; Fludernik 2003; Bell and Alber 2012). The metafictional mode has also been studied in many other media, from film to comics to digital literature, often with a focus on the phenomenon of metalepsis (e.g. Kukkonen and Klimek 2011).

These recent theories of metafiction add to the previous work new analytical details, as well as the conviction that the functions of metafiction can be much more varied than was suggested by the postmodern focus on the unsettling of representational conventions and ontological certainties. But most importantly, they draw attention to the rhetorical aspects of metafiction, and to the effects that this particular mode of writing can have on its readers. When Mark Currie brought out his edited collection of the central writings on metafiction in 1995, he took up in his introduction Scholes's (1979) suggestion that the metafictional mode is first and foremost a kind of literary theory, even if it takes the form of fictional literature. One consequence of such a definition is that the writers of metafiction become also literary critics and thus representatives of 'Literatureland […] where texts and acts of interpretation constitute the world of experience' (1995b, 3). With this description, Currie gives metafiction a strong communicative function and the readers of metafiction a role as individuals willing to engage in contemplating literary critical and theoretical issues in this peculiar form of discourse.

This rhetorical perspective on metafiction links readily with cognitive study of narrative, and together they have generated fruitful discussions on the aims and effects of literary self-reflection. In this conversation, Monika Fludernik (1996) has been crucial in connecting fictional stories to the communicative dimensions of everyday oral storytelling, and further, all storytelling to the concept of 'experientiality'. In order to be a narrative, Fludernik has argued, a text must present to its readers a particular experiential quality. It is the analysis of the varying ways in which fictions present and manipulate this quality that constitutes what Fludernik has called '"natural" narratology' – the word 'natural' signifying the close connections formed between narratological analysis and the actions and experiences of human minds as they relate to everyday narrative representations. Fludernik (2003) has also made detailed narratological categorisations of different forms of metafictional and metanarrative techniques and their various perspectives towards the aesthetic illusion generated in readers' minds, and shows how the lines of such categories are drawn differently in different languages (in this case English and German). The connection between narrative and and the human mind as formulated by Fludernik has continued to be examined by

scholars in the field that came to be known as cognitive narratology, and where the work ranges over the many effects that narratives can have on readers, from the 'natural' to the seemingly 'unnatural' (see Alber, Nielsen and Richardson 2013; Alber and Richardson 2020).

Alongside Fludernik's natural narratology, James Phelan's rhetorical model of narrative also includes metafictional strategies as part of the 'synthetic' component of narrative. Alongside the 'mimetic' and the 'thematic' components, the synthetic functions as a textual quality that directs readers' interest towards the world and the characters of the narrative 'as artificial constructs' (Phelan 2007, 6). As the three sets of components take up various prominence in different texts, they also direct readers' attention in various ways to the three main rhetorical resources of narratives: their world-likeness, their ideological, ethical or cultural meanings, and their artificiality. While Phelan's interests do not particularly lie with metafictional texts, and in his analyses the synthetic component gets less attention than the other two (see also Clark and Phelan 2020, 5–8; Mikkonen 2022), his definition of the synthetic component as a constant, if often less prominent, element in all fictional narrative underlies my argumentation in this volume.

Metafiction, as a literary form that self-consciously reflects on the nature of fiction, is a phenomenon clearly dependent on what fiction itself is. More specifically, the rhetorical stance of each individual metafictional text is based on what kind of a phenomenon that particular text assumes fiction to be. Fictionality, as has been shown by centuries of debate, is a complex and historically contingent concept, and it is clear that since metafictions have all that variety of conceptualisations at their disposal, our understanding of the effects of metafiction must develop equal variety, rather than take all metafiction to be aiming at a single effect. As Richard Walsh (2007, 41) points out, readings that assume the generic qualities of metafiction to arise from a single rhetorical aim – for example that of destabilising the border between fiction and history – 'do so at the expense of the specific effects of a given text, beyond any such generic theme'. Thus, in addition to Phelan's discussions concerning the synthetic component in narrative, Walsh's work on the quality of fictionality provides another stay for discussing literary self-reflection in rhetorical narratology. It is on the basis of these elements from the more traditional literary studies – studies of metafiction, cognitive narratology and rhetorical narratology – that I set out to examine the consequences of metafictional strategies to the experience of reading a fictional narrative.

One further element that I want to bring up here in the Introduction is the way in which literary self-reflection appears in forms of literature that do not match the traditional genre characteristics of metafiction. If the rhetorical aims of literary self-reflection can vary, one of the discernible patterns in that variation arises from the particular genres in which the self-reflection appears. In this volume I focus on two genres in particular: 'mainstream' literary metafiction and speculative fiction. While the former texts (those by A.S. Byatt and John Barth, for example) are easily recognised as examples of self-reflective literature as it is traditionally understood, the latter ones (texts by e.g. China Miéville or Ted Chiang) are not called metafictions, nor do they tend to get discussed in terms of literary self-reflection. Speculative fiction or SF – here including the genres of science fiction and fantasy – is not often discussed in narrative theory, nor, on the other hand, have speculative fiction studies often focused on narrative form. Generic plot-types and techniques of exposition have received their fair share of attention within the study of the speculative genres (see e.g. Malmgren 1991 for science fiction; Mendlesohn 2008 for fantasy), but there are still very few

scholars (e.g. Caracciolo and Ulstein 2022; Roine and Suoranta 2022) focusing on the details of narration in SF texts, let alone on narrative structures that challenge conventional literary techniques and their naturalised interpretations.

Within narratology, on the other hand, speculative narratives have been discussed within the paradigm of unnatural narratology (see Alber, Nielsen and Richardson 2013; Alber and Richardson 2020). In this discussion, however, the specific genre features of SF and the self-reflective use of those conventions are not in focus; usually they are covered with general statements about how the deviation of SF conventions from real-world parameters is an accepted part of the genre (e.g. Alber 2013, 51; Heinze 2013, 36). Narratology's general silence on speculative fiction has, however, been broken by Brian McHale, whose work from *Postmodernist Fiction* (1987) onwards has engaged science fiction as a genre where literary form – including techniques of worldbuilding and fluctuating ontologies, or the impossibilities of omniscient narration or temporal disordering – is often self-reflectively on show (McHale 1987, 2010, 2018, 2021). Similarly to McHale's work, the material analysed in this volume includes an array of texts in which conventional forms of representation are self-reflectively turned back upon themselves, but in which the exact characteristics of that self-reflection depend on not only the connections and histories of 'high' literary metafiction but also those of speculative fiction. As the readings in the following chapters aim to show, SF is a genre where narrative self-reflection – or 'literal narratology' (McHale 2018) – is easy to accomplish and its effects frutiful for narrative scholars to examine.

Finally, metafiction can also be seen to approach speculative fiction in the sense that it partakes in the sense of the uncanny. As Nicholas Royle (2002, 19) notes, 'as a figure and experience of what is at once inside and added on, always already at home yet an outsider, constitutive yet supplementary, the "meta-" itself may be uncanny'. Thus metafictions and speculative fictions are connected by their explicit engagement with both the conventions of narrative discourse and with the fundamental strangeness of fictionality: with the doubled experience of being inside and outside a world at one and the same time, or of perceiving objects and individuals even as we also perceive their imaginariness and their artificiality.

Second-Generation Cognitive Narratology

As I mentioned in the opening of this Introduction, cognitive literary study has grown into a wide and diverse field. This variety could be roughly mapped by focusing on two different axes of interest: one for scholars' interest in either theoretical or empirical study, and another for their adoption of either 'first-generation' or 'second-generation' cognitive science models. The axis of theory and empirical study has been present in literary studies already in the differences of, for example, poststructuralist theory and reader-response studies. But it has become much more visible within cognitive literary studies, as scholars have gravitated, on the one hand, towards philosophy of mind and consciousness, or, on the other hand, towards empirical psychology and neuroscience. My own work in this volume is motivated more by the philosophical than the empirical aims, as I wish to focus on conceptual re-definitions of fundamental issues such as perception, cognition and imagination. I find great value in empirical testing, but I also believe that at this moment in the development of the field, our understanding of both human minds and literary works is well served by scholarship that aims to query the philosophical basis on which such tests are designed. As will be discussed in more detail in Chapter 5, I believe that some of the background assumptions of the empirical work done so far on the human imagination in general and the experience of

literature in particular need to be readjusted before the testing will produce the kinds of results I feel we need. Thus, this book does not present empirical work of my own, nor does it always build directly on the existing empirical evidence. Rather I want to work on our models of mind and literature, query some of the models used so far, and present testable claims appear here only tangentially or perhaps to be picked up by other scholars for further work.

As to the second axis of interests, I take my cue from the second, embodiment-oriented, rather than the first, computational generation of cognitive science. This generational division is by no means unproblematic, since it implies a line of succession that is not really representative of cognitive sciences as they now stand. First of all, most of empirical neuroscience continues to work within the classical computational and representational paradigms, and thus calling those paradigms the first generation, as if their time was past, gives a wrong impression of the field. The second generation, on the other hand, has roots that go all the way back to early 20th-century phenomenology, and thus implying that it consists of only recent innovation is just as mistaken.[3] However, since the labels have already received some mileage, I use them here to position myself and the theories I draw on.

To generalise, the classical or 'first-generation' theories of cognitive science rely on computational theories of cognition. Such views of mental functioning (also known as 'cognitivist' theories) present the mind as information processing and take the computer as their paradigmatic model for cognition (Harre 2002, 113–143). The 'second generation', as termed by Mark Johnson (1997, 148; 2018, 626), is seen as a wave of theories particularly interested in the embodied mind and in the dynamic interaction of such a mind with its environment. In *Stories and the Brain* (2020, 13–25) Paul B. Armstrong traces the moves made from the first-generation to the second-generation theories, and shows how the adoption of the two paradigms in literary studies – narratology in particular – matches other moves made there as the field changed its structuralist models to more phenomenologically and pragmatically minded theories. Central in both the cognitive and the literary theoretical dichotomies is the difference between, on the one hand, modelling structural relations in stable and logical systems (associated with the vocabulary of frames, scripts and schemata), and, on the other hand, describing mind and narrative as situated, embodied interaction between actors in complex environments. Armstrong sides with the latter – as do I – and argues (with reference to Hubert Dreyfus) against the traditional analogy of the brain as a symbol-manipulating computer:

> Computers lack context, background, and prior experience that we as embodied conscious beings typically employ in testing hypotheses about how to configure a situation we encounter, whether in a text or the world, and replacing this deficiency by positing preset mental constructs that do the work only displaces the problem that needs to be solved. (2020, 18)

To replace this computer-model of the brain Armstrong (2020, 2, 15–25) presents the 'neurophenomenological' tradition of second-generation cognitive theories, and links it with a similarly phenomenologically influenced tradition within narratology that runs through Wolfgang Iser and Paul Ricoeur.

In this book I build specifically on the second-generation theories, and suggest that they present an interesting and fruitful basis for a view of mind and literature. Thus, without wading any deeper into the debates about what is valid within the cognitive sciences themselves, I identify myself as having adopted the perspective of the phenomenologically inflected second-generation models for theorising fictional narratives. The theories presented

in the context of the second-generation cognitive literary studies, particularly their description of the dynamic interface between texts and the minds of readers, help me tackle the problem of textual self-reflection in a way that does not set reflective action in opposition to the other, more commonly studied cognitive-emotional effects of fiction.

Second-generation approaches to literature mostly refer to a combination of theories known as the 4E-theories of the mind (see Kukkonen and Caracciolo 2014). These theories take cognition as *embodied*, *embedded*, *enactive* and *extended*, and they distance themselves from the representational and computational paradigms of cognition. The four central (and nicely alliterating) concepts of the 4E theories may have been originally selected somewhat by chance (Menary 2010, 459–360; Newen, De Bruin, and Gallagher 2018b, 4), but they all present challenges to the internal, amodal and representation-based models of the mind common in computational cognition. *Embodiment* asks us to abandon the earlier computer metaphor and to take into account the full somatosensory system as well as the body's hormonal traffic as parts of the thinking being. *Embeddedness* asks us to also take into account how the surrounding world impinges on the thinking body; and the *extended mind* paradigm goes further to claim that not only does the world affect our cognition, but that in using parts of the world in the action thinking, the mind in fact extends beyond the limits of our physical bodies. Finally *enactive* theories suggest that our minds and the world are structurally coupled (Maturana and Varela 1972) so that our embodied action constitutes the world we live in, just as the world allows us to be only in certain ways and not in others. All of these four Es have been taken up within second-generation cognitive literary scholarship with varying emphases.

Central work within second-generation cognitive literary studies include, among others, studies by Lars Bernaerts, Marco Bernini, Marco Caracciolo, Terence Cave, Ellen Esrock, Kaisa Kortekallio, Karin Kukkonen, Anežka Kuzmičová, Yanna Popova, and Emily T. Troscianko (see also Kukkonen and Caracciolo 2014). While there are differences within this group that will come clear in the chapters that follow, few of those differences amount to more than variations in emphasis. Caracciolo, for example, has focused on enaction, and his volume *The Experientiality of Narrative: An Enactivist Approach* (2014), lays out the dynamics of embodiment as represented in the experience of fictional characters, as well as how texts work to transfer those dynamics to the experience of readers. Popova's volume *Stories, Meaning, and Experience: Narrativity and Enaction* (2015) covers many of the same issues, as well as the relationship between metaphor and narrative. Kukkonen, on the other hand, has laid out a detailed model of analysis based on both embodied and extended cognition, particularly as they appear in Andy Clark's (2016) predictive processing model (Kukkonen 2017, 2019, 2020). Particularly useful in Kukkonen's work is her emphasis on form and her description of literary works as 'designed sensory flow' (2020, 5), which draws attention to how the crafted shapes of narrative move readers in rhetorically purposeful ways. Ellen Esrock (2001; 2004) has worked specifically on readers' enacted perception, and suggests that such responses should be thought of in terms of simulatory processes. Troscianko (2013), on the other hand, draws attention to how conceptual structures influence experiences of vividness, and Kuzmičová (2012) has studied, for example, readers' motor enactments of fictional spaces. Second-generation cognitive approaches to speculative fiction are a very recent endeavour, but at least two strong PhD dissertations have already appeared: Kaisa Kortekallio's reading of embodied figuration in SF (2020) and Gry Ulstein's discussion of environmental narratives and the Weird genre (2021).

Many of these scholars working have their roots in narratology, and therefore form does play an important role in their work. However, the issue of literary self-reflection is rarely focused on as such. The exceptions to this are Kukkonen, who theorises metafiction as part of the the designed flow of sensory experiences offered by narrative (2020, 96–104), Bernini (2014, 2018), whose work draws parallels between the actions of reading and writing in extended and enactive environments, and Bernaerts (2017), who has studied the abstract spaces of experimental narratives. The current volume builds on all this earlier work on literature and 4E cognition and applies such thinking more explicitly and extensively on the questions of formal self-reflection. However, rather than take on all of the different varieties of 4E theories equally, my focus is on enactive cognition and on the ways that it conceptualises our cognitive environments as constituted through processes of embodied sense-making.

Enactive Cognition and Literary Self-Reflection

Enactive cognition, as described by Varela, Thompson and Rosch (1991/1993) was originally conceived as an embodied and phenomenologically aware alternative to mainstream cognitive science. It presents cognition as a process of sense-making built on autopoiesis; that is, as a dynamic relationship between a life-form and its environment. Enactive cognition is embodied cognition – as opposed to being abstract computation – but even more importantly, it is a form of embodied being that enacts the environment it exists in. As living and thinking beings we know 'how to negotiate our way through a world that is not fixed and pre-given but that is continually shaped by the types of actions in which we engage' (Varela, Thompson and Rosch 1991/1993, 144). All thought, feeling and action is here seen as skilful embodied engagement with our environments – engagement where those environments themselves are enacted or brought forth by the creature as it uses, for example, its sensorimotor skills to make certain parts of its environment available to further perception and action (Colombetti and Thompson 2008, 55). Cognising beings thus establish for themselves a world of meaning; an *Umwelt* in Jakob von Uexküll's terms (1934/1957), and in that meaningful environment questions of relevance are constantly in flux, depending on the kind of activity the cogniser engages in, and on what is available in the environment. In such an interaction, the environment is not just something the creature passively exists in, but the creature's movement, its consumption of available materials and its behaviour with other creatures change that environment from moment to moment. As Shaun Gallagher puts it, for the proponents of enactivist theory,

> the explanatory unit of cognition (perception, action, etc.) is not just the brain, or even two (or more) brains in the case of social cognition, but dynamic relations between organism and environment, or between two or more organisms, which include brains, but also include their own structural features that enable specific perception-action loops, which in turn effect statistical regularities that shape the structure and function of the nervous system. (2017, 11)

On the basis of such conceptualisations, enactive theory can be seen to be useful in three areas of cognitive science in particular: firstly, in relating 1st-person phenomenological experience to the 3rd-person scientific study of cognition; secondly, in linking together the various domains of cognitive science that focus on different levels of cognition from cellular interaction to complex social cognition; and thirdly, in reminding cognitive science to be reflective about itself and about its own perspective as a form of human cognition studying human cognition (Gallagher 2017, 10–12; see also Thompson 2007). While enactive

cognition originally arose within biology and cognitive science as a way of modelling very basic processes of perception and movement, and while it is still often thought of as a theory of 'low-level' cognition contemporary work includes studies that scale enaction up to the higher levels of cognition such as memory, reasoning and imagining (Stewart, Gapenne and Di Paolo 2010/2014, viii; Gallagher 2017, 187).

The enactive continuum between the various domains of cognitive action, from cellular interaction to complex social cognition is based on the concept of *sense-making*. An autopoietic creature that regulates its structural coupling with the world in the way described by enactive theory also 'establishes a *perspective on the world* with its own normativity, which is the counterpart of the agent being a center of *activity in the world*' (Di Paolo, Rhode and Jaegher 2010/2014, 39, emphases original; see also Colombetti 2014, 15–19). This combination of a perspective on and centred activity in a world makes the embodied interaction with the world meaningful for the agent. 'Sense-making' means this dimension of cognition whereby we engage and enact the world through perception. The process is best described as bringing forth a world through skilful sensorimotor action, rather than receiving data from the outside and then representing a world in our brains. In developing the theories of enactive perception and 'sensorimotor coupling', Alva Noë (2004, 1) has influentially argued that 'perception is not something that happens to us, or in us. It is something we do'. Thus all living creatures act to make various parts of the world perceptually accessible at various times, and this fundamentally action-dependent view of the world is what enactive theory brings to the cognitive sciences.[4]

For literary scholarship, theories of real-world perception are less central than those of the imagination, as we want to talk about how readers engage with texts as something more than their physical manifestation as marks on a page. From the enactive perspective, imagining is a complex congregation of practices, which together form the embodied and affective experience of things that are not present. When imaginative practices are understood in an enactive way, they are, like all other forms of cognition, seen to take place in a constitutive interaction with an environment, but this time that environment is not physical; for example, it can consist of concepts, images and cultural practices (Gallagher 2017, 195). Enactive imagining can of course be fully embodied, as when gesture and movement in space couple with culturally coded forms so that by moving, we can actually do cognitive work. Edwin Hutchins's famous study of a ship's navigator is one example of this: the navigator does complex calculations not just in their head or on paper, but through the physical movements they make with the map and the measuring instruments (Hutchins 2010/2014, 442–445). Playful or artistic imagination, when it uses the physical environment in a similar fashion, turns into pretense or playacting: for an actor pretending to be a tiger, the imagining 'is accomplished in the playacting', rather than the actor first forming a mental representation of themselves as a tiger and then acting it out (Gallagher 2017, 193).

Enactive imagining can, of course, also forego actual bodily movements altogether. For enactivists such as Gallagher, Evan Thompson, or Dan Hutto and Erik Myin, this kind of abstracted imagination involves mental action on re-enacted perceptions. As Gallagher argues in his discussion of Gilbert Ryle's theories, 'the imaginative practice is to manipulate thoughts, images – take them up and play with them, move them around, in order to solve a problem, or map them onto novel affordance spaces' (Gallagher 2017, 196).[5] This kind of enactive manipulation of thoughts and images is what my model of enactive cognitive narratology builds on. In order to think of thought and images as parts of an environment that can in this way be 'manipulated' in the imagination, I also draw on the concept of *affordance*.

Affordances have become a central element in an enactive model of imagination, since they define what actions the cognitive environment makes possible for the cognizer. As I mentioned earlier in this Introduction, the concept comes from the ecological psychology of James J. Gibson (1979), and it refers to the way environments 'offer' or afford certain behaviours and do not afford others. An affordance 'is both an external phenomenon covered by the laws of nature' that sets the physical properties of an object, 'and a potential for action held from a point of view' (Kramnick 2018, 11). In addition to 'laws of nature', affordances can also be seen in less concrete cognitive actions such as abstraction and future projections (Di Paolo, Buhrmann and Baradiaran 2017, 218–249), or in language as social sense-making (Cuffari, Di Paolo and De Jaegher 2014). Written text, therefore, can function for writers as a way of constructing, rather than expressing, thought, and for readers as an opportunity for performing imaginative action rather than receiving perceptual data. As examples of these more abstract cognitive objects we can think of the linguistically encoded scholarship stored on my laptop, the notes I have made in libraries, the bulleted outlines and drafts I have created of this chapter – all of these afford me to write the text you are reading, and, indeed, to think things I would not be able to think without their support. Furthermore, as Gallagher points out, imaginative action is also affordance-based, and these affordances can be either physical props used in pretend play, or social and cultural 'instituted practices' (2017, 194) that make certain imaginings possible, and that are constantly changed and extended by the imaginative actions we undertake.

Affordances and imaginative action have been explicitly related to literature by Terence Cave (2016), who suggests that reading is 'thinking with literature', or a process of creating and making use of cognitive affordances. In Cave's brand of 'cognitive pragmatics' (24), affordances enable a way of thinking of the relationship between form and content as one of ecological relevances. Following the original theory of affordances, Cave's approach is pragmatic: no individual formal element affords the same cognitive action for all readers, but instead, an affordance is '*underspecified*', and it is the relevance of one quality or another in a particular cognitive situation with a particular thinker that defines its level of presence. Affordance in a text is 'a thing that adumbrates a purpose or indefinite set of purposes; only a particular use and a particular context can select the *relevant* purpose' (Cave 2016, 51; emphases original). As an example of how affordances could be used in literary analysis Cave discusses the development of literary genres; how qualities such as the form of the sonnet appear in the ecology formed by cultural contexts, and prove themselves to have relevance and use-value for certain cognitive actions. The *volta* – the formal and conceptual turn in the sonnet – can, Cave suggests, be seen to afford the 'articulation' or flexible connection of 'the argument or state of feeling elaborated in the quatrains' to the more 'aphoristic formula' of the final couplet (56). Thus affordances 'are at once immanent dynamic form and *emergence*, the phenomenon as it takes shape in the moment' (62; emphasis original). It is exactly their refusal to be extracted from their particular environments that makes the concept valuable for a literary scholar.

While Cave's focus is on the affordances of one of the most iconic literary forms, the sonnet, the concept has also been discussed in relation to the textual characteristics of speculative fiction by Kaisa Kortekallio. In *Reading Mutant Narratives* (2020), Kortekallio examines the affordances certain SF texts have for changing readers' 'experiential habits' and allowing readers the opportunity for encountering more-than-human experientialities. This model of 'performative enactivism' takes literature as an affective technology where the particular genre features of SF allow for different kinds of embodied engagements with narrative than mainstream fictions tend to do. Kortekallio's focus is particularly on the figure of the mutant,

a textual pattern that offers an opportunity for readers to reconfigure their own experiential relationship with the nonhuman world.

These two brief examples from Cave and Kortekallio of how affordances can be used to analyse both traditional and new literary forms reveal the usefulness of the concept for reading the abstract patterns of thought and rhythm in the sonnet, as well as the immediately and intuitively perceivable embodiment of characters in even the most wildly fantastic texts. In the chapters that follow, I will build on such work in the second-generation paradigm and approach literary works from the perspective of enacted affordances. I ask what cognitive action narrative fictions afford, and how the study of affordance and literary form can help us to connect the abstractions of self-reflective fictions to the embodied experiences of both characters and readers. My answers emerge from recognising the artificiality of literature and from emphasising how readers' skilful navigation of that artificiality is one aspect of the cognitive environment they are in as they read. This view of the literary work as a cognitive environment makes it possible for the work's various aspects – from bodily movement of characters to self-reflective artificiality – to be experienced more or less seamlessly. As discussed above, despite its early focus on low-level cognition, enactive theory does also have a handle on complex cognition and reflection, and it posits that reflective thinking is not a separate process from or antithetical to cognition that takes place nonconsciously. Deliberation, decision-making, belief and embodied imagining all mesh in literary experience to support, rather than oppose each other.

As my main focus is on the affordances and cognitive ecologies, in this book the 'environments' of narrative fiction are understood not in the sense of readers enacting the world represented in the text, whether a dystopian New York or Dickens' London, but instead in the sense of readers enacting an environment that is a complex cognitive space. In such a space, fictionality, artefactuality and narrative structure are explicitly and in-eliminably part of the environment. In this, I take part in the conversation within cognitive literary studies that tries to move away from seeing literary effects as naturalised on the basis of what they represent, and more towards seeing them as generated by the exceptional formal and ontological qualities of narrative fiction (see Kukkonen and Nielsen 2018). The aim is to get past the thought that narrative fiction does its work most effectively when readers are immersed in an illusion of a world, or that explicit artificiality would create a perceptual and emotional distance between fiction and its readers that interferes in that work. By focusing on effects that are thought to exist in spite of the fiction's artificiality, much of cognitive literary studies has paid less attention on those embodied and emotionally engaged effects that arise precisely from the artefactual nature of the text – from its form rather than its content, to use the traditional although contested division. As a result, those aspects of literary experience that have to do with a full-on cognitive-emotional engagement with fictions as fictions, rather than as illusions, are left unexplored. In this book I argue that what is encountered during reading is experienced both as being *like* a world, but also *as* a fiction; that is, it is approached and engaged as it is because of what it is – an artefact, rather than an illusion of reality. Therefore, my particular interest in the role of the clearly signalled artefactuality of the literary work arises from wanting to understand how readers' minds are able to assume a perspective that is fully cognitively and emotionally engaged with both the fictional events and with their artefactuality.

Contents of this Volume

All of the chapters that follow turn around two major issues concerning fiction and self-reflection: firstly, readers' experience of fictions as abstract environments that have various degrees of spatiotemporally worldlike qualities; and secondly, readers' ontological and emotional positioning towards such environments. These two central themes are looked at through perspectives typical to the study of literature and cognition: the action of metaphorical 'seeing as' in Chapter 2; readers' experience of space in Chapter 3; their experience of narrative time in Chapter 4; and finally the roles of belief and emotion in readers' engagement with fiction in Chapter 5. Throughout, I set theory in dialogue with literary texts that are self-reflective about their own artifice. As I mentioned, these texts represent both more traditional mainstream metafictional qualities and the literalisation of form typical to speculative fiction. Within these wider categories, I have also selected texts that have a particular rhetorical aim with their self-reflectiveness: rather than wanting to flaunt their artificiality for the sake of undermining their own immersive effects (which was often the case with postmodern metafiction) these texts have a built-in assumption that readers are able to maintain engagement and awareness of artificiality at the same time. It is the literary theoretical premises of that assumption that this book will try to outline.

Following this Introduction, Chapter 2 focuses on enactive cognition and the doubled perception of seeing one thing as another, as well as the immediate perception of a thing simultaneously with of its mode of representation. Most recognisably this form of perception is present in metaphor. With Paul Ricoeur's (1975/2003) discussion of metaphor and narrative as two expressions of the same representational dynamic as my starting point, and moving on to contemporary enactive theories of metaphorical cognition (e.g. Jensen and Greve 2019), I argue that the double vision of metaphor is a process of enactive sense-making whereby we modify the affordances of our cognitive environments – whether those environments are physical or conceptual. My case studies in this chapter are A.S. Byatt's 1985 novel *Still Life* and Catherynne M. Valente's 2011 novella 'Silently and Very Fast', both of which examine the role of metaphor in human thinking. The chapter also lays out some of the groundwork for the two forms of self-reflective fictions I study here: mainstream metafiction and speculative fiction. I argue that the rhetorical aims of Byatt's and Valente's texts are similar despite the fact they come from very different traditions of literary representation, and that they both present their readers with the kind of literary self-reflection that combines immediate, embodied perceptions with experiencing the artifice with which those perceptions are brought about. The double vision discussed here will be also taken up in Chapter 5, where, instead of the perception of individual objects and metaphorical phrases, the focus is on our perception of an entire work and of its fictional mode of presentation.

Chapter 3 turns the focus on readers' spatial perception. I examine the concepts of immersion, transportation and 'worldness' that express the underlying spatial metaphors widely adopted within cognitive literary studies for the experience of fiction. These metaphors are used in the models of possible worlds theory and deictic shift theory (DST) as well as in the common-sense vocabulary of experiencing fiction, and they form a basis for many cognitive explanations of how fictional worlds are perceived as spaces readers imaginatively enter (Herman 2004; Ryan 1991, 2001; Segal 1995; W. Wolf 2004). One consequence of the spatial metaphors has been a claim about literary self-reflection: that since reflective awareness implies an external view on the fictional world, it must also dismantle readers' sense of being immersed in that world. In contrast, I suggest that readers regularly encounter the seemingly paradoxical combination of experiencing immersion

together with an explicit engagement with unreality, and that this combination does not involve them in an alternation between two incompatible perspectives. Enactive perception, I argue, provides an explanation as to how such an experience is possible. On the basis of work by Alva Noë (2004, 2012, 2015), as well as theories of cognitive affordances (Gibson 1979; Cave 2016), I explore what happens to the embodied experiences of imagined spaces when fiction is explicit about the constructedness of its worlds. What exactly is being experienced as present and accessible to us: the crafted, communicative fiction, the world it seems to represent, or both? In this chapter, I also analyse China Miéville's conceptual thriller *The City & The City* (2009) and Ali Smith's metafictional *There but for the* (2011). Where Miéville's speculative fiction uses readers' perception of generic structures to thematise and estrange their perceptual experience of space, Smith's more mainstream novel focuses on how language creates both controlled spaces and human freedoms, and on how readers take up their own space within a text. I suggest that a combination of affordance theories, reader models from second-generation cognitive studies, and my readings of these two self-reflective novels offers a way of describing narrative spaces as abstract cognitive environments, and that the affordances of such a space both are and are not like those of our physical reality.

Chapter 4 extends my focus from the spaces of narrative fiction to spatiotemporality as a larger phenomenon. In our critical discourse, experiences of self-reflection are often presented not only as moments of stepping out of the fictional world, but also as suspension and pause that stall the narrative movement. In contrast, my aim is, firstly, to show how textual self-reflection can be described as an embodied experience of speed, generated by our movement through the forms, concepts and imaginings offered by the cognitive environment that is the narrative. Here I focus on John Barth's 'On With the Story', and combine a reading of the metafictional elements in the short story with an enactive, embodied theory of literary self-reflection as *vection*. Secondly, I will examine the phenomena of narrative suspense (Sternberg 1978, 1992, 2003) by analysing the self-reflective temporal structures of Ted Chiang's science fiction novella 'Story of Your Life'. Here the slow revelation of the narrator's simultaneous experiencing of past, present and future literalises the conventional separation of story time and discourse time. This temporal perspective in Chiang's story also leads me to examine the phenomenology of re-reading, and the ways in which that cognitive process complicates our understanding of the role that knowledge about the artefact plays in our experience of narrative fictions.

Chapter 5 takes a deeper look at fictionality in the experience of literary artifice. It suggests ways in which the combination of enactive cognition and rhetorical narratology is able to describe the experience of maintaining an active imaginative and emotional connection with narratives when their artificiality is made explicit. I first approach the issue via the so-called 'paradox of fiction', and the ways in which empirical work within cognitive literary studies has dealt with the combination of emotion and artifice. The role of emotion in fiction has often been conceptualised through ontological structures and the rational beliefs and propositions attached to such structures (Konrad, Petraschka and Werner 2018). This is true also with empirical studies of the reading of fiction, many of which inherit their conceptual layout from psychological studies of the imagination, and this causes problems for their ability to study the effects of self-reflective fictions. My suggestion is that the idea of enacted cognitive environment can explain how readerly engagement with fictional characters and the act of fiction-making can both root themselves in interpersonal cognition, and yet bypass belief as a necessary element in engaging with fictions. In this chapter, therefore, the focus is on fictionality as an aspect of narrative artifice that – along with spatiality, temporality and

the lamination of metaphorical thinking – does not deny our basic cognitive capacities, but instead engages them in a way that is different from what we do in real-world contexts. The touchstone text in the first part of the chapter is Dave Eggers's *A Heartbreaking Work of Staggering Genius* (2000), a memoir that embeds an autobiographical narrative within self-reflective commentary, and which explicitly calls attention to the issues of emotional sincerity and rational distance. By constructing a hyper-aware scaffolding around the raw emotion in the memoir, Eggers makes use of the affordances of the various author and audience positions available to fictional narrative, and engineers a situation of joint attention or participatory sense-making with his authorial audience. In enacting the self-reflective scaffolding, readers experience both the pain and the self-deprecation of Dave Eggers in a manner that would not be possible without the cognitive moves structured by that scaffolding.

My final textual companion is Christopher Priest's exploration of the effects of fiction and stage magic in his 1995 novel *The Prestige*. Priest's narrative mixes historical romance and science fiction to examine audiences' relationship with the act of magic – whether belief in the supernatural or engagement with the artifice of the magician is more relevant to their engagement with the act. Most central to the novel and to my analysis is the audience's expectation of unreality and their engagement with the entire act of conjuring, rather than just with the results produced. Fiction, I suggest, similarly engages readers the representation as well as in what is represented, and out of that full enaction emerges a cognitive state of 'lucid self-delusion' (Landy 2012). This experience would not exist without readers' active complicity in a fiction as fiction, or without their knowledge of the meaning-making actions required in the performance of that literary magic.

Our studies of the seemingly irrational and contradictory ability readers have of engaging fully with worlds and beings they know are constructs have not, I suggest, been fully successful when they build on structures of belief, on the opposing hierarchies of cognition and emotion, or on the spatial metaphors usually associated with experiencing fictional worlds. The second-generation cognitive sciences, in moving away from the computational paradigm, are less interested in thinking of the imagination as either simulated visuals or rational information processing, and much more focused on the seemingly self-contradictory sides of the imagination. The combination of the enactive paradigm and the theories of fictionality within literary studies can, I argue, help each other in forming a more nuanced view of the imagination for both literary studies and the cognitive sciences. The conception of literary imagining offered in *Self-Reflective Fiction and 4E Cognition* helps to cut through the earlier dichotomies and to question those psychological explanations of readers' engagement with narrative fiction that rely on notions of illusion or belief, on human universals disconnected from cultural contexts, or on models that dismiss the level of language and artful construction in literary works.

---

[1] I have also chosen the term 'self-reflection' in favour of 'self-refentiality', since the latter is often used in psychological discussions for the human metacognitive ability to reflect on one's own thoughts – an issue that is not in focus here.

[2] For studies on some of these authors see e.g. Butler and Gurr (2008); Huber (2014); Johnston (2017) and Phelan (2007).

[3] For discussions on first and second-generation approaches within the cognitive sciences, see e.g. Newen, De Bruin, and Gallagher (2018b); Carney (2020).

[4] Gallagher (2017, 21–24) also notes how, in order to get a handle on enactive cognitive science, the best way is to take it less as an empirical scientific research paradigm, and more as a 'philosophy of nature', in that it is holistic and does not lend itself easily to clear operationalisation.

[5] For an argument about how the pragmatist philosophy of e.g. Peirce and Dewey also acts as precursor to both enactive and extended cognition, see Gallagher (2016).