[Manuscript version. Please cite published version in T. Demeter, T. Parent and A. Toon (Eds.) *Mental Fictionalism: Philosophical Explorations* (pp. 52-69). Abingdon, Oxfordshire: Routledge]

Fictionalism and intentionality

Adam Toon
University of Exeter

Abstract

This chapter offers a defence of mental fictionalism. Its central claim is that the notion of the mind as an inner world of representations is merely a useful fiction. Mental fictionalism is often said to suffer from "cognitive collapse", since stating the fictionalist's position itself involves reference to mental states, such as imagination or make-believe. This chapter shows how mental fictionalism can avoid cognitive collapse. To do so, it explores fictionalism's broader implications for the nature of intentionality. The key to avoiding cognitive collapse is to see that fictionalism can grant the existence of external, public representations with content, such as written and spoken language. In contrast, the notion of inner representations is what the early fictionalist Hans Vaihinger called a "real fiction": it is an idea that is not merely false, but incoherent.

1 Introduction

What are thoughts and how do they represent the world? The representational theory of mind (or *representationalism*) offers a straightforward and influential answer to this question: thoughts are inner representations (or *mental representations*). Our thoughts have content because our mental representations do. Indeed, according to representationalism, *all* intentional phenomena—from language to maps, diagrams and road signs—ultimately gain their content from mental representations. Unfortunately, the question of how mental representations gain *their* content remains so far unanswered. *Mental fictionalism*—at least, as I understand the approach (Toon 2016, 2021a, 2021b)—claims that mental representations are useful fictions. People do not really have representations inside their heads, but talking *as if* they do helps us to make sense of their behaviour. Mental fictionalism has many advantages. For one thing, it solves the problem of explaining how mental representations gain their content. (Answer: they don't, because there are none.) And yet mental fictionalism also faces serious challenges. Perhaps most serious is the worry that it suffers from "cognitive collapse" (e.g. Joyce 2013, Wallace 2016, 2022).

The worry about cognitive collapse can be put as follows. Mental fictionalism argues that the inner states posited by folk psychology, like beliefs and desires, do not exist. According to the fictionalist, talk about such states is merely a useful fiction. And yet treating something as a fiction seems to require certain kinds of mental states: we are asked to *imagine* what the fiction says is true or *make-believe* that such and such is the

case. As a result, the critic alleges, mental fictionalism is incoherent: it denies the existence of mental states while assuming the existence of at least some of these states, such as imagination or make-believe. And if the fictionalist grants that *these* states exist, why deny the existence of other mental states? What justifies the unequal treatment? The worry can also be put in terms of representation. The fictionalist claims that inner representations do not exist and that our talk about them is a useful fiction. And yet a fiction is itself a representation: our folk psychological fiction, for example, *represents* people *as* having inner representations. To avoid collapsing like a house of cards, it seems that fictionalism must therefore allow that at least one representation exists and has content. Once again, if fictionalism allows this much, why deny the existence of mental representations in particular? Why the unequal treatment?

In what follows, my aim will be to show how mental fictionalism can avoid this worry about cognitive collapse. To do so, I will first introduce mental fictionalism (Section 2) and then consider its implications for the nature of intentionality (Section 3). My discussion will focus largely on thoughts (e.g. beliefs, desires or intentions), setting aside other aspects of the mind (e.g. sensations or emotions). As we will see, fictionalism's approach to thought and intentionality differs sharply from that taken by the representational theory of mind. It also differs from Daniel Dennett's *intentional stance* (1987), a view that would otherwise seem to have much in common with fictionalism. The key to avoiding cognitive collapse, I will argue, is to see that fictionalism can grant the existence of *public* representations with content, such as written and spoken language (Section 4). In contrast, the notion of *inner* representation found in folk psychology is

deeply problematic. In fact, it is what the early fictionalist Hans Vaihinger called a *real fiction*: it is an idea that is not merely false, but incoherent (Section 5). The upshot is that we have no problem treating the mind *as if* it were an inner world of representations. Indeed, we cannot help but do so. And yet, once we try to take this idea seriously, we find ourselves at a loss to make sense of it.

2 Folk psychology

Let us start once again with Wilfrid Sellars' oft-told myth about the origin of talk about mental states (Sellars 1956). The myth begins with our Rylean ancestors, whose language talks only of overt behaviour. Its hero is Jones, who introduces into this society a new theory that posits internal, psychological episodes he calls *thoughts*. Jones' theory is modelled on overt verbal behaviour: thoughts are said to be like speech in important respects (e.g. they have content), though not all (e.g. there is no inner tongue). Told in this way, Sellars' myth fits with the idea that folk psychology is a proto-scientific theory of our inner world, with mental representations some of the most important inhabitants of that world. On this view, what Jones gives the Ryleans is a powerful new theory to explain and predict people's behaviour. The fictionalist myth is rather different (Toon 2016). It begins with the same Rylean ancestors. In the fictionalist version, though, what Jones gives the Ryleans is not a theory, but a metaphor. He suggests treating people *as if* they had inner episodes analogous to overt speech. In fact, we can imagine that Jones introduces a host of different metaphors, each dedicated to a different aspect of mental life: memory is treated like an inner notebook, reasoning like an inner argument, desire

like an inner shopping list, and so on (Toon, 2021a). This stock of metaphors greatly enriches the Ryleans' language. And yet Jones is careful to point out that these inner episodes (or notebooks, arguments or shopping lists) do not really exist; they are all merely useful fictions for making sense of people's behaviour.

We can develop this idea using Kendall Walton's influential analysis of metaphor in terms of pretence and make-believe (1993). Suppose we remark that "George's money troubles are a heavy burden on him". In Walton's analysis, when we say this we invoke a familiar game of make-believe in which we imagine that someone's problems are physical objects they must carry. This game is governed by implicit rules (e.g. the more serious the problem, the heavier the object). Our utterance is an act of pretence in this game. We do not claim that George is (literally) carrying a heavy object; we only pretend to assert this. And yet, by doing so, we also indicate that pretending in this way is appropriate. What makes it appropriate (or inappropriate)? A whole range of facts about George and his finances (e.g. his bank account is overdrawn; he lies awake at night worrying about it, etc.) When we say "George's money troubles are a heavy burden on him", what we really assert are these facts about George and his financial position. If these facts did not hold (e.g. if George had just won the lottery, or didn't give two hoots about his bank balance), then our pretence would be inappropriate and our assertion false.

The fictionalist understands talk about mental states in a similar way. Consider memory and standing beliefs. Suppose we say that "Tom believes the train to Exeter departs at 12.03". According to the fictionalist, much of our everyday talk about memory is metaphorical: we treat memory *as if* it were a kind of private, inner notebook that guides

our actions. Of course, this doesn't always work—people sometimes forget things or get confused—but much of the time it gives us a valuable means of making sense of people's behaviour. Our utterance about Tom is an act of pretence within this game. We do not claim that Tom (literally) has an inner notebook telling him when the train to Exeter departs; we only pretend to assert this. And yet, by doing so, we indicate that pretending in this way is appropriate. What makes it appropriate (or inappropriate)? A whole range of facts about Tom and his behaviour (e.g. when he sees it is already 12.02, he hurries to the platform; when you ask him what time the train leaves, he says "12.03", etc.). When we say "Tom believes that the train to Exeter departs at 12.03", what we really assert are these facts about Tom and his behaviour. If these facts did not hold (e.g. if Tom calmly sauntered to the waiting room, or replied "12.13" instead), then our pretence would be inappropriate and our assertion false.

All of this means that fictionalism understands talk about the mind rather differently to representationalism. According to the representationlist, when we attribute a belief or desire to someone we are claiming that they have an inner representation that plays a particular role in their mental machinery. According to the fictionalist, we pretend that people have inner representations in order to make claims about their behaviour. Roughly speaking, when we attribute a belief or desire to someone we are claiming that they behave *as if* they had such an inner representation—even though, in fact, they do not. As a result, it is misleading to say simply that fictionalism denies the existence of mental states, or thoughts in particular (Toon, 2021b; cf. Toon, 2016). *If* mental states are presumed to be inner representations, then fictionalism does indeed deny their existence.

But the fictionalist can allow that there are perfectly "real patterns" (Dennett, 1991) in people's behaviour that are described by our folk psychological metaphors. In this sense, fictionalism can be taken to offer an account of what certain mental states are and how they are picked out by folk talk about the mind.

3 Intentionality

What does fictionalism mean for our approach to intentionality? Let us start with the intentionality of thought (Section 3.1) and afterwards consider public representations, such as written and spoken language (Section 3.2).

3.1 Thoughts

The first point to notice is that, for the fictionalist, our concepts concerning intentionality apply, first and foremost, to public representations, especially language. The core intentional notions are semantic categories such as meaning, truth and reference. According to fictionalism, when we talk about mental states, we transfer these semantic categories metaphorically to the mind as an inner realm. Noting that people can say and write down sentences that have meaning—are about the world, are true or false, and so on—we begin to talk as if people also had such things inside their heads guiding their behaviour. An important consequence is that fictionalism pre-supposes that we can grasp these semantic notions without invoking mental states such as thoughts, beliefs and intentions. This commitment is not unique to fictionalism, however. In fact, it lies at the

heart of Sellars' myth. Our Rylean ancestors can engage in overt verbal behaviour before they have any notion of thoughts as inner episodes. It is only afterwards that Jones comes along and uses this overt verbal behaviour as the model for his theory of inner states.

Sellars' myth serves to highlight a further important point: to say that our concepts concerning the intentionality of the mental are based on those concerning public representations is not yet to say that the intentionality of the mental derives from that of public representations. To see this, notice that a representationalist might follow Sellars in thinking that public language serves as the model for our theory of inner states. And yet the representationalist will argue that these inner states *exist*. Using the model, we *discover* that people have inner representations and that it is these representations that explain why people behave as they do—including their ability to use public language. In this way, the representationalist will claim that, even if our concepts concerning public intentionality come first, it is still our inner representations that provide the basis for all intentional phenomena. In a similar manner, someone might first possess a range of concepts relating to everyday objects like billiard balls (e.g. hardness, position, velocity), while knowing nothing whatsoever about atoms or molecules. In due course, however, they might use billiard balls as a model to construct a theory of atoms and molecules and conclude that it is these hidden particles that ultimately explain the behaviour of ordinary objects, including billiard balls.

For the fictionalist, the intentionality of mental states cannot be grounded in the content of mental representations, since she argues that they do not exist. Instead, the intentionality of mental states is grounded in facts about a person's behaviour: it is

because someone behaves in the way that they do that they can be said to possess certain mental states. Of course, the fictionalist claims that we make sense of people's behaviour using metaphors involving public representations, like notebooks. But the fact that we choose to describe someone's behaviour in a certain way does not mean their behaviour depends upon our descriptions. For the fictionalist, a person possesses a given mental state in virtue of exhibiting the relevant pattern of behaviour, not because we pick out that pattern of behaviour using particular metaphors. As long as their behaviour remained the same, they would possess the same mental states—even if there was no one else around to see it. In virtue of their behaviour, Sellars' Rylean ancestors had *minds* even before Jones introduced his remarkable linguistic innovation—although this innovation might also have changed their behaviour in important respects too.

There is an obvious and important complication to this basic picture, however. For the behaviour required to exhibit a given mental state will often involve the use of public representations, especially language. For example, one sort of behaviour associated with believing that the train to Exeter leaves at 12.03 is that, if someone asks you when the train leaves, you'll say "12.03" (or "just after 12", "in a few minutes", etc.). In some situations, you might write the time down instead, or point to a clock, or draw on a train timetable. Of course, this is not to say that mental states can be *reduced* to linguistic behaviour. Fictionalism rejects behaviourism's attempt to reduce talk about the mind into talk about behaviour, whether linguistic or otherwise (Toon, 2016). The point is simply that exhibiting the right pattern of behaviour to count as possessing a particular mental state will often involve using language or other sorts of public representation.

Notice that I say that exhibiting the right pattern of behaviour will often involve the use of language, not that it will always do so. Consider animals and pre-linguistic infants. We often attribute mental states in such cases. We say that a dog knows where it lives or that a baby wants milk, even if neither can tell us as much. Fictionalism has no problem explaining such attributions. Consider the metaphor of memory as a notebook. This metaphor is most apt when applied to creatures that can use language. After all, people who use actual notebooks can typically read their contents if you ask them. But the metaphor can still be useful in other cases. In some respects, it is useful to treat a dog as if it had an inner notebook saying where it lives: it will help you to predict where the dog might end up if it gets lost. In other respects, of course, the metaphor is less apt: unlike someone who had such claims written down in a notebook, the dog cannot tell you its address or point it out on a map. The fictionalist need not insist on a sharp divide here, however. The aptness of any metaphor is a matter of degree (Toon, 2021a; cf. Dennett, 1996, 2013).

The upshot is that fictionalism need not claim that the ability to use language is necessary for the possession of mental states. For some mental states, like knowing where you live or wanting some milk, it is possible to exhibit the right pattern of behaviour without the use of language. For other mental states, however, the ability to use language is necessary. It is hard to know what a dog could do to convince us that it believes that current levels of inequality are a threat to liberal democracy, for example, or how a baby could show us that it wants the government to adopt a Keynesian economic policy. Let us mark this distinction by talking about *language-dependent* and *non-language-dependent* mental

states. There will be considerable debate about which side of this divide certain mental states fall, of course. Can animals without language have intentions? Can they have reasons for their actions? Fortunately, we need not enter into these debates here. The important point is simply that fictionalism need not pre-judge these issues by ruling out the possibility of minds without language.

3.2 Language

How might the fictionalist explain the content of language and other public representations? It is important to acknowledge at the outset that fictionalism does not provide a theory of public intentionality. It is not a theory of meaning. Fictionalism also places constraints on the theory of meaning that we might adopt. In particular, it will rule out any attempt to explain meaning in terms of the content of an accompanying mental state. For example, consider a simple Gricean theory that says that, in uttering *U*, a speaker *S* means *p* if and only if *S* utters *U* intending to produce in a hearer *H* the belief that *p*. Such a theory aims to reduce linguistic meaning to the contents of the mental states of the speaker and hearer. For those who adopt the representational theory of mind, this move is perfectly legitimate, since they will then look to explain the content of the speaker and hearer's mental states in terms of their inner representations. For the fictionalist, however, this move is not available and the theory risks circularity. According to the fictionalist, the speaker and hearer's mental states are grounded not in any inner representation, but in their overall pattern of behaviour. If this behaviour includes linguistic behaviour—if, that is, the relevant mental states are what I have called

language-dependent—then our theory of meaning will be circular: we will have explained an utterance's meaning in terms of the speaker and hearer's mental states and then explained these mental states in terms of the speaker and hearer's utterances.

So fictionalism cannot explain the content of our sentences by appealing to our mental states. At first glance, this might seem like an insurmountable problem. After all, a set of marks on paper just sitting there on the page would seem to be entirely inert or "dead" (Wittgenstein, 1953). If we want to explain how such marks gain their meaning, surely we must appeal to the mental states of people who write them down or write them? The fictionalist can agree that, without people around, marks on paper would indeed be meaningless. It is only because they are taken up and used in certain practices that they come to possess meaning. For the fictionalist, the crucial point is that, when it comes to giving an account of these practices—when it comes to explaining exactly how it is that the use of marks bestows meaning—our explanation must not rely upon the prior content of our mental states. Or, to be more precise, it must not rely upon the content of our language-dependent mental states—those mental states that already rely upon the use of language. Otherwise, our account will be circular. This still leaves a range of alternative explanations open to us, however. In particular, it leaves open the possibility that we might explain how meaning arises from norm-governed social practices.

What might such an explanation look like? Consider the following sketch of one possible account (adapted from Haugeland, 1990). Recall Sellars' mythical society. His "Rylean" ancestors were already able to use language, although this language was impoverished by its lack of psychological terms. According to the fictionalist, these Rylean ancestors

already have (both language-dependent and non-language-dependent) mental states in virtue of patterns in their behaviour, even if Jones has yet to give them the metaphors to pick out these patterns. However, let us now consider a set of earlier "pre-Rylean" ancestors, who lack the ability to use language. Since they cannot use language, the fictionalist must also conclude that the pre-Ryleans lack any language-dependent mental states. She cannot claim, as the representationalist might, that these thoughts are already lodged somewhere inside their heads, just waiting to be said out loud. The fictionalist can allow, however, that the pre-Ryleans have non-language-dependent mental states—those more basic beliefs, desires and intentions that we might be willing to attribute to animals or infants. How might language and meaning arise in such a community?

Suppose that the pre-Ryleans are what Haugeland (1990) calls *conformists*: they tend to act alike, and to encourage others to act alike, rewarding them if they fall into line or punishing them if they don't. Within such a community, *norms* will arise—socially-sanctioned forms of behaviour (*customs* or *practices*). Some practices will involve tools. Since practices are norm-governed, these tools have prescribed *roles*—ways in which they are *supposed* to be used. For example, the role of screwdriver is to turn screws. This is what a screwdriver is *for*. The basic idea is that language itself can be understood a tool within these social practices. Some of the Ryleans' practices involve making certain sounds or marks while they do certain things. Like the screwdriver, these sounds or marks have prescribed roles within these practices: their use can be correct or incorrect, appropriate or inappropriate, and so on. Again, like the screwdriver, the sounds or marks are *for* something. For example, some of them might be for naming things. But there will

be many different language-using practices. To a first approximation, meaning *is* this norm-governed use of sounds and marks.

Of course, this is the briefest possible sketch of such an approach to language. There is much that must be done—and, indeed, has already been done—to develop such an approach (e.g. Wittgenstein, 1953, Brandom 1994). For our purposes, the important point is that fictionalism need not make meaning entirely mysterious. Instead, it presents us with a challenge: if representationalism asks us to naturalise mental representation, fictionalism asks us to naturalise meaning. As we have noted already, fictionalism itself is not an answer to this challenge. Fictionalism is not a story about the pre-Ryleans and how they come to acquire language. Instead, it is a story about the Ryleans: it is a story about how, once a community has acquired language, it might then acquire the idea of mental states. According to fictionalism, this happens when public intentionality is projected back on to a metaphorical inner realm. A community that already uses language as an *external* tool in its social practices—to name things, to make assertions, to ask questions—begins to talk as if it they had such things inside their heads.

3.3 Taking stock

The upshot is an approach to intentionality that differs radically from that taken by the representational theory of mind. For the representationalist, all intentionality stems ultimately from mental representations. For the fictionalist, there is no one source for intentionality. Instead, there are two: behaviour and social norms. The intentionality of the

mental is grounded in behaviour. In its basic form, it can be possessed by creatures without language. In contrast, the intentionality of public representation is grounded in social norms. These two forms of intentionality are fundamentally different, but they are also intricately related. With public representations come more complex forms of behaviour and, therefore, more complex mental states. Also with public representations come the metaphors we use to pick out the patterns of behaviour that ground our mental states. As Haugeland (1990) points out, any approach that admits the existence of two (or more) fundamentally different kinds of intentionality invites the question of what they have in common. Why call them all *intentionality*? For the fictionalist, the answer lies in the metaphors that figure in our ordinary talk about the mind. At the heart of folk psychology is not merely the notion of behaviour, but behaviour *as if* it were governed by inner representations. It is this metaphor that unites these two forms of intentionality, even if they are, at bottom, fundamentally different phenomena.

4 Cognitive collapse

4.1 The problem

We can now return to the problem of cognitive collapse. Let us recall the worry here. The fictionalist says that the inner states described by folk psychology, like belief and desire, do not exist. She also tells us to regard talk about such states as a useful fiction. And yet treating something as a fiction would itself seem to require certain sorts of mental states, like imagination or make-believe. So fictionalism is incoherent. To respond by granting

the existence of some mental states—like imagination or make-believe—while denying reality to the rest of the mind seems like an arbitrary and rather desperate attempt to get out of jail free.

It might be tempting to dismiss this problem out of hand. As we noted in Section 2, it is misleading to say simply that fictionalism denies the existence of mental states—at least as I understand the view. Although she denies the existence of inner representations, the fictionalist allows that there are patterns in our behaviour that render our attributions of mental states true or false. In this sense, mental states are perfectly real. If that's right, perhaps there is no difficulty in the fictionalist appealing to imagination or make-believe after all? Sadly, things are not that easy. Although this response is correct as far as it goes, it would leave fictionalism incomplete as an approach to the mind. In particular, it would mean that it could not be applied to imagination or make-believe. Fictionalism does not merely say that our attributions of mental states are grounded in behaviour. It offers an analysis of *how* this takes place. According to the fictionalist, we talk about behaviour *via* the fiction of inner representations. If our analysis of this process involves imagination or make-believe, then it cannot make sense of our attributions of *these* mental states themselves, or else it will be circular. The upshot is that, even if the fictionalist is entitled to assume the *existence* of mental states such as imagination or make-believe without fear of incoherence, she could not explain our talk about them. The fictionalist analysis would have to be abandoned for such states. And if we were to abandon fictionalism for imagination or make-believe, the critic might insist, why not abandon it across the board? Once again, why treat these states differently?

The worry about cognitive collapse mirrors a well-known objection to eliminativism. Like the fictionalist, the eliminativist argues that the inner states described by folk psychology do not exist. And yet, the critic objects, asserting something involves believing it. So eliminativism is incoherent: the very act of asserting the position shows that it to be false. Eliminativists are able to offer a compelling response to this objection, however. Churchland (1981) argues that the charge of self-refutation begs the question: it assumes that we must explain what happens when someone makes an assertion (puts forward an argument, defends a position, etc.) in folk psychological terms. And yet this is exactly what the eliminativist denies. Eventually, according to eliminativism, we will come to possess a proper neuroscientific theory of activities such as asserting a position, putting forward an argument, and so on and this theory will find no place for the categories of folk psychology.

The challenge facing mental fictionalism is more troubling, however. The incoherence facing eliminativism is alleged to arise not so much from the *content* of the eliminativist position as from the *act* of asserting it (Joyce 2013). Eliminativism is the claim that mental states do not exist. In itself, this position does not assume the existence of mental states; indeed, its sole claim is that they do not exist. Instead, the trouble is supposed to arise because the possibility of asserting any claim whatsoever—whether about philosophy or the weekend's football results—is said to require the existence of mental states. The eliminativist can justly reply that this assumption begs the question. The challenge facing fictionalism, is more worrying, however, since the content of the fictionalist position *does* seem to assume the existence of mental states. The fictionalist

does not simply claim that inner psychological states do not exist; she also tells us to regard talk about such states as a useful fiction. It is this further claim that generates the worry of inconsistency, for it seems to assume the existence of particular sorts of mental states, such as imagination or make-believe.

In this respect, the objection facing fictionalism is closer to that often levelled against Dennett's intentional stance (Bennett and Hacker 2003, p. 426, Adams and Aizawa 2001, 49). According to Dennett, when we attribute beliefs and desires to people, we do not claim that people have states bearing these contents inside their heads. Instead, we adopt the intentional stance: we attribute those beliefs and desires that allow us to make sense of their behaviour. Here critics detect a problem. After all, adopting the intentional stance towards another person (or creature or object) would itself appear to be an intentional act: it is an interpretation that we use to make sense of their behaviour. The result is that Dennett is charged with much the same sort of incoherence as the fictionalist: the intentional stance would seem to assume the existence of exactly the phenomena whose existence it denies. As Adams and Aizawa (2001) put it, "[t]he content of Mike's attitude seems to depend on Ike's attitude, but whence comes the content of Ike's attitude?"

4.2 Public pretence

To see how fictionalism can avoid cognitive collapse, we can begin by recalling a point we have already discussed in Section 3—namely, that fictionalism can acknowledge the existence of external, public representations with content. As we saw, the fictionalist must

deny that the content of such representations stems from the prior content of mental states. Instead, the most promising alternative looks to the role that these representations play within norm-governed social practices. Fictionalism has little else to say about this public form of intentionality; its scope is limited to the intentionality of the mental. It is here that fictionalism would seem to depart from Dennett's view. As I understand it, the intentional stance is intended to provide the whole story about intentionality. For instance, Dennett (2009, p. 345) rejects Robert Brandom's (1994) claim that only social creatures are capable of genuine belief. Instead, Dennett envisages a continuous spectrum of cases, from thermostats to Sherlock Holmes, with "no theoretically motivated threshold distinguishing the 'literal' from the 'metaphorical', or merely 'as if', cases" (e.g. 2009, p. 343). As we saw in Section 3.1, the fictionalist can agree with Dennett that the intentionality of the *mental* is indeed a continuous spectrum: there is no clear line beyond which our metaphors cease to apply. For the fictionalist, however, the intentionality of *public* representations stands apart: in this sense, fictionalism acknowledges that the emergence of social norms ushers in a new, and fundamentally distinct, form of intentionality. It is these 'literal' cases of intentionality that are projected back to yield the 'as if' intentionality of mental states.

How does this allow fictionalism to avoid cognitive collapse? The key point to notice is that, perhaps somewhat surprisingly, the fictionalist's analysis need not refer to imagination, make-believe or any other mental states. Instead, it relies upon the notion of *pretence*. Walton's analysis of metaphor relies on the idea that, in "pretending to say one thing, one may actually be saying, asserting, something else" (Walton 2000, p. 95). In the

case of folk psychology, in pretending to describe inner representations, we are actually saying something about behaviour. Does the notion of pretence itself pre-suppose that of mental states? Does pretending that *p* is true involve imagining *p*, for example, or believing that *p* isn't true? Not necessarily. An actor playing the lead in J. B. Priestley's *An Inspector Calls* at the end of a long tour might well have ceased to imagine anything much as he appears onstage. In quiet moments, perhaps he's thinking about his plans for the weekend, or reminding himself to pick the kids up from school. And yet surely he is still pretending to be the mysterious Inspector Goole. In moments when he does find himself attending to the content of his pretence, he might well believe some of it too. For example, perhaps he agrees with Goole's parting sentiment that we are responsible for the fate of others and that "if men will not learn that lesson, then they will be taught it in fire and blood and anguish" (Priestley, 2000, p. 207).

Instead of focusing on its supposed connections with mental states like imagination or disbelief, the best way to understand pretence is as a public, rule-governed activity found in distinctive social practices, like putting on plays. As we saw in Section 3, Sellars' myth already assumes that our concepts concerning public intentionality are prior to those concerning mental states. Our Rylean ancestors can talk before Jones comes along; they already possess concepts of meaning and truth, as well as related ideas such as assertion, questioning, promising, and so on. To avoid cognitive collapse, the fictionalist must insist that the Ryleans could also engage in pretence. As well as making assertions or asking questions, they could also play games and tell stories. If this is an embellishment of the myth, it is a fairly minor one. Of course, it must be conceded straightaway that

fictionalism does not offer a theory of pretence, much in the same way that it does not offer a theory of meaning or assertion. But that is hardly surprising. It simply reminds us that fictionalism is not a general theory of intentionality. In other words, it shows that fictionalism is incomplete, not that it is incoherent.

5 Real fictions

In essence, I have suggested we can avoid cognitive collapse by distinguishing between external and internal representations: while the former exist, the latter do not. Crucially, the fictionalist's key notion (pretence) falls into the former category. All this invites the question: why the unequal treatment? Why think external representations are perfectly respectable, while inner ones are merely fictions? Is this all a rather convenient ploy to get the fictionalist out of trouble?

Here it is helpful to introduce a distinction made by the early fictionalist Hans Vaihinger in his *The Philosophy of "As If"* (1924). In Vaihinger's terminology, *fictions* are claims that are false and taken to be false by those that use them. Vaihinger distinguishes between what he calls *semi-fictions* and *real fictions*: while semi-fictions are false and known to be so, real fictions are not merely false but incoherent. If we ignore the effects of friction, for instance, we invoke a semi-fiction: we know our assumption is false, but there is nothing incoherent in the idea of an object moving without encountering friction. Other fictions are more puzzling. Discussing atoms in nineteenth century physics, Vaihinger writes:

> "If […] we designate the atoms as centres without extension, we are merely creating substantial basis for the relationships of force, a basis that, upon more accurate scrutiny, turns out to be a very strange construction indeed. For an entity without extension that is at the same time a substantial bearer of forces—this is simply a combination of words with which no substantial meaning can be connected" (1924, p. 219).

For Vaihinger, the notion of atoms as centres without extension is a real fiction. When we stop to think about it, we simply cannot make much sense of the idea that all the mass of an object could be located at a point without extension. And yet this does not prevent this idea from playing an important role in physical theory. As Vaihinger puts it, "the concept in question is contradictory, but necessary" (1924, p. 72; for further discussion of Vaihinger's ideas, see Fine 1993, Suarez 2009 and Appiah 2017).

Why does the fictionalist regard inner representations as fictions? In some cases, it is easy to tell whether we're dealing with a fiction: if we run our hand across the lab bench, we can feel there *is* friction—we just think we can safely ignore it. We certainly cannot simply look and see that people do not have representations inside their heads. The reason we should regard mental representations as fictions, I suggest, is that they are real fictions, in Vaihinger's sense: the notion of inner representation that we find in folk psychology is not merely false, but incoherent. Our ordinary concept of representation concerns external, public forms of representation that represent through social conventions, like spoken and written language, maps or diagrams. And yet it is clear that mental representations cannot be representations in this sense: they are supposed to be

locked away inside people's heads and are certainly not subject to any social conventions. For the fictionalist, mental representations are much like point masses. Talking as if the mind were an inner world containing such representations is an enormously productive way of making sense of people's behaviour. And yet, when we stop to think about it, we see that the idea that people could really have such things inside their heads makes little sense.

In this context, we might recall Wittgenstein's famous remark that "only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious" (1953, para. 281). This remark has inspired Maxwell Bennett and Peter Hacker (2003) to argue that neuroscientists' talk about the brain containing beliefs, knowledge or, indeed, inner representations is simply incoherent. After all, a brain does not resemble a living human being. On this central point, the fictionalist can agree. Notice, however, that immediately making this remark, Wittgenstein considers an objection: "'But in a fairy tale the pot too can see and hear!'" (1953, para. 282). If we tell stories like this, doesn't this show that we *can* make sense of attributing mental states to inanimate objects after all? The same point can be made about mental representations. Philosophers often dream up thought experiments about creatures with sentences or pictures inside their heads (e.g. Sprevak 2010). Indeed, the fictionalist must rely on this fact: we must be able to *pretend* that people possess such inner representations, even if they do not. If we can engage in this pretence, doesn't this show that talk about mental representations *is* coherent after all?

This would be too quick. In the first place, notice that a fairy tale will usually have us

imagine that a pot can see or hear by making it *behave* like a human being in certain respects—it might be able to shout or run away if someone tries to fill it with hot water (Wittgenstein, 1953, para. 282; see also McGinn, 1997). If *this* is what we imagine, then the fairy tale will not show that it makes sense to attribute mental states to inanimate objects, since the pot is not inanimate. A similar lesson applies in the case of mental representations. When we are asked to imagine a creature with inner representations, we often find ourselves imagining an inner agent (or "homunculus") who *reads* these representations. If *this* is what we imagine, then such thought experiments do not show that it makes sense to talk about mental representations. This scenario might well be coherent, but it is not what the representationalist needs. After all, most will not want to countenance inner homunculi.

More importantly, though, it is simply false to say that, whenever we engage in pretence, our pretence must be coherent—if, by "coherent", we mean that it must make sense if taken literally. In fact, we often have little idea what it would mean to take our pretence literally. Consider the fairy tale with the talking pot. Wittgenstein asks, "Is it false or nonsensical to say that a pot talks? Have we a clear picture of the circumstances in which we should say of a pot that it talked?" (1953, para. 282). Arguably, we do not. Our use of language in fairy tales stands somewhat apart from that in ordinary life, and it can be difficult to see what it could mean to take it literally. Think of children's games. In the midst of playing a game, children will suddenly find they can perform amazing feats of magic, turn themselves into ghosts or monsters, disappear into thin air, travel back in time or turn into a rocket and launch themselves into space. In many cases, it is hard to see

what it could mean to take these ideas literally. And yet we are still able to engage in these sorts of pretence perfectly well. Indeed, even small children can play games like these.

The same is true in the case of metaphors. Some metaphors can be understood literally. People can literally, as well as metaphorically, carry a heavy burden. Many metaphors do not make sense if taken literally, however (Walton 2000, p. 96). Suppose someone remarks that the clouds are angry today. It is hard to know what it could mean to take this literally. Our ordinary concept of anger simply does not apply to objects like clouds. And yet we can still use this idea figuratively. Saying that clouds are metaphorically, rather than literally, angry is perfectly meaningful. It can even help us to pick out genuine facts about the state of the weather. Many metaphors that we use to describe the mind are like this: they simply do not make sense if taken literally. Chief amongst these, I suggest, is talk about inner representations. Our ordinary concept of representation simply does not apply to representations inside the head. And yet we can still use this idea figuratively. Our metaphor is perfectly meaningful. As we have seen, it can even help us to pick out genuine facts about people's behaviour.

For the fictionalist, the important point is that, even if the notion of inner representations cannot be understood literally, we can still pretend that people have such things. The notion of representation has a literal use when applied to external forms of representation, like notebooks, maps or to do lists. It also has a metaphorical use when applied to the mind as an inner world. Both of these uses are perfectly legitimate. The trouble arises only when we begin to confuse them. Three caveats are in order at this point, however.

First, our discussion has focused on the notion of mental representation as it appears in ordinary talk about the mind. It is this notion that I have argued is a real fiction. The fictionalist need not deny that proponents of the representational theory of mind might one day develop a new, technical notion of representation that does not suffer from this incoherence. In a sense, of course, this is precisely what representationalists have tried to do: they have tried to show how inner representations might gain their content through some other means, such as causal relations or evolutionary history. Up till now, this project has not been successful, but we should not rule out the possibility that it might succeed one day. Even if this project is ultimately successful, however, we might still ask what this technical notion has in common with our ordinary concept of representation, or what the existence of such representations tells us about our ordinary concept of mind. Certainly, if the fictionalist's analysis of our ordinary talk about the mind is along the right lines, it is debatable whether these inner representations—even if they do exist—would count as beliefs or desires.

Second, even if the notion of mental representation cannot be taken literally, this does not mean that talk about them cannot do useful work, even in scientific contexts. Opponents of mental representation often assume that, if talk about them cannot be true, then they can play no useful role in cognitive science—at least, no useful explanatory role (e.g. Bennett and Hacker 2007, p. 140; cf. Dennett, 2007). And yet *all* scientific models make assumptions that we know are not true, while others invoke ideas that are hard to take literally, like point masses, massless springs or infinite potential wells. Of course, the fictionalist can agree that talk of inner representations must be handled with care: we

must not take our metaphors *too* seriously. Still, we must take them seriously enough if we are not to overlook the vital role they play in much of our thinking. Surveying debates over the notion of the atom in nineteenth-century physics, Vaihinger writes,

> The defence was always anxious to show that the alleged contradictions were only apparent and that the concept therefore possessed objective validity and could be applied. Their opponents, on the other hand, demonstrated the contradictions and so refused to allow the concept any legitimate place in science; in other words, they poured out the baby with the bath, while the defence accepted it—un-washed (1924, p. 72).

Much the same could be said about contemporary debates over mental representation.

Third, notice that fictionalism—at least as I've presented the view—need not imply any wholesale opposition to realism. In particular, it need not impose a blanket ban on inference to the best explanation (cf. Sprevak 2013). I've suggested that there is a particular reason why the success of folk psychology should not lead us to infer the existence of mental representations: these inner representations are real fictions and so our folk stories about them cannot be true. This need not threaten inference to the best explanation, which requires that the best explanation is *good enough* to warrant our inference (Lipton 1991). At a minimum, this requires that it is coherent. This is not to say that fictions cannot explain but only to say that, if they are incoherent, they cannot even be candidates for inference to the best explanation. After all, inference to the best explanation tells us to infer what would, *if true*, provide the best explanation for the

evidence—and real fictions cannot be true (Levy, 2018). So fictionalism need not reject inference to the best explanation. It does, however, reject any defence of representationalism that appeals merely to the success of folk psychology. It is often said that, since folk psychology is successful, we can be confident that mental representations exist, even without any naturalistic account of mental representation in hand. This is too quick. Before we can appeal to the success of folk psychology to argue for the existence of mental representations, we must first show that we are not dealing with a real fiction. After all, our physical theories might enjoy any number of successes, but we do not take this to demonstrate the existence of point masses or massless springs.

Conclusion

Fictionalism is popular in many areas, from mathematics to morality. Mental fictionalism has few adherents, however. Much of the blame lies with the problem of cognitive collapse: if mental fictionalism assumes the existence of the very thing that it brands a fiction, it is hard to see how it could even to get off the ground. I hope to have shown that the situation is not as dire as it seems. Properly understood, mental fictionalism suggests a new approach to intentionality that is coherent, if incomplete. In the beginning was the deed. After that, came the word, which brought with it new deeds and new ways to describe them. Our ordinary talk about the mind is a metaphorical mapping of words onto deeds: we talk as if people had inner representations in order to make sense of their behaviour. Fodor (1975) famously defended representationalism on the grounds that it is the only game in town. In a sense, the fictionalist can agree. The trouble is that, like many games, representationalism is hard to take seriously.

Acknowledgements

An earlier version of this chapter was presented at the conference on *Mental Fictionalism* held at the Hungarian Academy of Sciences, Budapest, October 24th to 25th, 2019. I am very grateful to all the participants at the conference for making it such an enjoyable occasion. I learned a great deal from our discussions and I hope it has enabled me to improve some of the ideas presented here. I am especially grateful to Tamás Demeter and Ted Parent, for helping to organise the conference, and to Amber Ross, for insightful and constructive comments on an earlier draft of this paper. Ideas from the chapter were also presented at the *Reconceiving Cognition* conference, Antwerp, June 27th to 29th 2018, the Institute of Philosophy, London, March 12th 2019 and the 93rd Joint Session of the Aristotelian Society and the Mind Association, Durham, July 19th to 21st 2019. My thanks to audiences at each of these events for helpful comments and criticism.

References

Adams, F. & Aizawa, K. (2001). The bounds of cognition, *Philosophical Psychology*, 14 (2001), 43–64.

Appiah, K. A. (2017). *As If: Idealization and Ideals*. Cambridge, MA: Harvard University Press.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Bennett, M. R., & Hacker, P. M. S. (2007). The Conceptual Presuppositions of Cognitive Neuroscience: A Reply to Critics. In M. Bennett, D. Dennett, P. Hacker and J. Searle (Eds.) *Neuroscience and Philosophy: Brain, Mind, and Language* (pp. 127-162). New York: Columbia University Press.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.

Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67–90.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. (1991). Real Patterns. *The Journal of Philosophy*, 88(1), 27–51.

Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.

Dennett, D. C. (2007). Philosophy as Naïve Anthropology: Comment on Bennett and

Hacker. In M. Bennett, D. Dennett, P. Hacker and J. Searle (Eds.) *Neuroscience and Philosophy: Brain, Mind, and Language* (pp. 73-96). New York: Columbia University Press.

Dennett, D. (2009). Intentional systems theory. In B. P. McLaughlin, A. Beckermann & S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind* (pp. 339–350). Oxford: Oxford University Press.

Dennett, D. (2013). *Intuition Pumps and Other Tools for Thinking*. London: Penguin.

Fine, A. (1993). Fictionalism. *Midwest Studies In Philosophy*, 18(1), 1–18.

Fodor, J., 1975. *The Language of Thought*. Cambridge, MA.: Harvard University Press.

Haugeland, J. (1990). The Intentionality All-Stars. *Philosophical Perspectives*, *4*, 383–427.

Joyce, R. (2013). Psychological Fictionalism, and the Threat of Fictionalist Suicide. *The Monist*, 96(4), 517–538.

Levy, A. (2018). Modeling and realism: strange bedfellows? In J. Saatsi (ed.) *The Routledge Handbook of Scientific Realism*. Abingdon, Oxfordshire: Routledge

Lipton, P. (2004). *Inference to the Best Explanation*. Abingdon, Oxfordshire: Routledge. (1st Edition 1991)

McGinn, M. (1997). *The Routledge Guidebook to Wittgenstein and the Philosophical Investigations*. Abingdon, Oxfordshire: Routledge.

Priestley, J. B. (2000). *An Inspector Calls and Other Plays*. London: Penguin. (*An Inspector Calls* was first published in 1947).

Sellars, W. (1956). Empiricism and the Philosophy of Mind. In H. Feigl & M. Scriven (Eds.) *The Foundations of Science and the Concepts of Psychology and Psychoanalysis. Minnesota Studies in the Philosophy of Science, Vol. 1*. Minneapolis: University of Minnesota Press.

Sprevak, M. (2013). Fictionalism About Neural Representations. *The Monist*, 96(4), 539–560.

Suárez, M. (Ed.) (2009). *Fictions in Science: Philosophical Essays on Modeling and Idealization*. London: Routledge.

Toon, A. (2016). Fictionalism and the folk. *The Monist*, 99, 280–295.

Toon, A. (2021a). Minds, materials and metaphors. *Philosophy,* 1-23. doi:10.1017/S0031819120000406

Toon, A. (2021b) 'The Story of the Ghost in the Machine'. In S. Sedivy (ed.) *Art, Representation and Make-Believe: Essays on the Philosophy of Kendall L. Walton.* Abingdon, Oxfordshire: Routledge.

Vaihinger, H. (1924). *The Philosophy of 'As If': A System of the Theoretical, Practical, and Religious Fictions of Mankind*. Abingdon, Oxford: Routledge.

Wallace, M. (2022). Mental Fictionalism. In T. Demeter, T. Parent and A. Toon (Eds.) *Mental Fictionalism: Philosophical Explorations* (pp. 27-51). London: Routledge.

Wallace, M. (2016). 'Saving Mental Fictionalism from Cognitive Collapse'. *Res Philosophica*, 93(2), 405-424.

Walton, K. (1993). ‘Metaphor and Prop Oriented Make-Believe’. *European Journal of Philosophy*, 1(1), 39–57.

Walton, K. L. (2000). ‘Existence as metaphor?’ In Everett, A. & Hofweber, T. (Eds.) *Empty Names, Fiction, and the Puzzles of Non-Existence*. CSLI Publications (pp. 69-94). Reprinted in Walton, K., *In Other Shoes: Music, Metaphor, Empathy, Existence*. New York: Oxford University Press (pp. 89-113). (Page numbers cited in text refer to reprinted version.)

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell. (Translated by G. E. M. Anscombe)