

[Manuscript version. Final paper to appear in *The Monist* and available online at

<http://dx.doi.org/10.1093/monist/onw005>]

## **Fictionalism and the folk**

Adam Toon  
University of Exeter  
a.toon@exeter.ac.uk

### **Abstract**

Mental fictionalism is the view that, even if mental states do not exist, it is useful to talk as if they do. Mental states are useful fictions. Recent philosophy of mind has seen a growing interest in mental fictionalism. To date, much of the discussion has concerned the general features of the approach. In this paper, I develop a specific form of mental fictionalism by drawing on Kendall Walton's work on make-believe. According to the approach I propose, talk of mental states is a useful pretence for describing people and their behaviour. I try to clarify and motivate this approach by comparing it to well-known alternatives, including behaviourism, instrumentalism and eliminativism. I also consider some of the challenges that it faces.

### **1 Introduction**

How should we make sense of ordinary talk about the mind? When we say that someone has a particular belief or desire, are we claiming that they have a certain sort of causal state inside their heads? If we are, might future cognitive science show that we are wrong? Might it turn out that mental states, like beliefs and desires, do not exist? In this

paper, I explore one way to approach these longstanding questions concerning the status of folk psychology: *mental fictionalism*. Put simply, mental fictionalism is the view that, even if mental states do not exist, it is useful to talk as if they do. Mental states are useful fictions.

Recent philosophy of mind has seen a growing interest in mental fictionalism (e.g. Wallace 2007, Demeter 2013a). To date, much of the discussion has concerned the general features of this approach and the difficulties it faces.<sup>1</sup> In this paper, I develop a specific form of mental fictionalism by drawing on Kendall Walton's work on make-believe (Walton 1990, 1993). According to the account I will put forward, when we say someone has a particular belief or desire, we are not making a claim about their inner machinery. Instead, talk about mental states is a useful *pretence* for describing people and their behaviour. As a result, the legitimacy of ordinary talk about the mind does not depend upon whether future cognitive science will discover beliefs or desires inside our heads.

The structure of the paper is as follows. First, in Section 2, I give a brief outline of Walton's theory, focusing on his discussion of metaphor and prop-oriented make-believe. In Section 3, I show how we may draw on Walton's analysis to develop a fictionalist analysis of folk psychology. In Section 4, I further clarify and motivate this account by comparing it to a number of well-known alternative positions, including behaviourism, instrumentalism and eliminativism, as well as alternative forms of fictionalism. Finally, in Section 5, I consider some important challenges facing mental fictionalism.

---

<sup>1</sup> An exception is Demeter (2013b), who proposes a particular form of mental fictionalism that treats folk psychology as device for expressing an interpreter's affective reactions.

## 2 Prop-oriented make-believe

Imagine some children playing with a doll. They pick the doll up and hold it in their arms, push it down the hallway in a pushchair, cover it with a blanket, and so on. As they do so, the children imagine themselves picking up a baby and cradling it, taking it to the shops and lying it down in its cot to go to sleep. In Walton's theory, dolls and other objects used in make-believe are called *props* and the rules that govern their use in the game are called *principles of generation* (Walton 1990). Within a game of make-believe, the properties of the props, together with the relevant principles of generation, make propositions *fictional*. To say that a proposition is fictional, in Walton's sense, is simply to say that participants in the game are prescribed to imagine it. For example, if the doll's "eyes" are closed, the children are to imagine that the baby is asleep. It is fictional that the baby is asleep.

Notice that, since the content of a game of make-believe depends only on the props and principles of generation, it possesses a certain kind of "objectivity": if the doll's eyes are closed then it is fictional that the baby is asleep, even if none of the children happen to notice this.

One important feature of games of make-believe is *participation*. Children playing with a doll do not simply sit and look at it. Instead, they carry out various actions with the doll (e.g. putting a bottle to its "mouth") and these actions generate imaginings within the game (e.g. that the children are feeding the baby). Like the doll, the children themselves become props in the game. The children also participate verbally. For example, David might sing a lullaby as he rocks the doll in his arms, thereby making it fictional that he is singing a lullaby to get the baby to sleep. Importantly, acts of pretence can be used to make genuine assertions. Suppose that Anna looks at the doll and says "She's sleeping now!" When she says this, Anna pretends to assert that the baby is asleep. But she also indicates that pretending in this way is appropriate and, in doing so, makes a genuine assertion: she claims that the state of the props is such that to pretend in the way that she does is, fictionally, to speak the truth. In other words, Anna claims that the doll's "eyes" are closed. Once again, notice the objectivity of games: if, in fact, the doll's "eyes" are

open, then Anna's pretence is inappropriate and her assertion is false.<sup>2</sup>

The reason that children play with dolls, it seems, is not because they are interested in the properties of the doll as such. Dolls are interesting only in so far as they allow the children to immerse themselves in a make-believe world in which they can cradle a baby, take it out in its pushchair, feed it or rock it to sleep. Walton (1993) calls this kind of make-believe *content oriented*. He distinguishes it from a rather different use of make-believe, which he calls *prop oriented* (*ibid.*). Walton's central examples of prop oriented make-believe are cases of metaphor or figurative language:

Where in Italy is the town of Crotone?, I ask. You explain that it is on the arch of the Italian boot. 'See the thundercloud over there – the big angry face near the horizon,' you say; 'it is headed this way'. Plumbers and electricians distinguish between 'male' and 'female' plumbing and electrical connections. We speak of the saddle of a mountain and the shoulder of a highway.

All of these cases are linked to make-believe. We think of Italy and the thundercloud as something like pictures. Italy (or a map of Italy) depicts a boot. The cloud is a prop which makes it fictional that there is an angry face. Male and female plumbing or electrical connections are understood to be, fictionally, male and female sexual organs. The saddle of a mountain is, fictionally, a horse's saddle. But our interest, in these instances, is not in the make-believe itself, and it is not for the sake of games of make-believe that we regard these things as props. [...]

Make-believe [...] is useful in these cases [...] for articulating, remembering, and communicating facts about the props – about the

---

<sup>2</sup> Note that here I adopt a pragmatic, rather than semantic, interpretation of Walton's account (for more on this distinction, see Friend 2007).

geography of Italy, or the identity of the storm cloud, or functional properties of plumbing or electrical fixtures, or mountain topography. It is by thinking of Italy or the thundercloud or plumbing connections as potential if not actual props that I understand where Crotone is, which cloud is the one being talked about, or whether one pipe can be connected to another. (Walton 1993, pp. 40-41)

In content oriented make-believe, our interest lies in the content of a make-believe world. By contrast, in prop oriented make-believe, our interest lies in the props themselves; the role of make-believe is to help us to understand the props. Despite the difference between these forms of make-believe, Walton's analysis of utterances involving prop oriented make-believe parallels his analysis of utterances in the children's game. If Mark says "Crotone is on the arch of the Italian boot", he is involved in pretence, much like Anna when she looks at the doll's eyes and exclaims "She's sleeping now!". And yet, just like Anna, Mark also makes a genuine assertion: he claims that the state of the props is such that to pretend in the way that he does is, fictionally, to speak the truth. In other words, Mark asserts that Crotone is in such-and-such a position on the Italian coastline. Invoking a familiar game of make-believe in which Italy is imagined to be a boot provides him with a more colourful and memorable way of communicating this fact.

Walton's theory has been influential amongst fictionalists in many domains (e.g. Crimmins 1998, Joyce 2005, Kroon 2001, Yablo 2001). In the next section, I show how it can be used to develop a form of mental fictionalism.

### **3 Folk psychology as fiction**

I now want to suggest that we should understand ordinary talk about mental states in terms of pretence. One way to introduce this idea involves a twist on Wilfrid Sellars' famous myth about the origin of talk about mental states (Sellars 1956; cf. Yablo 2005). Sellars asks us to imagine a society who at first use only what he calls a "Rylean language" that is restricted to terms referring to overt behaviour. At some point, along

comes a visionary theorist, called Jones, who develops a theory of internal, psychological episodes, which he dubs *thoughts*. Jones bases his theory of thoughts on the model of overt verbal behaviour. However, like all models, Jones' model is accompanied by a commentary specifying which of the features of its source domain (overt verbal behaviour) are to be carried over to its target domain (internal psychological states). For example, Jones is clear that his theory doesn't claim that there is any hidden tongue that produces inner speech or that inner speech can be heard. Using his theory, Jones is then able to predict and explain the complexity of people's behaviour in a way that the Rylean language could not.

With Sellars' story in mind, I want to propose an alternative, fictionalist, myth. According to this myth, what Jones introduces to our Rylean ancestors is not a theory, but a useful game of prop oriented make-believe. Jones invents a game in which we are supposed to imagine that people undergo inner episodes, called thoughts, which are analogous to overt verbal behaviour. These inner episodes respond to experience, interact in certain ways and cause certain behaviours. According to this alternative myth, Jones still takes overt verbal behaviour as his model for thoughts. But the commentary he offers on this model is very different from that envisaged by Sellars. In the fictionalist telling of the tale, Jones no more claims that he is committed to the existence of these episodes of inner speech than he claims that they are uttered by a hidden tongue. Instead, the entire model is proposed merely as a useful metaphor for describing people and their behaviour. Thoughts are not theoretical entities of a new theory of the mind, but useful fictions.

We can now fill out our fictionalist myth using Walton's analysis. Folk psychology offers a fruitful and productive game through which we can make sense of people and their behaviour, in much the same way that the Italian boot is a useful game for understanding the geography of Italy. Ordinary talk about mental states should be understood along the same lines as an utterance like "Crotone is on the arch of the Italian boot". As we saw in Section 2, on Walton's analysis, when Mark says this he is not making a straightforward assertion; his utterance is an act of pretence. And yet, by pretending in this way, he does make a genuine assertion, namely that the location of Crotone is such that, fictionally, he

speaks the truth. I suggest that we understand ordinary talk about mental states in a similar manner. Suppose that Ruth says “John believes that planes can fly”. When she says this, I suggest, Ruth is not making a straightforward assertion; her utterance is an act of pretence. And yet, by pretending in this way, Ruth does make a genuine assertion, namely that John is in some state *S* such that, fictionally, she speaks the truth. Invoking the game in which Italy is imagined to be a boot provides Mark with a useful means for describing the location of Crotona. Similarly, invoking the game of folk psychology provides Ruth with a useful means for describing John and his behaviour.

On this view, then, ordinary talk about the mind involves acts of pretence within a folk psychological game of make-believe. Within this game, we are to imagine that people have certain inner states inside their heads, such as beliefs and desires. We are also to imagine that these states arise in certain circumstances, interact in certain sorts of ways, and produce certain sorts of behaviour. We can capture some of the rules of the folk psychological game in various truisms. If someone is looking at an object in good light, then normally we are to imagine that they have a particular inner state (a belief) which says that there is an object in front of them. If someone hasn’t eaten anything all day, then normally we are to imagine that they have a different sort of inner state (a desire) which says they should eat something. In general, however, the rules of the folk psychological game are notoriously difficult to specify. This need not pose a particular problem for the fictionalist, however. After all, the rules governing children’s games are rarely explicitly formulated.

Why might someone be attracted to mental fictionalism? The most familiar motivation for fictionalism is a worry about ontological commitment (Wallace 2007). Thus, the fictionalist might be persuaded by various arguments for eliminativism about mental states (e.g. Churchland 1981, Ramsey, Stich and Garon 1990). And yet she might also worry that we cannot do without folk psychological talk. The fictionalist therefore aims to show how we can keep such talk around without being committed to the existence of mental states. There is also a rather different motivation for turning to mental fictionalism, however. Rather than being persuaded by arguments for eliminativism, the

fictionalist might primarily be interested in making sense of our ordinary talk about the mind. When we examine such talk closely, she might argue, we find that it is best understood in fictionalist terms: despite appearances to the contrary, in fact the folk are *not* committed to the existence of beliefs and desires as discrete inner causes of behaviour; instead, folk discourse treats these entities as useful fictions. A fictionalist who takes this approach endorses a *hermeneutic*, rather than *revolutionary*, fictionalism (Stanley 2001): she presents fictionalism as a descriptive claim about folk discourse, rather than a prescription for how that discourse should be reinterpreted.

Tentatively, I wish to propose mental fictionalism as a hermeneutic fictionalism. Of course, folk discourse is complex and nuanced, and it might be difficult to capture this complexity in any single account (Dennett 1987, Godfrey-Smith 2005). Nevertheless, I want to suggest that a fictionalist analysis captures the spirit of much of our ordinary talk about mental states. The guiding motivation behind this approach is familiar from behaviourist or instrumentalist approaches in general: the legitimacy of ordinary talk about the mind does not depend upon speculation about the machinery inside people's heads. The fictionalist analysis of ordinary folk discourse requires further defence, however. After all, it might be argued, when we attribute beliefs and desires to people, we don't normally feel as if we are pretending or engaging in a game of make-believe. I return to this objection in Section 5. First, however, I hope to clarify and motivate the fictionalist view in more detail by comparing it to a number of well-known alternative approaches to folk psychology.



## 4 Contrasts

### 4.1 Behaviourism

First, let us compare fictionalism to behaviourism.<sup>3</sup> Like fictionalism, behaviourism also claims that ordinary talk about the mind is not committed to the existence of beliefs and desires as inner causal states. However, the behaviourist tries to avoid commitment to such internal states by giving a reductive analysis of mental talk in terms of behaviour. Thus, to say that John wants to go to Madrid is not to describe any state inside his head; it is to describe his tendencies in various circumstances: if he were browsing through travel guides, he would reach for the guide to Madrid; if he were in the airport, he would catch the plane for Madrid, and so on. The key problem for behaviourism, of course, is that filling out such analyses proves to be extremely difficult. The main reason for the failure of this project, it seems, is that there is no straightforward, one-to-one correspondence between individual mental states and behavioural dispositions. Instead, the behaviour that results from any individual mental state typically depends on many other mental states. John's desire to go to Madrid will only lead him to catch a plane for Madrid if he also believes that planes fly, that the plane at the gate will take him to Madrid, and so on.

Along with the behaviourist, the hermeneutic fictionalist also argues that folk psychology is not committed to the existence of beliefs and desires as inner causal states. The key

---

<sup>3</sup> Throughout this discussion, I will use the term "behaviourism" to refer to analytic, rather than methodological, behaviourism. While the analytical behaviourist offers an analysis of ordinary talk about mental states, the methodological behaviourist instead proposes a restriction on the proper object of psychological inquiry. For a classic defence of analytic behaviourism, see Ryle (1949). For methodological behaviourism, see Skinner (1974).

difference between fictionalism and behaviourism, however, is that the fictionalist does not try to avoid ontological commitment by giving a reductive analysis of mental state attributions in terms of behaviour. Instead, at least on the version of mental fictionalism I have proposed, Ruth's utterance "John believes that planes can fly" *means* exactly what the realist about mental states takes it to mean: it means that there is a particular internal state inside John's head with certain characteristic causes and effects. The reason that Ruth is not committed to the existence of such entities, according to the fictionalist, is simply because her utterance is not a straightforward assertion; instead, it is an act of pretence. Similarly, if we say, "Crotona is on the arch of the Italian boot" our utterance means the same as if we were (oddly enough) to claim that there really *was* an enormous boot floating in the Mediterranean. Within the context of the game, however, our utterance is recognised as pretence (Walton 1990, 1993, Yablo 1998).

Our pretence involving the Italian boot allows us to make a genuine assertion regarding the location of Crotona. Similarly, Ruth's pretence regarding John's inner states allows her to make a genuine assertion regarding John and his behaviour. There is an important difference between the two cases, however. In the case of the Italian boot, the use of make-believe is dispensable. We can offer an alternative, literal paraphrase that captures the assertion we make about Crotona: we are claiming that Crotona lies on the southern coast of Italy, somewhere roughly between Capo Colonna and Taranto. Not all metaphors can be given a literal paraphrase, however. Thus, Stephen Yablo (1998, p. 250) notes the possibility that

the language might have no more to offer in the way of a unifying principle for the worlds in a given content than that they are the ones making the relevant sentence fictional. It seems at least an open question, for example, whether clouds we call *angry* are the ones that are literally *F*, for any *F* other than 'such that it would be natural and proper to regard them as angry if one were going to attribute emotions to clouds'.

Yablo calls metaphors that resist literal paraphrase *representationally essential* (*ibid.*; see also Walton 1993). For the fictionalist, the failure of the behaviourist project suggests that

the metaphors of folk psychology are representationally essential. If we say John wants to go to Madrid, we are saying that he is in some state *S* such that it is appropriate to pretend in this way. There are indefinitely many different forms of behaviour that might make this pretence appropriate: running eagerly to board a plane (because he believes it will take him to Madrid); taking the ferry to Santander and then the train (because he is scared of flying); wistfully thumbing through travel brochures (because he believes he can't afford to go to Madrid); and so on. Each of these counts as a situation in which the rules of folk psychology make it fictional that John desires to go to Madrid. And yet, just as there might be no way of capturing what is common to all clouds we call "angry" (apart from that they each make it fictional that they are angry), so there might be no way of capturing what is common to each of these forms of behaviour (apart from that they each make it fictional that John desires to go to Madrid). The problems faced by behaviourism thus turn out to be an instance of a more general phenomenon, namely the difficulty of giving a literal paraphrase for metaphors and figurative language.

## 4.2 Instrumentalism

It is also helpful to compare fictionalism to instrumentalism. Instrumentalism can be characterised in a number of different ways. *Intentional systems theory* aims to capture the principles underlying our ordinary belief and desire attributions (Dennett 1987). According to David Braddon-Mitchell and Frank Jackson (2007, p. 159), instrumentalism is the view that "[t]here is nothing more to being a believer and a desirer, a thing with beliefs and desires, than being a being whose behaviour is well predicted by [intentional systems theory]". Thus, the instrumentalist claims that

‘S believes P and desires Q’ is true iff (a) the behaviour of S is well predicted by intentional systems theory, and (b) the best belief and desire hypotheses according to the principles of intentional systems theory for predicting S’s behaviour are that S believes P and desires Q (*ibid.*)

An important challenge for instrumentalism is provided by the Blockhead thought

experiment (Block 1981). Suppose Blockhead's behaviour matches our own, but its inner machinery involves a giant 'look-up' table determining its response to each possible input. Intuitively, it seems, we would not count Blockhead as a genuine thinker. And yet, if all we mean when we say that someone has beliefs or desires is that their behaviour can be predicted by intentional systems theory, then instrumentalism cannot account for this intuition. As Braddon-Mitchell and Jackson (2007, p. 163) observe, fictionalism can avoid this worry by rejecting the instrumentalist's semantic thesis. We have seen already that the fictionalist takes an utterance like "John believes that planes can fly" to mean what the realist does: taken literally, it claims that there is a particular sort of internal state inside John's head. As a result, fictionalism can account for our intuitions regarding Blockhead: if Blockhead does not possess the right sort of internal machinery, then he doesn't count as a true believer. To be a true believer one *would* have to have a certain internal organisation, which Blockhead lacks. As Braddon-Mitchell and Jackson see it, by developing along fictionalist lines, "instrumentalism is no longer the comfortably uncommitted doctrine it presented itself as: a cautious refusal to traffic in speculation about what goes on inside us. The only viable option turns out to be a version of the radical thesis that all ascriptions of belief and desire are false" (2007, p. 164). But, as I understand her, the fictionalist can remain agnostic regarding whether we humans have the required internal organisation to count as genuine believers. She simply insists that, regardless of whether we have this internal organisation or not, it is useful to talk as if we do.

Daniel Dennett is often taken to be the leading proponent of instrumentalism. Of course, Dennett himself has rejected the label "instrumentalism" for his position, as well as "fictionalism" (Dennett 1987, 1991). Despite this, I think that the version of mental fictionalism that I have proposed shares interesting similarities to Dennett's view, and that it offers a way to interpret aspects of Dennett's position that critics have found problematic. For example, Dennett (1987, p. 72) distances himself from fictionalism, which he takes to be the claim that theoretical statements are useful falsehoods. And he also denies that talk about mental states is neither true nor false. Instead, Dennett argues,

talk about beliefs and desires can be true, although this is a truth which “one must understand *with a grain of salt*” (1981, pp. 72-73; emphasis in original). To explain this idea, Dennett draws a parallel with centres of gravity:

I should have forsworn the term [instrumentalism] and just said something like this: My *ism* is whatever *ism* serious realists adopt with regard to centres of gravity and the like, since I think beliefs [...] are *like that* – in being *abstracta* rather than part of the “furniture of the physical world” and in being attributed in statements that are *true* only if we exempt them from a certain familiar standard of literality. (1981, p. 72)

Thus, Dennett suggests, claims about beliefs and claims about centres of gravity can both count as true, even if neither beliefs nor centres of gravity exist. The analysis I introduced in Section 3 offers a way to make sense of this idea (cf. Hutto 2013). When we say that a system has certain centre of gravity, we pretend that its mass is located at a certain point. Taken literally, this is false. And yet, when we attribute a centre of gravity to a system, we also make a genuine assertion: we claim that the system’s mass is distributed such that it is appropriate to pretend that all the mass is located at a particular point. And this claim can be straightforwardly true. Similarly, when we claim that someone has a certain belief, we pretend that they have a particular inner state. Taken literally, this is (or might be) false. And yet, when we attribute a belief, we also make a genuine assertion: we claim that they are in a particular state such that is appropriate to pretend in this way. And this claim can be straightforwardly true.

Notably, Dennett also insists that, although beliefs and desires might not exist, the patterns picked out by the intentional stance are nevertheless entirely real, objective features of the world (Dennett 1987, 1991). Moreover, Dennett argues, these patterns would be overlooked by Martians who knew all there was to know about the physical world, but failed to adopt the intentional stance. Once again, I think that a fictionalist analysis can help us to understand how this might be the case. Suppose that we say “the clouds over Exeter were angry all day”. There are no (literally) angry clouds. And yet when we say this we might capture an entirely real, objective feature of the world: the

fact that the clouds over Exeter on a particular day were such that our pretence is appropriate. And if describing clouds as angry is a metaphor that resists literal paraphrase—that is, if we are unable to give a description in the language of the physical world that specifies what it is that all clouds that we would fictionally count as angry have in common—then the pattern exhibited by clouds over Exeter today might be inaccessible to someone who doesn't recognise this particular game. Similarly, even if beliefs and desires do not exist, pretending that they do might allow us to pick out perfectly real and objective features of the world. And if talk of beliefs and desires is a metaphor that resists literal paraphrase—that is, if we are unable to give a description in the language of the physical world that specifies what it is that all people who count as possessing a particular belief or desire have in common—then these patterns might be overlooked by beings that did not take part in the game of folk psychology.

### 4.3 Prefix-fictionalism

Many discussions of mental fictionalism focus mainly on a version of the approach that we might call *prefix-fictionalism* (Armour-Garb and Woodbridge 2015). On this view, utterances such as “John believes that planes can fly” are taken to be elliptical for prefixed claims like “In the folk psychological fiction, John believes that planes can fly” (Wallace 2007, Parent 2013). By contrast, I have proposed a form of *pretence-fictionalism* (Armour-Garb and Woodbridge 2015): talk about mental states should be understood as an act of pretence, rather than involving an implicit prefix. While both prefix and pretence versions of mental fictionalism try to avoid commitment to mental states, they offer different analyses of folk psychological talk. In this section, I argue that we ought to prefer the pretence form of mental fictionalism.

The prefix form of mental fictionalism is a version of what Yablo (2001) calls *meta-fictionalism*: according to the prefix-fictionalist, utterances that appear to be about people's mental states turn out to be claims about the contents of a particular theory, namely folk psychology (see also Eklund 2007). Yablo (2001, pp. 75-76) presents a

number of challenges for meta-fictionalism, focusing primarily on mathematical fictionalism. First, Yablo argues, meta-fictionalism has problems dealing with modal claims. We regard claims like “ $2+2=4$ ” as necessarily true. According to the prefix-fictionalist, however, when we say “ $2+2=4$ ” we are claiming that “In standard mathematics,  $2+2=4$ ”. And yet, arguably, this claim is not necessarily true, since standard mathematics might perhaps have turned out differently. Second, Yablo argues that meta-fictionalism faces “problems of concern”. Suppose that we say “the number of starving people is large and rising”. When we say this, what we care about is people and their desperate situation. And yet, according to the prefix-fictionalist, what we are really talking about is the content of standard mathematics. Third, Yablo raises a related, phenomenological worry. When we say “the number of starving people is large and rising”, we do not *feel* remotely as if we are talking about the content of standard mathematics; instead, we feel that we are talking about people.

Each of these difficulties confronts prefix versions of mental fictionalism. First, consider modality (cf. Wallace 2007, pp. 20-21). Of course, unlike a claim like “ $2+2=4$ ”, we don’t take attributions such as “John believes that planes can fly” to be necessarily true. And yet prefix-fictionalism still seems to run into difficulties here. Consider a possible world in which folk psychology turned out differently (perhaps our visionary Jones never lived, or took a different model as the inspiration for his revolutionary account of people’s behaviour). According to prefix-fictionalism, in such a scenario, a claim such as “John believes that planes can fly” would be false, even if all of the facts about John and his aeroplane-related behaviour were exactly the same. This seems wrong: intuitively, I think, we should still count this as a world in which John believes that planes can fly. Or, to take another example, suppose that we grant that, since Freud’s work, the notion of unconscious desires has become part of folk psychology. Again, it seems wrong to say that, had Freud never written, then all of our attributions of unconscious desires would be false. Second, consider the problem of concern. When we talk about people’s mental lives (rather than, say, discuss philosophy of mind), what we care about are the people we are talking about, not the contents of folk psychological theory. Finally, when we say “John

wants to go to Madrid” or John wants a new job”, it certainly *feels* as if we are talking about John, not about the way the folk talk about John.

The pretence version of mental fictionalism avoids these worries. On the analysis I have proposed, attributions of mental states are not claims about the contents of folk psychological theory; instead, they are acts of pretence that serve to make claims about people and their behaviour. This allows the pretence fictionalist to respond to each of the difficulties facing prefix fictionalism. If we say (correctly) that “John believes that planes can fly”, the claim we make is true in virtue of facts concerning John and his behaviour. If the contents of folk psychology were to change while the facts about John remained the same, then our claim would still be true. Similarly, our attributions of unconscious desires would retain their present truth values, even if Freud had never written. And the pretence-fictionalist has no problem explaining why it is that, when we attribute mental states to John, it is John himself that we care about and whom we feel we are talking about: according to the pretence analysis, that is precisely what we are doing, although we do so via the use of pretence.

The pretence version of mental fictionalism also avoids a further difficulty confronting prefix-fictionalism. As we have seen, the prefix-fictionalist claims that talk about mental states is talk about a folk psychological fiction. The guiding analogy here, of course, is with an utterance such as “Holmes smokes a pipe” which, according to prefix-fictionalism, is short for “In the Sherlock Holmes stories, Holmes smokes a pipe”.<sup>4</sup> And yet there seems to be an important difference between the two cases. After all, it is clear which fiction underpins our talk about Sherlock Holmes: we can point to our copy of Conan Doyle’s stories. But there is no text that sets out the principles of folk psychology. Indeed, as we saw earlier, such principles are notoriously difficult to formulate explicitly. In this respect, mental fictionalism might seem to be in a worse position than fictionalism in other domains. For example, Gideon Rosen’s (1990) modal fictionalism is able to draw

---

<sup>4</sup> For a discussion of the problems facing this approach to fiction, see Friend (2007).



on David Lewis' *On the Plurality of Worlds* (1986) to set out its guiding fiction. Of course, there are ways that the prefix fictionalist might try to respond to this worry (see Wallace 2007, pp. 9-11). For our purposes, the important point is that the problem does not arise for pretence fictionalism. All the pretence fictionalist requires is that we have a reasonably coherent set of rule-governed practices for attributing mental states to people based on their behaviour. It does not require that we are able to distil any folk psychological principles or laws by reflecting on those practices. Indeed, as we noted above, the rules of most make-believe games—even children's games with dolls—are complex and difficult to formulate explicitly.

#### **4.4       Eliminativism**

Finally, let us compare fictionalism to eliminativism. Eliminativists claim that mental states, like beliefs and desires, do not exist. The eliminativist might offer various arguments for this claim: she might argue that folk psychology is a bad theory (Churchland 1981), for example, or that it stands fundamentally at odds with particular developments in cognitive science (Ramsey, Stich and Garon 1990). As we have seen, the fictionalist need not follow the eliminativist in denying the existence of mental states; instead, she might remain agnostic on the matter. The distinctive feature of fictionalism is that it allows us to grant that, even if mental states do not exist, we can nevertheless continue talking as if they do. This represents a clear advantage of fictionalism over eliminativism. Abandoning talk about beliefs, desires and other mental states would require an enormously dramatic and far-reaching transformation in our language. The fictionalist aims to show that the legitimacy of ordinary talk about the mind does not depend upon beliefs and desires finding a place in the theories of future cognitive science.

Why talk about mental states if they don't exist? According to the fictionalist, such talk shares the advantages of metaphorical and figurative language more generally. We have seen already that metaphors can allow us to make claims that we are unable to express in a straightforward literal description. But metaphorical and figurative language also brings

further advantages. Metaphors introduce a “framing effect” (Moran 1989; see also Demeter 2013b): we are asked to “see” our primary subject (e.g. Italy, clouds) in terms of a secondary subject (e.g. a boot, emotions) (Beardsley 1962; see also Hills 2011). This yields a host of different cognitive benefits. Metaphorical claims are often especially vivid and memorable, for example, and prompt a range of further inferences about their primary subject. Consider the metaphor of theories as buildings (Lakoff and Johnson 1980). If we are told that someone is hard at work constructing a new theory, we are immediately led to wonder what the theory’s foundations are, whether it requires further support, whether it will ultimately stand or fall, and so on. By asking us to see people’s behaviour as the product of inner states interacting in various ways, folk psychology introduces an enormously powerful framing effect. It provides us with a way of describing people and their behaviour that is vivid and memorable (e.g. “John had an overwhelming desire to go to Madrid”), prompts us to make a range of further inferences (e.g. “if John thinks there’s any way he can get to Madrid, he’ll take it”), allows us to contrast different people’s behaviour in new ways (“Ruth’s desire to go to Madrid is nowhere near as strong as John’s”), and so on. The key point, according to the fictionalist, is that folk psychology can possess these virtues even if there are no beliefs or desires.

## **5 Challenges**

I now turn to consider some objections to mental fictionalism. Each of these represents an important challenge facing the position, and I cannot hope to do justice to them properly here. But I hope at least to show that these objections need not be fatal and to indicate some possible lines of response.

The first, and most obvious, challenge concerns the phenomenology of talk about the mental (Wallace 2007; see also Eklund 2007). As we noted earlier, we don’t normally feel as if we are engaging in pretence when we attribute mental states. When we say “John believes planes can fly” or “John wants to go to Madrid”, we don’t feel as if we are engaging in make-believe; we feel as if we are making a straightforward assertion about

John. Of course, one response to this worry is to endorse a revolutionary fictionalism. But there are also ways that we might look to defend hermeneutic fictionalism against this objection. The first point to stress is one made a number of times already: on the pretence version of mental fictionalism, mental state attributions *are* used to make genuine assertions about people and their behaviour, albeit via pretence. Moreover, we need not insist that the folk pretend in the same manner as children participate in their games; for example, perhaps they merely indicate the relevant pretence, without engaging in it (Walton 1990, 1993).

More positively, the fictionalist might point out that we often don't notice when we are speaking metaphorically, especially when metaphors are familiar to us (Yablo 2000; see also Eklund 2007). Once we reflect more closely on our talk about mental states, she might argue, we see that it bears striking similarities to metaphorical language. To support this claim, the fictionalist can draw on some well-known arguments for instrumentalism. For example, Yablo (2000) notes that objects invoked in make-believe have problematic identity conditions (e.g. "is the fuse you blew last week the same as the one you blew today?"). In a similar vein, Dennett notes that "common intuition does not give us a stable answer to such puzzles as whether the belief that 3 is greater than 2 is none other than the belief that 2 is less than 3" (1987, p. 55). Another characteristic feature of metaphors is that they invite "silly questions" (e.g. "What species is the monkey on your back?"). Many realist questions about mental states, I suggest, also strike us as silly questions (e.g. "Where exactly is John's belief about planes? Is it to the right or left of his desire to go to Madrid?").

The mental fictionalist might also support her interpretation of folk discourse by appealing to an "Oracle argument" (Eklund 2007). Yablo (2000) offers the following argument in support of fictionalism about abstract objects. Suppose an omniscient Oracle told you that there are no abstract objects. Would you stop saying things like " $2+3=5$ "? Arguably not. Instead, you would go on talking exactly as before. If that's right, then it seems that such talk did not commit you to the existence of abstract objects in the first place. Again, the fictionalist might draw on familiar instrumentalist themes to mount a

similar argument in defence of mental fictionalism. Thus, Dennett (1987) writes:

“[s]uppose, for the sake of drama, that it turns out that the sub-personal cognitive psychology of some people turns out to be dramatically different from that of others. One can imagine the newspaper headlines: “Scientists Prove Most Left-handers Incapable of Belief” or “Startling Discovery – Diabetics Have No Desires.” But this is not what we would say, no matter how the science turns out.

And our reluctance would not be just conceptual conservatism, but the recognition of an obvious empirical fact. For let left- and right-handers (or men and women or any other subsets of people) be as internally different as you like, we already know that there are reliable, robust patterns in which all behaviourally normal people participate—the patterns we traditionally describe in terms of belief and desire and the other terms of folk psychology” (1987, p. 235).

Like the eliminativist, the fictionalist can agree that the scenario that Dennett describes is possible: it *might* turn out that people do not literally have beliefs or desires inside their heads. And yet, even if scientists were to make such a startling discovery, the fictionalist argues, folk psychological talk would continue unabated. What this shows is that talk about mental states serves a different purpose, which doesn’t depend on the existence of mental states as inner causes, namely providing a valuable metaphor for identifying patterns in behaviour.

A second important objection to mental fictionalism is that it cannot account for the fact that mental states appear in causal explanation (cf. Sprevak 2013). If we say “John’s desire to go to Madrid caused him to spend all his savings”, we appear to cite a mental state as the cause of John’s behaviour. And yet, if mental states are fictions, it seems that they cannot be causes. Is this a serious difficulty for the fictionalist? Consider other uses of metaphor in causal explanations. Suppose we say “the angry clouds caused Ruth to fetch her umbrella”. Following Walton’s analysis it seems that, when we say this, we are

saying that the clouds were in a certain state *S*—whatever it is—such that, fictionally, we speak the truth, and this state *S* caused Ruth to fetch her umbrella. Even if there are no angry clouds, it is arguable that this is still a genuine causal explanation. Similarly, if we say “John’s desire to go to Madrid caused him to spend all his savings”, we are claiming that John is in a certain state *S*—whatever it is—such that, fictionally, we speak the truth, and this state *S* caused him to spend all his savings. Once again, even if there are no desires states, it is arguable that this is still a genuine causal explanation. But, of course, it falls far short of the idea that folk psychological explanations pick out discrete inner causes of behaviour, and there is certainly much more to be said here (for a related discussion, see Dennett 1987, pp. 56-57).

Finally, and perhaps most worryingly, a critic might argue that mental fictionalism is incoherent (e.g. Wallace 2007, 2014, Daly 2013, Joyce 2013, Parent 2013). The charge of incoherence mirrors a well-known objection to eliminativism. The eliminativist claims that mental states do not exist. And yet, the critic claims, asserting something involves believing it. So eliminativism is incoherent: the very act of asserting the position shows it to be false. Fictionalism faces a related worry. The fictionalist claims that talk about beliefs and desires is pretence within a game of make-believe. And yet pretence and make-believe are themselves folk psychological notions. So mental fictionalism is incoherent: the very act of stating the position requires commitment to the existence of mental states. This is perhaps the most serious objection facing mental fictionalism, and I cannot hope to deal with it properly here (for fuller discussions, see Wallace 2014, Joyce 2013, Parent 2013). The most promising response for the fictionalist, I believe, echoes Churchland’s (1981) well-known response on behalf of eliminativism (see also Wallace 2007, Joyce 2013).

Churchland (1981) argues that the charge of self-refutation begs the question against eliminativism: 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 such activities

which is radically different from that offered by folk psychology. At present, however, we have no alternative but to describe activities such as assertion using folk psychological notions. The mental fictionalist might take a similar line. The fictionalist's central claim is that the practice of folk psychology should be understood in a similar way to activities such as playing games or engaging with fiction. Eventually, we might come to possess a proper neuroscientific account of such activities which is radically different from that offered by folk psychology. At present, however, we have no choice but to describe these activities using folk psychological notions such as pretence and make-believe.

### **Acknowledgements**

Versions of this paper were presented at the conference on Mental Fictionalism at the University of Edinburgh, July 14<sup>th</sup>-15<sup>th</sup> 2014, 'Shaping the Trading Zone: Bringing Aesthetics and Philosophy of Science Together' at the University of Leeds, September 5<sup>th</sup>-6<sup>th</sup> 2014, and the workshop on Scientific Fictionalism at the Institute of Philosophy in London, June 12<sup>th</sup>-13<sup>th</sup> 2015. Thanks to the participants at each of these events for helpful comments and criticism. And a special thank you to Fiora Salis for all her encouragement and hard work in editing this volume. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant agreement No. 331432.

## References

- Armour-Garb, B., & Woodbridge, J. A. (2015). *Pretense and Pathology*. Cambridge: Cambridge University Press.
- Beardsley, M. (1962) The Metaphorical Twist. *Philosophy and Phenomenological Research*, 22(3), 293–307.
- Block, N. (1981). Psychologism and Behaviorism. *The Philosophical Review*, 90(1), 5–43.
- Braddon-Mitchell, D., & Jackson, F. (2007). *Philosophy of Mind and Cognition: An Introduction*. Oxford: Blackwell (2<sup>nd</sup> edition).
- Burgess, J. & Rosen, G. (1997). *A Subject With No Object*. Oxford: Clarendon Press.
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67–90.
- Crimmins, M. (1998). Hesperus and Phosphorus: Sense, Pretense, and Reference. *Philosophical Review* 107, 1-47.
- Daly, C. (2013). Psychology and Indispensability. *The Monist*, 96(4), 561–581.
- Demeter, T. (Ed.) (2013a). Mental Fictionalism [Special Issue]. *The Monist*, 96(4).
- Demeter, T. (2013b). Mental Fictionalism: The Very Idea. *The Monist*, 96(4), 483–504.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Eklund, M. (2015). Fictionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 edition). Retrieved from <http://plato.stanford.edu/archives/win2015/entries/fictionalism>
- Friend, S. (2007). Fictional Characters. *Philosophy Compass*, 2(2), 141–156.
- Godfrey-Smith, P. (2005). Folk Psychology as a Model, *Philosopher's Imprint*, 5(6), 1–16.

- Hills, D. (2012). Metaphor. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition). Retrieved from <http://plato.stanford.edu/archives/win2012/entries/metaphor>
- Hutto, D. (2013). Fictionalism about Folk Psychology. *The Monist*, 96(4), 582–604.
- Joyce, R. (2005). Moral Fictionalism. M. E. Kalderon (Ed.) *Fictionalism in Metaphysics* (pp. 287-313). Oxford: Oxford University Press..
- Joyce, R. (2013). Psychological Fictionalism, and the Threat of Fictionalist Suicide. *The Monist*, 96(4), 517–538.
- Kroon, F. (2001). Fictionalism and the Informativeness of Identity. *Philosophical Studies* 106, 197-225.
- Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
- Moran, R. (1989). Seeing and Believing: Metaphor, Image, and Force, *Critical Inquiry*, 16(1), 87–112.
- Parent, T. (2013). In the Mental Fiction, Mental Fictionalism Is Fictitious. *The Monist*, 96(4), 605–621.
- Ramsey, W., Stich, S. P., & Garon, J. (1990). Connectionism, Eliminativism and The Future of Folk Psychology. *Philosophical Perspectives*, 4, 499–533.
- Rosen, G. (1990). Modal Fictionalism. *Mind* 99: 327-54.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.
- 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.



- Skinner, B. F. (1974). *About Behaviourism*. London: Jonathan Cape.
- Sprevak, M. (2013). Fictionalism About Neural Representations. *The Monist*, 96(4), 539–560.
- Stanley, J. (2001). Hermeneutic Fictionalism. *Midwest Studies in Philosophy* 25(1), 36–71.
- Wallace, M. (2007). Mental Fictionalism. Unpublished manuscript.
- Wallace, M. (2014). Saving Mental Fictionalism from Cognitive Collapse. Unpublished manuscript.
- Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard: Harvard University Press.
- Walton, K. L. (1993). Metaphor and Prop Oriented Make-Believe. *European Journal of Philosophy*, 1(1), 39–56.
- Yablo, S. (1998). Does Ontology Rest on a Mistake? *Proceedings of the Aristotelian Society, Supplementary Volume*, 72, 229–261.
- Yablo, S. (2001). Go figure: A path through fictionalism. *Midwest Studies in Philosophy*, 25(1), 72–102.
- Yablo, S. (2005). The myth of the seven. In M. E. Kalderon (Ed.) *Fictionalism in Metaphysics* (pp. 88–115). Oxford: Oxford University Press.