

## Mental Fictionalism: the costly combination of magic and the mind

Amber Ross (forthcoming)

in *Mental Fictionalism: Philosophical Explorations* (Routledge)

Tamas Demeter, Ted Parent and Adam Toon (eds.)

### 1. Introduction

Recent philosophy of mind has seen a growing interest in mental fictionalism, and there are several compelling reasons for its current popularity. Some philosophers think that the best way to make sense of our mental state discourse is to treat that discourse in a way quite similar to how we treat our discourse about fiction, rather than treating the terms of folk psychology as if they intended to refer to actual entities. Relatedly, mental fictionalism (if successful) might help resolve the tension between adopting a physicalist view of the mind and the inclination to treat many of our mental state attributions as *truth-apt*. Statements such as “He believes it will rain soon,” “She wants more information before making this decision,” and “He’s been feeling a bit down lately,” all seem to be evaluable for truth or falsity. A perennial problem for philosophy of mind is that it is far from obvious just *what it is* that provides the truth conditions for these statements. Given that both linguistic and ontological behaviorism have failed as satisfactory accounts of either the nature of mental states or the meanings of our mental terms, it seems clear that, *whatever it is* that constitutes the truth conditions of certain mental state attributions, observable behaviors (on their own) are *not the right sort of things* to constitute them.<sup>1</sup>

Another reason to adopt a fictionalist approach to mental states might be that it allows a physicalist to more easily adhere to Quine’s criterion for ontological commitment: *To be is to be the value of a bound variable*. A philosopher who wants to embrace both physicalism and Quine’s criterion is likely to be concerned that our discourse about mental states (in particular, our folk psychological

---

<sup>1</sup> Observable behavior may well provide *reasons in favor of* certain mental state attributions, but behavior does not, on its own, make it the case that certain mental state attributions are *true* (see Putnam, 1993; Lewis, 1983)

mental state attributions) commits us to the existence of certain metaphysically suspect entities such as beliefs, desires, and feelings.

Though Quine's dictum explicitly commits a speaker only to the entities referred to in certain *types* of discourse (specifically, discourse involving whatever entities our best scientific theories quantify over), this is sufficient to cause problems for physicalists who would like to retain mentalistic discourse, since folk psychology- the common-sense theory that employs beliefs, desires, and feelings in order to predict and explain the actions of ordinary individuals- certainly appears to be (1) a scientific theory (as it is both systematic and non-coincidentally successful at predicting agents' actions), and (2) proceeds by referring to mental "entities" (our beliefs, desires, and feelings) which seem to play a causal role in generating the actions of individual agents. Mental discourse also seems indispensable to our everyday lives, and it may be indispensable to empirical science as well: it is unclear whether psychology, neuroscience, or the social sciences could proceed without use of the terms that apparently refer to these entities (such as *memories* and *sensations*). The likelihood that we will discover a straightforward reduction of our folk psychological mental states to neurological states seems to be steadily diminishing over time, which makes it less likely that folk psychology will simply be replaced by more advanced neurological (or biological, or chemical) theories.

As originally proposed by Wallace (2007) these latter concerns have been central to the motivation behind mental fictionalism. Mental fictionalism may occupy just the right location in logical space for the Quinean philosopher of mind who seeks a principled way to retain our folk psychological discourse. Quine's dictum does not commit us to the entities referred to in *all* types of discourse; in particular, Quine makes explicit mention of *fictional* discourse as the kind of discourse in which being the value of a bound variable has no ontological strings attached.

One way in which a man may fail to share the ontological commitments of his discourse is, obviously, by taking an attitude of frivolity. The parent who tells the Cinderella story is no more committed to admitting a fairy godmother and a pumpkin coach into his own ontology than to admitting the story as true. (Quine, 1980, p. 103)

On the face of it, *fiction* and *fictional discourse* seem to be just the sort of areas in which we might find appropriately analogous strategies for dealing with the challenges mentioned above. Richard Joyce characterizes philosophical fictionalism, in general, as “not just the negative thesis that utterances that appear to make ontological commitments do not do so, but also the positive (though vague) thesis that the role that these utterances do play is substantively similar to the role of familiar fictional discourse” (Joyce, 2013, p. 520). If this is correct, then a successful mental fictionalism should serve a tidy dual purpose: it should allow the speaker to use referring terms in her folk psychological theory without thereby being committed to the existence of their referents, and it should illuminate or improve our understanding of *what speakers are doing* when they engage in this discourse (since fictionalism denies that speakers are making *existence* claims, they must be doing *something else*).

For these reasons alone, mental fictionalism deserves serious consideration. It is an intriguing position, a cunning move into a long-overlooked opening in logical space, and is explicitly encouraged by Quine’s own mention of fiction as a paradigmatic case of non-committing discourse. However, I suspect- for reasons that I will give below- that the cost of adopting mental fictionalism may outweigh its benefits, and at the very least we ought to acknowledge what would be lost by adopting such a position as well as what could be gained. In the process, we may also illuminate several important features of our own folk psychology that often go unnoticed.

## **2. Toon’s Mental Fictionalism as Prop-Oriented Make-Believe**

The specific account of mental fictionalism on which I will focus is the account developed by Adam Toon (2016); a version modeled on Kendall Walton’s theory of “prop-oriented make-believe” (1990, 1993). According to Toon’s account, talking of people’s mental states, or making mental state attributions, should not be interpreted as “making a claim about their inner machinery” (Toon, 2016, p. 280). Instead, “talk of mental states is a useful pretence for describing people and their behaviour” (ibid). Toon’s Waltonian account differs from other forms of mental fictionalism that could be categorized as *meta-fictionalism* (Yablo, 2001), according to which claims that

appear to be about mental states are actually claims about the contents of our folk psychological theory.<sup>2</sup> On Toon's account, the principles of folk psychology are conceptualized not as a story themselves but as Waltonian *principles of generation*, a loose set of rules which establishes a kind of pretense that, along with objects that serve as props for the pretense, allow us to engage in a game of make-believe. "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)" (Toon, 2016, p. 281). In the pretense, or in the make-believe, props have features that they do not have in the actual world, and these features are specified (to the required extent) by the principles of generation for that game of make-believe.

Consider Walton's example, discussed by Toon, in which someone is using the pretense that *Italy is a boot* in order to describe the location of different Italian cities:

Mark [engaging in the pretense that Italy is a boot] ...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. (Toon, 2016, p. 282)

In the case where a speaker says "Crotone is on the arch of the Italian boot," an analysis of the act in terms of prop-oriented make-believe seems straightforward. We can identify the prop- Italy- and the principles of generation- the rule that, in the make-believe, Italy is a boot and the Italian cities occupy different locations on the boot. The prop and the principles of generation are reasonably clear, and engaging in this pretense may be prop-oriented in the sense that it improves our understanding of Italian geography (or eases our learning it). Mark asserts a truth about the actual world- the location of Crotone- by using certain phrases in an act of make-believe pretense. A speaker can use fictional discourse to assert something true about the prop.

Though it may not be an unmitigated success, Toon's proposal has clear advantages over other versions of mental fictionalism. Modeling his view on prop-oriented make-believe allows us to retain at least two important features of folk psychology: certain statements employing mental

---

<sup>2</sup> For examples of this alternative form see Wallace (2016), Parent (2013), and Eklund (2007)

state terms will be truth-apt, and the truth value of these statements will depend not only on the principles of folk psychology (treated as “principles of generation”) but on these principles in combination with certain (objective) facts about the actual world (the way things are with the “props” of the make-believe). This second feature is an advantage of Toon’s proposal over other forms of mental fictionalism, as it allows for a level of objectivity that could otherwise easily be lost. As Toon explains, using the example of a baby-doll in a make-believe game,

...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. (2016, P. 281)

To show how this framework might be applied to folk psychological discourse, Toon points to Sellars’s “Myth of Jones” and recasts Jones’s crucial turn as a fictionalist move. Rather than introducing a new way of talking about people that employs mental state terms to refer to (supposedly actual) hidden interior states, Jones introduces mental discourse as a kind of make-believe; a useful metaphor.

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. (Toon, 2016, p. 283)

### **3. A partial defense of Prop-Oriented Mental Fictionalism: Responding to Bourne and Caddick Bourne (2020)**

As valuable as it would be to retain the objectivity and connection to the actual world which prop-oriented mental fictionalism could provide, there are several problematic disanalogies that need to be addressed. The first is between (a) what is actually asserted when making claims about the props

in a make-believe game and (b) what is actually asserted when making claims about individuals' mental states.

In the Italy/boot make-believe, Mark may actually assert that Crotone is in such-and-such a position on the Italian coastline when he says, "Crotone is in the arch of the boot". But there are no easily identifiable *non-fictional* statements (ones void of mentalistic terms) which would be asserted with the (pretense) utterance that contains mental state terms. According to Toon, if we interpret a discourse as a kind of make-believe, then "acts of pretence can be used to make genuine assertions" (Toon, 2016, p. 281). In that case, we would like some account of what it is that Sia *actually* asserts when she says, "Sergio feels happy when he helps others." Whatever she asserts, it cannot commit her to the existence of the mental state of *happiness*. To make sense of the view Toon offers, we would like some account of *what statement* one would genuinely asserting when (in the pretense) one says something like, "He believes it will rain soon," or "She feels lucky today". No obvious candidates emerge.

Bourne and Caddick Bourne (2020) object to Toon's account of mental fictionalism on related grounds. In particular, they claim that, "genuinely prop-oriented make-believe demands a particular relationship between props, content, and understanding. It is not clear that this relationship can be borne out in the case of FP" (2020, p. 173). As they elaborate,

We can well understand the structure and flight of the folded paper and the objectives of throwing it and adjusting it through seeing them as props for imagining the behavior of a plane rather than something else. Making the analogous claim about behavioral props and folk-psychological make-believe is much more difficult. What exactly are we identifying about behavioral props in seeing them as suited to prescribing those imaginings rather than others? (2020, p. 175)

Whereas my objection above focused on content, Bourne and Caddick Bourne object that there is no analogous *understanding* generated by engaging in prop-oriented mental fictionalism. With this element absent, they claim that Toon has, at best, argued for interpreting folk psychology as metaphor (2020, p. 177). If this is correct then Toon's theory has lost a large part of its appeal. The

prospect of a Waltonian prop-oriented model of mental fictionalism was particularly enticing because it gave the *actual world* a substantial role to play in determining the truth value of particular mental state attributions. And intuitively it does seem as though the facts about the actual world (specifically, facts about the individuals to whom we are attributing mental states) are an ineliminable part of what constitutes the truth conditions of our mental state attributions. The *prop-orientation* aspect of prop-oriented make-believe allowed us to easily integrate the actual world into the grounds for the truth value of folk psychological statements. Without this anchor, folk psychological claims would be unmoored from the world, from the individuals they are supposedly about, from the actual features that should adjudicate whether, for instance, elephants are capable of feeling sad, or Sergio feels happy when he helps others.

Though Bourne's and Caddick Bourne's criticism is plausible, I believe that Toon's version of mental fictionalism can respond to their objection (as well as mine above) with some slight augmentation. Toon has not specified any particular genre of fiction for these make-believe games, and there are legitimate fictional genres to which he could appeal that would be immune to these critiques. Bourne and Caddick Bourne ask what exactly it is that we should be identifying about behavioral props in seeing them as suited to prescribing those (mentalist) imaginings rather than others. I ask what statement is being asserted when one says "Elephants can feel sad" while engaging in the mental-pretense. Folk psychology does not address these questions- while the folk have an intuitive notion of which behaviors are evidence for the presence of which mental states, there is no account of *exactly what it is about the behavior* that makes it reasonable for us to ascribe certain mental states. Instead, folk psychology (it has been suggested) treats mental entities *as if they are magic*. And magic, by its nature, must be left unexplained (else it would cease to be *magic*).<sup>3</sup> Toon might plausibly argue that the games of make-believe after which we should model our interpretation of folk psychology as fictional discourse belong to the fantasy genre of fiction, in which magic is a central feature.

---

<sup>3</sup> This characterization of magic as something that- by its nature- must resist exhaustive explanation bears a significant similarity to Yablo's and Gallois's (1998) characterization of certain metaphors as "representationally essential" (i.e., the kind of metaphors that are not amenable to exhaustive explanation). I find it plausible that, that if we interpret folk psychology as metaphor, then- given how we typically conceive of the mind- these metaphors may be rightly considered representationally essential. However I am disinclined to view folk psychology as a practice that, at its most fundamental level, trades in *metaphor*. Below I will give reasons to believe that this characterization of folk psychology overlooks some of its most important features.

If we take the fantasy genre into consideration, then Bourne's and Caddick Bourne's criteria for what qualifies as a proper instance of prop-oriented make-believe is simply too strict- there could be no prop-oriented fantasy make-believe. And there is abundant evidence in popular culture that fantasy make-believe can be a prop-oriented fiction, with sufficient principles of generation to allow hoards of fantasy-aficionados to engage in elaborate make-believe involving magic and magical items. Recreational societies have been established in which players cooperatively engage in these games and understand their rules, and real money is spent on props that are- in the make-believe- magical items that (fictionally) act in rule governed ways with no further explanation of what it is about the item that allows it to behave as it does (it is magic).

As Toon notes, "In general... 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" (2016, p. 284). And in fantasy fiction, the rules by which magical objects perform their magic cannot be exhaustively formulated (since this would be to *remove the magic* from the make-believe). Just *what it is* about the magic wand that allows it to perform certain sorts of magic, and just *what it is* about a magic potion that allows it to perform other sorts of magic, need not (and perhaps cannot) be laid out in exhaustive detail by the *principles of generation* for a prop-oriented fantasy make-believe. Nevertheless, there may be a shared sense in which players can understand that wands and potions have particular magical powers.

Bourne and Caddick Bourne claim that, "Toon's proposed folk psychological game *has not been shown* to be a genuine case of prop-oriented make-believe" (2020, p. 175). Whether or not his original formulation met its aim, Toon is free to adopt prop-oriented *fantasy* make-believe as the model for his mental fictionalism without losing its essential *prop-orientation*. Consider the following example in which children reenact the Turkish Delight scene from S.C. Lewis's *The Lion, The Witch, and The Wardrobe*:

*The child dressed as The Snow Queen offers the child playing young Edmond a plate of sweets- in the make-believe, these are enchanted Turkish Delight. The*



*objects themselves- the sweets- have magical powers: when Edmond eats the Turkish Delight, the magical food makes him care less about the welfare of his siblings and care only for eating more Turkish delight.*

What are we to understand *about* a gelatinous sweet that would allow us to grasp the powers that this food has over young Edmond? This question has no answer beyond “it is magical”, yet the presence of magical items is not a problem for prop-oriented fantasy make-believe. The make-believe proceeds successfully without specifying *what it is* about the wand, or the Turkish Delight, that gives it magical powers. And if ‘prop-oriented fantasy make-believe’ is a legitimate prop-oriented genre of fiction, then the mental fictionalist can use this as their model. The magic of magical items is left unexplained, just as the relation between behavior and mental state attributions is left unexplained. The fact that folk psychology does not point to *what it is* about certain behaviors that motivates our mental state attributions should not, therefore, cause problems for the prop-oriented mental fictionalist.

#### **4. The Make-Believe Mind as *Magic***

In the context of discussing magic as it appears in fiction, perhaps it is appropriate to say that we “understand magic”. But in other contexts it might be more appropriate to say that magic is something *beyond understanding*. Rather than *understanding* magic, we merely *suspend our disbelief*; we withhold judgment. I have suggested that, if folk psychology is to be understood as a form of fiction, then prop-oriented fantasy make-believe would be particularly attractive model. The benefit of incorporating magic within the fiction goes beyond providing mental fictionalism wide leeway regarding what sorts of explanations the fictionalization should provide. Fantasy make-believe may be the appropriate type of fiction to illuminate our folk psychological practices because, if minds did operate by our folk psychological principles, minds would be *magical*. If a fiction provides an exhaustive explanation of the mechanisms behind the “magic” in its make-believe, this explanation serves to *remove the magic* from the fiction. In a similar vein, once we have a full causal story of how a system performs its actions, we are intuitively *less inclined* to attribute mental states to that system. Where, folk psychologically, we would have once posited mental states in relation to an individual’s behavior, an exhaustive causal explanation of that

behavior seems to *remove the mental* from that individual. Daniel Dennett has long held the view that there is an uncanny parallel between our conception of magic and our folk conception of *the mind*.

It seems to many people that consciousness is a mystery, the most wonderful magic show imaginable, an unending series of special effects that defy explanation.... If you actually manage to explain consciousness, they say, you will diminish us all, turn us into mere protein robots, mere things. (Dennett, 2003, p. 7-8)

When we think about the phenomena of consciousness and wonder how they are accomplished in the brain, it is not at all unusual to fall back on the hyperbolic vocabulary of “magic”... And when one of these effects [i.e. mental states] is explained, one can sometimes observe the same disappointment, the same resistance: to explain an effect is to diminish it. (ibid, p. 10)

If a proper interpretation of folk psychology is one in which it treats mental states as if they are *magic*, then adopting the fantasy genre not only removes the restrictions Bourne and Caddick Bourne invoke in their objection to prop-oriented mental fiction, but positively strengthens the analogy between fictional discourse and mental discourse.

## **5. Additional Challenges for Prop-Oriented Mental Fictionalism**

### **5.1 The problematic disanalogy between Waltonian *principles of generation* and the principles of folk psychology**

Fantasy make-believe, complete with its magical elements, may be the best candidate model of fiction for mental fictionalism. But I would suggest that there are still significant disanalogies between fantasy make-believe and folk psychology, and that these differences are sufficient grounds for rejecting mental fictionalism whole cloth, or at least seriously reconsidering whether the costs of adopting fictionalism are worth its potential benefits.

Consider the role of *principles of generation* in a Waltonian theory of make-believe. Whereas these principles of generation- in addition to the facts about the props- make Sally's statement, "the baby is sleeping," fictionally true, we cannot say the same for the principles of folk psychology. It would be a mistake to interpret the latter as having an analogous role in *establishing* the truth value of mental state attributions. One way to put this might be that folk psychological principles do not have the same *strength of authority* as the principles of generation in a game of make-believe. It would be inaccurate to say that the principles of folk psychology, in combination with facts about the world, *provide* the truth conditions (even fictional truth conditions) for mental state attributions. Folk psychology works (when it does) because it *tracks* something true about the world. In a game of make-believe, principles of generation establish (more or less explicitly) which real-world states of affairs makes which statements fictionally true. The principles of generation *make it the case* that certain statements are true in certain contexts- there are no grounds on which to challenge the authority of these principles. The principles of folk psychology operate in a different way- folk psychological principles can be *improved*, in an objective sense, because folk psychology has an objective goal- it aims to track truth (however loosely) about minded creatures.

Objectively "improving" the *principles of generation* in a game of make believe is incoherent— there are no "better" principles of generation, because there is no goal external to the make-believe which guides the creation of these principles. The purpose of the principles is to allow players to engage in make-believe- the game is its own goal, an end in itself. If we deny that the make-believe game has no further goal, and instead claim that the goal of prop-oriented make believe is to better understand the real-world objects that serve as the props, then the theory will need to confront Bourne's and Caddick Bourne's objection head-on, and explain *what it is* that we better understand by engaging in this make-believe. Engaging with folded paper as a paper airplane may help us better understand the mechanics underlying some forms of flight; engaging in folk psychology does not help us to better understand anything that can be specified as *underlying* the agents' actions.

## 5.2 The Different Types of Indefiniteness in Fiction and Folk Psychology

This disanalogy between the function of the principles of generation in make-believe and the principles of folk psychology is connected to a further disconnect between folk psychology and fiction- the difference between the *indefiniteness* that may be found in fiction and in folk psychology. Bourne and Caddick Bourne point to this problem in their discussion of mental fictionalism, though it falls outside the scope of their criticism:

Mental fictionalists may also see the indefiniteness in FP (Dennett, 1991, p. 49) as having a precedent in fictional representation. Whether indefiniteness in FP is the same phenomenon as the indefiniteness found in fictions, which inevitably leave their fictional worlds incomplete in some respects (e.g., the color of a character's socks), requires further argument. We shall not discuss indefiniteness in this chapter. (2020, p. 169 fn1)

The indefinite aspect of both fiction and folk psychology may have encouraged some to describe folk psychological entities (mental states) as *something like* useful fictions. But there is a difference between being (i) a kind of fiction, and being (ii) *something like a useful fiction*. While the latter may be a helpful and accurate description of mental states for some purposes, this does not mean that the former, stronger, claim will hold. Mental states may well be *something like useful fictions* without this legitimizing a move to mental fictionalism.

It is unlikely that anyone would argue that the simple fact that folk psychology and fiction both have an indefinite aspect to them is *sufficient* to establish mental discourse as a type of fictional discourse. There are many ways for something to have indefinite features. In some domains, the truth value of a statement is indefinite only until the concepts involved become more precise. The truth value of the statement, "This noise is loud," may be indefinite unless or until our criteria for *loudness* is made more precise. Non-technical terms often have vague boundaries, and folk psychology is thick with non-technical terms. But further research into the objects of that discourse (the type of research conducted in developmental or comparative psychology, for example) and closer scrutiny of those concepts, can reduce some of the initial indeterminacy, if not eliminate it completely. In the domain of fiction, no further research or clarification of our concepts will suffice to eliminate- or even diminish- once we encounter the indeterminacies found there. In a practical

sense, there are always further facts to be discovered regarding the actual world, and this is not the case with fiction. In fiction, there is a point at which the “facts” simply *run out*, and the indeterminacy that remains cannot be resolved- even to a degree- by further research or conceptual analysis.

The truth value of certain statements in folk psychology (such as mental state attributions) may be genuinely indefinite under a given conceptualization of folk psychology. But, as noted above, the concepts within a folk psychological theory- our mentalistic concepts- can be objectively *improved*; there are objective standards by which changes in our conception of mental states may be *better* or *worse*, rather than merely (neutrally) different. Our conception of *emotion*, a folk psychological term, may be improved or made more precise, so that a statement which was genuinely indefinite under one interpretation of *emotion* can be given a determinate truth value under an improved conceptualization. In mental discourse, unlike in fiction, there may be cases of *genuine, but temporary, indeterminacy*.

We improve, and precisify, our mentalistic concepts not only by reflecting on those concepts themselves but by reconsidering those concepts in light of empirical discoveries made concerning creatures in the actual world.<sup>4</sup> There may be stubborn boundary cases, in which (it seems) no amount of calibrating our concepts with empirical discoveries will resolve certain questions about minds or mental states, but in the domain of folk psychology the indefinite status of many statements may be merely temporary, and might be resolved by additional work. Conceptual precision will not move the needle towards irradicating indefiniteness from fictional domains- any genuine indefiniteness in a fiction is permanent. Fictional domains have pre-determined boundaries. By their very nature, the information one can gather about a fictional domain is limited. Certain claims are made about a fictional domain, establishing its nature, and certain truths are entailed by those claims. Other fictional states of affairs are more-or-less likely, depending on what has been explicitly established in the fiction and how closely the fiction is intended to resemble the actual world.<sup>5</sup>

---

<sup>4</sup> This method by which we may improve our mentalistic concepts is what Kristin Andrews has dubbed the “Calibration Method” (see Andrews, 2014).

<sup>5</sup> This way of characterizing fiction applies more straightforwardly to the model of the fictional narrative than an actively-evolving game of make-believe. The indefiniteness of make-believe fiction may be more akin to the

I suggest that these two areas of disanalogy- (1) the difference in the *authoritative strength* of the principles of generation (for make-believe) and the principles of folk psychology (for mental state attributions), and (2) the discrepancy between the *indefiniteness* found in folk psychology and the indefiniteness found in fiction- would need to be resolved by any tenable version of mental fictionalism.

## 6. A positive suggestion:

### **Exchanging *fictions* for *levels of abstraction***

Given the problems for mental fictionalism discussed above, a more attractive option might be to treat folk psychology as a high-level explanation of agential behavior. Viewed in this way, folk psychology ceases to be a failed theory and is merely an explanation appropriate to one of many levels of abstraction, all of which are real; multiple theories may explain the same phenomenon at different level of abstraction, some higher and some lower. At this particular level of abstraction, explanations of why an agent does what she does “bottom out” in terms of beliefs and desires and sensations (or propositional attitudes and conscious experience). When a *person’s* actions are our explanandum, beliefs and desires and sensations may be explanatory bedrock; at this level of abstraction, the existence of a person’s mental states may be a *brute fact*. At lower levels of abstraction- levels at which our explanandum is not a person, *per se*, but a biological organism- explanations will bottom out in terms of biological or neurological processes rather than mental states. There will be no precise mapping between folk psychological explanations and neurological explanations because these explanations illuminate their subject matter in different ways, treating different partitions of the world as constituting the appropriately identifiable patterns within the same phenomenon.

This allows us to treat mental states as *real patterns* (Dennett, 1991), existing at a higher level of abstraction than neurological states. In doing so we can retain something important about mental

---

indefiniteness found in folk psychology. In both make-believe and in mentalistic discourse, there may be genuine temporary indefiniteness, which can be resolved by revising the principles of each. Of course, as noted above, the kind of revision that would constitute an *improvement* on those principles differs significantly between the two cases.

state attributions that fictionalization is forced to discard. If folk psychology is a high-level explanation that succeeds when it tracks truth, then folk psychology remains *fallible*, as it should. And it is consistent with our actual folk psychological practices that we can *wrong* about some of our mental state attributions, even if those mental state attributions would be deemed correct according to our current folk psychological principles.<sup>6</sup> Folk psychology takes itself to be tracking truth, not *creating* it. When we ask “Can elephants feel sad?” this question isn’t answered simply by referring to our folk psychological conceptions of sadness and seeing whether *what’s going on* with elephants fits our extant notion of sadness. Our notion of *what sadness is* should be open to revision- and there should be a difference between a *mere change* in the theory and a *genuine* improvement in the theory.

This returns us to the previous objection to mental fictionalism: an essential feature of folk psychology is that it has room for genuine *improvement*, rather than mere alteration. Prop-oriented make believe is an attractive model for mental fictionalism because it allows facts about the actual world to play a role in determining the (make-believe) truth. But to be properly analogous to our practice of folk psychology, there must be room for our principles of generation to be *getting it wrong*, too. Mental *fictionalism*, modeled on any form of fiction, may concede too much authority to the principles of folk psychology.

In practice, it may certainly be helpful to occasionally conceive of mental states as a kind of *useful fiction*, but we should be wary of taking this notion of *the mental as fiction* too seriously. It is difficult to conceive of multiple levels of explanation for a single event, each of which illuminates the event at a different level of abstraction, and none of which are “closer to the truth” than others. Do parents care for their offspring *because they love their children*, or *because they have a biological drive to create and protect their offspring*, or *because they are driven by chemical reactions in their brains*? Intuitively, these explanations seem to compete with one another- if the answer to one is “yes”, it seems that the answer to the others must be “no”. It is hard to conceive that all three can be true, simultaneously, without any explanation undermining the others. In response to this difficulty, we may sometimes characterize beliefs and desires, and love, as *useful*

---

<sup>6</sup> The idea that folk psychology takes itself to be fallible in this way is admittedly at odds with the characterization of folk psychology within some forms of mental fictionalism (see Wallace, 2013).

*fictions*, but this is best understood as a hermeneutical crutch- a characterization of something complex in a way that is both more familiar and *in the same neighborhood* as what we're actually trying to grasp. But there is a danger in leaning too hard on these conceptual crutches. Some metaphors and other hermeneutical devices may act as what Daniel Dennett calls a *boom crutch*, "the ones that only seem to aid in understanding but that actually spread darkness and confusion instead of light" (Dennett, 2013, p. 14). I would suggest that thinking of folk psychology- and mental states- as akin to *some kinds* of fiction may aid in understanding *to a certain extent*. For instance, it may help us move past the trap of assuming that, if physicalism is true, then everything *real*- like beliefs and sensations- must systematically reduce to more fundamental physical elements. (Money is *real*, but not in the same way that this desk is real. *Romeo and Juliette* is a *real* play, while *Burgfil the Terrible* is not. Nonetheless, both *money* and the play *Romeo and Juliette* are types of fiction.)

Insofar as mental states do not belong in the same ontological category as tables and chairs and arms and legs, thinking of them as *something like fiction* may genuinely aid our understanding of what they are (or, at the very least, what they are not). But if we lean too hard on the fiction analogy, we are guaranteed to lose a vital feature of our mental state attributions- *that they aim to track truth*, even if that truth may be found only at a fairly high level of abstraction. If this is correct, then the benefit of adopting mental fictionalism- either that it gives us a way to treat our mentalistic statements as *truth-apt*, or that it eases the tension between Quine's criterion for ontological commitment and the indispensability of our mental discourse- simply isn't worth the cost.

Mental fictionalism offers a novel and potentially fruitful position in the logical landscape. If successful, it could provide us with resolutions to longstanding philosophical puzzles. Though I have raised challenges to the theory here, I am hopeful that these challenges can be met. Despite the problems I describe above I feel the intuitive draw of placing mental entities and fictional entities in the same general category (or at least the same general neighborhood). The idea of mental fictionalism remains appealing in many ways, as thinking of mental states as *something akin to fiction* can- at times- genuinely aid our understanding of the mental.



## References

- Andrews, K. (2014) *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*. Abingdon, Oxon; New York, NY: Routledge.
- Bourne, C. and Caddick Bourne, E. (2020) 'Folk Stories: What has fiction to do with mental fictionalism' in Armour-Garb and Kroon (eds.) *Fictionalism in Philosophy*. Oxford: OUP, 168-186.
- Churchland, P. M. (1981) 'Eliminative Materialism and the Propositional Attitudes', *The Journal of Philosophy*, 78(2), p67–90.
- Demeter, T. (2013) 'Mental Fictionalism: The Very Idea', *The Monist* 96, p493-504.
- Dennett, D. C. (1991) 'Real Patterns', *Journal of Philosophy*, 88, p27-51.
- \_\_\_\_\_. (2003) 'Explaining the "Magic" of Consciousness', *Journal of Cultural and Evolutionary Psychology*, 1: p7-19.
- \_\_\_\_\_. (2013) *Intuition Pumps and other tools for thinking*. New York: W.W. Norton & Company.
- Hutto, D. (2013) 'Fictionalism About Folk Psychology', *The Monist*, 96, p582– 604.
- Joyce, R. (2013) 'Psychological Fictionalism, and the Threat of Fictionalist Suicide', *The Monist*, 96, p517– 538.
- Levy, A. (2011) 'Information in Biology: A Fictionalist Account', *Noûs*, 45, p640– 657.

- Lewis, D. (1983) 'Mad Pain and Martian Pain', *Philosophical Papers, Vol 1*, Oxford: OUP, p122-130.
- Parent, T. (2013) 'In the Mental Fiction, Mental Fictionalism Is Fictitious', *The Monist*, 96(4), p605–621.
- Putnam, H. (1963) 'Brains and Behavior', in Butler R. J. (ed.) *Analytical Philosophy, Second Series*, Oxford: Blackwell, p211-235.
- Quine, W. V. O. (1948) 'On What There Is', *The Review of Metaphysics*, Vol. 2, No.5, p21-38
- \_\_\_\_\_. (1980) 'Logic and the Reification of Universals', in *From a Logical Point of View*. 2<sup>nd</sup> edn. Cambridge, MA: Harvard University Press.
- Toon, A. (2016) 'Fictionalism and the Folk', *The Monist*, 99, p280-295.
- Wallace, M. (2007) *Mental Fictionalism*, Unpublished manuscript.
- Wallace, M. (2016) 'Saving Mental Fictionalism from Cognitive Collapse', *Res Philosophica*. 93, p405-424.
- Walton, K. (1990) *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.
- \_\_\_\_\_. (1993) 'Metaphor and Prop-Oriented Make Believe', *European Journal of Philosophy*, 1, p39-57.
- Yablo, S. and Gallois, G. (1998) 'Does Ontology Rest on a Mistake?', *Proceedings of the Aristotelian Society, Supplementary Volume*, 72, p229-261.

