From monkeys to infants: the empirical challenges facing mental fictionalism

Gavin Robert Foster

Published online: 09 Sep 2025.

Submit your article to this journal ↗

Article views: 14

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

From monkeys to infants: the empirical challenges facing mental fictionalism

Gavin Robert Foster

Department of Philosophy, Purdue University, West Lafayette, IN, USA

ABSTRACT

In recent decades, a novel theoretical account of folk psychology has emerged that challenges traditional assumptions: mental fictionalism. Rather than treating mental states as things in creatures' minds that we are really tracking, mental fictionalists argue that folk psychology is best understood as a collective fiction that we partake in. On this account, it is this *fiction* rather than an ability to track mental states that enables us to predict/explain one another's behavior, or to coordinate our affective states to grow and maintain social relationships. I provide an overview of the evidence from comparative and developmental psychology that is inconsistent with mental fictionalism, arguing that both views considered here are unable to account for the mindreading abilities of non-linguistic primates and children. I then offer an ontogenetic challenge, arguing that mental fictionalism faces the difficult task of explaining how mindreading could emerge from what are thought to be more sophisticated cognitive mechanisms, such as those underlying participation in either pretense or collective narratives. I conclude by highlighting the specific body of evidence that a more robust version of mental fictionalism should account for, especially as it continues to be refined and emerges as a serious competitor amongst interpretations of folk psychology.

1. Introduction

The ability to track and attribute mental states to other creatures is an indispensable part of all of our lives. We predict how our friends will react to exciting news, explain how your child's remark hurt someone's feelings, or predict what kind of drink your colleague wants. We do this seamlessly, often automatically, and it helps structure our social relationships. This capacity, termed mindreading or theory of mind, lies at the heart of our *folk psychology*, the everyday process of talking and thinking about the mental states of others, thereby enabling us to predict and explain their

CONTACT Gavin Robert Foster ✉ foste362@purdue.edu 💻 Department of Philosophy, Purdue University, West Lafayette, IN, USA

behavior. In recent decades, a novel theoretical account of folk psychology has emerged that challenges traditional assumptions: mental fictionalism. Rather than treating mental states as things in creatures' minds that we are really tracking, mental fictionalists argue that folk psychology is best understood as a collective fiction that we partake in. On this account, it is this *fiction* rather than an ability to track mental states that enables us to predict/ explain one another's behavior, or to coordinate our affective states to grow and maintain social relationships.

The primary motivation behind mental fictionalism is to carve out a middle ground between eliminative materialism about the mind on the one hand and a robust realism on the other. Advocates of the former (e.g., Churchland, 1980) argue that the propositional attitudes picked out by folk psychology do not really exist and should instead be eliminated from our vocabulary. Advocates of the latter (e.g., Fodor, 1980) deny this, insisting that folk psychological talk refers to real entities in our minds. Fictionalists agree with the ontological conclusions of the eliminativist: it is wrong to think there is some "belief box" in the brain that we are referring to when we engage in mindreading. However, they disagree with the eliminativist that we should, or even *could*, eliminate folk-psychological practice from our lives. Instead, they agree with the realist that folk psychology is indispensable. Thus, mental fictionalism belongs to a family of views, such as interpretivism or Dennettian intentional stance theory, that attempts to retain the attractive components of eliminativism and realism while avoiding their undesirable consequences.[1]

Most accounts of folk psychology have attempted to make sense of the vast literature on mindreading in both developmental and comparative psychology. The last few decades have produced robust evidence that non-human primates and pre-linguistic infants are adept at tracking the mental states of other creatures. In response, researchers have developed theories of mindreading that chart its developmental and evolutionary history (e.g., Andrews 2012; Schlicht, 2023; Spaulding, 2018; Wellman, 2014). Importantly, mental fictionalists have so far not provided *any* interpretation of this literature, and I argue that this is a fundamental problem with the view as it is currently defended. My aim in what follows is to expound on this, arguing that two core components of mental fictionalism, *factualism* and *fictionality*, force it to confront the empirical literature. By considering two leading strains of mental fictionalism – Toon's pretense-based view and Demeter's affective-based view – I suggest that mental fictionalism generally overestimates the cognitive abilities required for mindreading.[2] These fictionalists construe folk-psychological practice as complex acts of make-believe or affect-coordination through participation in shared linguistic narratives. However, it remains unclear how to make sense of the evidence for mindreading in non-linguistic animals and infants, as well as the

developmental story of how mindreading could evolve from such cognitive mechanisms. Ultimately, I argue that if we take the comparative and developmental psychology literature seriously, then mental fictionalism must provide answers to these challenges in the way that competing accounts of folk psychology do.

After introducing what both versions of mental fictionalism have in common, I provide a brief sketch of each of their individual views. Next, I provide an overview of the evidence from comparative and developmental psychology that is inconsistent with mental fictionalism, arguing that both views are unable to account for the mindreading abilities of non-linguistic primates and children. I then offer an ontogenetic challenge, arguing that mental fictionalism faces the difficult task of explaining how mindreading could emerge from what are thought to be more sophisticated cognitive mechanisms, such as those underlying participation in either pretense or collective narratives. I conclude by highlighting the specific body of evidence that a more robust version of mental fictionalism should account for, especially as it continues to be refined and emerges as a serious competitor amongst interpretations of folk psychology.

2. Two varieties of mental fictionalism

Fictionalism in philosophy encompasses a wide range of views covering distinct domains. In its most uncontroversial form, there is literary fictionalism: the view that the content of literary sentences (e.g., as contained in novels) does not entail a commitment to existence of the states of affairs they describe. There is also fictionalism in philosophy of mathematics, which argues that talk about numbers is really a disguised fiction that does not entail a commitment to the literal existence of numbers as such. Fictionalism about possible worlds holds that while modal talk is indispensable for counterfactual reasoning, it is merely part of a shared narrative we accept – without committing to an ontology of possible worlds. Similarly, moral fictionalism treats moral discourse as a convenient and practically useful fiction without assuming that moral values or duties have objective force. Following Kroon et al. (2019, p. 17), one can identify three core elements of fictionalism in general: (i) factualism – the statements made possess truth-value, albeit within a fictional domain; (ii) utility – there is a kind of indispensability involved in committing to the fictional discourse that would be lost if one attempted to excise it from our ordinary, everyday practice; and (iii) fictionality – the entities of the relevant domain do not actually exist but are rather spoken of *as though* they existed, and so are fictional. Mental fictionalists explicitly accept all three elements of general fictionalism, though they differ with respect to how their own particular interpretation implements them. Mental fictionalism integrates these three

elements by construing statements about "beliefs" or "desires" as metaphorical utterances that do not refer to anything real in the head, but that hold considerable practical value.

The tension between mental fictionalism and the empirical evidence stems from its commitment to both factualism and fictionality. In particular, the problem is that while mental fictionalists treat folk-psychological claims as *truth-apt* within their fictional frameworks, they also deny that genuine mental states exist as internal representations in other agents. This move commits the fictionalist to the claim that mindreading is a form of make-believe or narrative practice, rather than the genuine tracking of inner mental states.

As such, the fictionalist's stance makes folk-psychological practice dependent upon the ability to participate in a linguistically mediated game: only agents who can formulate truth-apt propositional thoughts can play it. Accordingly, attribution of inner mental representations like states of knowledge or desire is impossible for non-linguistic creatures, since their fictional truth-conditions require linguistic competence (e.g., Demeter, 2022, p. 142; Toon, 2022). However, as I argue, the empirical evidence from non-human animals and pre-linguistic infants show that folk-psychological creatures *are* attributing or are sensitive to mental states absent the required pretense or linguistic ability.

2.1. Toon's pretense-fictionalism

According to Toon's *pretense-fictionalism*, folk-psychological attributions function as acts of pretense rather than literal ascriptions of inner representations in other agents. His account is directly influenced by Kendall Walton's work in philosophy of aesthetics and in particular on his account of how agents engage in games of pretense/make-believe. Such games of pretense are often used *metaphorically*, utilizing a secondary domain of discourse to draw attention to features of a primary domain of discourse. Games of pretense thus employ metaphorical language suited to different purposes.

What is required are both *props*—things in the real world that serve as anchors for imaginings – and *principles of generation*—the rules governing how these props prompt the relevant fictional imaginings. These imaginings amount to sets of fictional truths for the purposes of engaging in a game. To take one of Walton's examples, consider a pair of boys, Eric and Greg, who are playing together in the woods. The boys agree that, in their world, any stumps they come across will actually be *bears*. The props of the game, then, are the *stumps*, and the principle of generation is that *stumps are bears* (Walton, 1990, p. 38). The set of stumps in the forest, along with the agreed-upon principle of generation,

produce a set of *fictional truths* for Eric and Greg (and for whoever else opts into the game). Recognition (imaginings) of these propositions are prompted through encountering the relevant props of stumps, creating a fictional world for Eric and Greg. Walton also distinguishes between *prop-oriented* pretense and *content-oriented* pretense. In the former, the game of pretense is used to understand properties and features of the props themselves, whereas in the latter, the aim is to use pretense for the sake of the content that arises in the fictional world. The stump game of pretense is content-oriented: the stumps are adopted as props merely to "fill out" the imaginative world, as opposed to try and understand features of the stumps themselves.

Toon argues that Walton's framework applies equally to how agents participate in folk-psychological practice, construing it as a form of prop-oriented pretense. He takes the relevant props to be other agents (human *or* non-human) and their sequences of behavior. Thus, the primary domain (the props) to be elucidated or understood is agential behavior and this is done through the secondary domain of folk-psychological discourse. The purpose of this discourse is to "explain and predict people's behavior," the props of the game of make-believe, and is therefore classified as an instance of prop-oriented pretense. Upon witnessing the right kind of behavior, Toon claims that the principles of generation prompt us to "pretend that people possess . . . inner worlds, just as we pretend that the events in some-one's life are stages in a journey, or Italy is a boot" (Toon, 2023, p. 23). Factuality is combined with fictionality because "by pretending to say that people harbour inner representations, we manage to say something true about their behaviour" (Ibid.). Following Walton, Toon emphasizes that participants in the game of folk psychology need not be (and usually *are not*) aware of their engaging in pretense (2016, 283).

To illustrate pretense-fictionalism, consider a modified version of one of Toon's recurring examples: Jane and Emma just ran into Mark at the bus station and, when asked, he says he is in a rush to get to the Marble Arch on Oxford Street. Shortly thereafter, the No. 73 bus pulls up, he waves goodbye, and he gets on the bus. On this basis, Jane and Emma can reasonably think to themselves "Mark remembered that the No. 73 bus goes to Oxford Street" (Toon, 2023, p. 23). On Toon's account, when they form this thought, they are not *actually* attributing an inner representation to Mark – there is no internal memory in his mind whose content involves the No. 73 bus and Oxford Street. Instead, they are using the folk-psychological concept of "memory" to explain his behavior: getting on the bus when it pulled up. Equally, one could use such concepts to help predict future behavior. As Toon puts it, "by invoking a game of pretense that treats memory as an inner notebook, we give ourselves a convenient . . . way of describing someone's behavior" and that "[p]ut simply, Mark behaves *as if* he had 'the No. 73 goes

to Oxford Street' written in his inner notebook – even though there is no such notebook" (Toon, 2023, pp. 22–3).

Though the principles of generations underlying forms of pretense may often be implicit (Walton, 1990, p. 41) it must be the case that there is a deep stability underlying folk-psychological practice. Consider the alternative: if the principles of generation underlying folk psychology were transient, then one person's use of terms like "belief" or "desire" could cease to cohere with another's, leading to radical confusion. Though error is possible, Toon seems to maintain that participants in folk-psychological pretense are not engaging in some kind of error beyond mistaking the discourse for describing the inner-representations of other agents. It is hard to envision how it could be otherwise. The stability is what guarantees that someone attributing "memory" to Mark, for example, is genuinely participating in the same pretense game as another who listens and understand the utterer. Without this shared, stable framework, the collective make-believe essential to Toon's prop-oriented pretense would dissolve into individualized imaginative play without shared principles of generation.

2.2. Demeter's affective-fictionalism

In contrast to Toon's pretense-fictionalism, Tamás Demeter's affective-fictionalism rejects the claim that folk-psychological practice is primarily aimed at explaining or predicting the behavior of other agents. Instead, Demeter insists that "psychological practice is a narrative practice" in that "it is a tool for weaving narratives that make the social world familiar to us" because "coherent narratives are valuable to us in themselves" (Demeter & Sugden, 2013, p. 489). Unlike Toon, he does not take folk-psychological practice to have epistemic ends. According to Demeter, two agents partaking in folk-psychological practice amounts to a form of *affective communication* in which "the parties conform to shared conventions while attempting to convey affective states" (Demeter, 2022, p. 143). He states that there must be "two sets of conventions": (a) those that "stabilise paradigmatic narrative structures as expressions of affective states" (Ibid., 143); and (b) those that "stabilis[e] the interrelation of various narrative elements" (Ibid., 145). The former are established through repeated instances of linguistic communication that successfully coordinate affective states between two agents.

To see how affective-fictionalism works, consider again the previous example of Mark at the bus stop. Suppose that rather than just forming the thought, Jane turns to Emma and utters that "Mark remembered that the No. 73 bus goes to Oxford Street." On Demeter's view, one might claim the following: Jane is not actually aiming to predict Mark's future movements nor represent his inner representational memory states.

Instead, this utterance serves a narrative role in weaving a coherent picture of Mark's behavior that makes sense to her and Emma (and whoever else is in the vicinity). Jane is initiating a "story" that aims to orient both her and Emma's affective states. Perhaps, for example, she is attempting to communicate that Mark is not someone who is easily lost in the city of London, or, similarly, that he would be a trustworthy individual to follow if one wished to get around London's public transit system. Her utterance thus helps both her and Emma form a consistent set of affective states directed toward Mark: that he is trustworthy, reliable, and so on. Alternatively, Jane could be aiming to communicate her affective state to Emma: she may want Emma to share in her relief that Mark is going to arrive at his destination. In that case, the utterance is not about "tagging" Mark as someone who is reliable but rather involves Jane's trying to shape (or perhaps clarify) Emma's affective states as directed toward Mark. According to Demeter, this kind of folk-psychological talk aims at securing affective alignment/cohesion rather than predicting or explaining behavior.

3. The challenge from non-linguistic mindreading

Though the bulk of mindreading research is now done by developmental psychologists, it actually originated in the comparative domain with the publication of Premack and Woodruff's (1978) seminal "Does the chimpanzee have a theory of mind?" In the decades following, both comparative and developmental psychologists have drawn and improved upon each other's experimental designs to test for different facets of theory of mind in non-human primates and children. On the developmental side, Wimmer and Perner's (1983) false-belief protocol initiated a wave of studies investigating the emergence of mindreading abilities in preschool-aged children. Later, developmental psychologists developed non-verbal mindreading experiments involving looking-time and eye-tracking methods for use on pre-linguistic infants, which were then adapted by comparative psychologists for use on apes, monkeys, and other animals. This iterative exchange of experimental design has enabled a more nuanced picture of how more sophisticated forms of mindreading (like false-belief understanding) can arise prior to linguistic ability both ontogenetically and phylogenetically (for good overviews, see Lewis & Krupenye, 2022; Wellman, 2014).

In this section, I highlight some examples of non-linguistic mindreading from both the comparative and developmental literatures. I first give examples of *cognitive* mindreading involving the tracking of informational states such as knowledge and awareness, and then I provide examples of *conative* mindreading involving the tracking of action-motivating states such as goals, desires, and intentions. For the sake of space, I will give examples

from the comparative literature for cognitive mindreading, and examples from the developmental literature for conative mindreading. However, in a footnote I provide corresponding studies showing the same findings in the other field.

3.1. Cognitive mindreading in non-human primates and pre-linguistic infants

One of the earliest forms of mindreading involves the capacity to track what another agent knows about where objects are located in the environment.[3] For example, in a classic experiment by Hare et al. ([2000](#)), a submissive and a dominant chimpanzee were placed behind gates on opposite sides of a large rectangular room. Two high-value food items were placed in the chamber-: one behind an opaque barrier (only visible to the submissive chimpanzee) and the other behind a transparent barrier (visible to both). At the start of each trial, each chimpanzee's gate was raised slightly so that they could look into the chamber and also observe each other's gaze patterns. The authors hypothesized that if the submissive chimpanzee was capable of tracking the dominant chimpanzee's perceptual state, then it should go to the food hidden behind the opaque barrier, thereby avoiding confrontation over the food visible to both chimpanzees adjacent to the transparent barrier. Controls were implemented to rule out deflationary interpretations, such as the possibility the submissive chimpanzee was merely tracking the other chimpanzee's line-of-gaze. Based on their results, as well as similar protocols showing the same findings, the authors concluded that there is good evidence that in the right kind of scenarios, "chimpanzees know what conspecifics do and do not see . . . " (2000, 783). Competitive experiments like Hare et al. ([2000](#)) provide strong evidence that mindreading emerged long before the linguistic practices or kinds of pretense required for the participation in group fictions that the mental fictionalists state exhaust the function of folk psychology.[4]

In the last decade and a half, psychologists have pursued a second approach by using eye-tracking to detect implicit mindreading – an automatic, lower-level sensitivity to the mental states of other agents. In the initial study (Marticorena et al., [2011](#)), experimenters approached adult macaques with an apparatus consisting of two boxes on either side of a platform, and a lemon that could be surreptitiously moved across the platform into either box by the experimenter behind the apparatus. The macaques watched as the experimenter observed the lemon move from the left box to the center of the platform, before returning to the left box immediately after. Next, in the "consistent" condition, the experimenter then reached into the box they had observed the lemon enter (left),

whereas in the "inconsistent" condition, the experimenter reached into the box they had not observed the lemon enter (right). As would be expected if the macaques were successfully tracking the perceptual state of the experimenter, they looked for a statistically significant longer period of time during in the inconsistent condition than in the consistent condition. This suggests that the macaques had attributed a knowledge state to the experimenter – and were surprised when the experimenter's actions ran contrary to that presumed knowledge. This experiment, and variations on it, have been repeatedly replicated over the last decade (e.g., Drayton & Santos, 2018; Horschler et al., 2019, 2021; Martin & Santos, 2014).[5]

3.2. Conative mindreading in non-human primates and pre-linguistic infants

In one prototypical example by Carpenter et al. (1998), twelve- to eighteen-month-old infants watched an experimenter perform a sequence of actions that led to a particular outcome (turning on a light). In the intentional condition, the experimenter made clear they were trying to turn on the light, whereas in the other condition they were marked as accidental. The children tended to imitate the unsuccessful action (turning on the light) in the intentional condition but not in the accidental condition. In a related study utilizing the previously discussed violation-of-expectations paradigm, ten- and twelve-month-olds were more surprised when a previously unsuccessful agent failed to perform their action when an obstacle was removed, even though the infant had never seen the action successfully completed (Brandone & Wellman, 2009). This and experiments like this demonstrate that at a young age, children are sensitive to the intentions of other agents.[6]

While infants are sensitive to desires states in others that they share (objective desires) as young as twelve-months-old (Gergely et al., 1995), the ability to recognize that another agent may wish to act on desires different from their own (subjective desires) emerges slightly later, at eighteen months. In Repacholi and Gopnik's (1997) seminal experiment, an experimenter expressed desire for broccoli as compared to goldfish. When the child was prompted to give the experimenter what they wanted, eighteen-month-olds (but not fourteen-month-olds) overrode their own preference for goldfish to give broccoli to the experimenter. Further, research has shown that children also attribute emotional states depending on the degree of desire satisfaction (e.g., Harris et al., 1989). As Wellman summarizes, these studies "show ways in which 2-year-olds [and younger] appreciate desires and emotions as internal mental states separate from external objects or overt behaviors . . . " (2014, p. 76).[7]

3.3. Mental fictionalism and non-linguistic mindreading

The swathe of evidence of mindreading in non-human primates and pre-linguistic infants poses a challenge for the mental fictionalist. On either of the views discussed here, children and primates should not engage in folk-psychological practice until they have the requisite cognitive abilities – pretense in Toon's framework, or collective narrative participation in Demeter's – that allegedly underwrite folk psychology. In particular, the combination of factualism and fictionality commits the mental fictionalist to the claim that folk psychological practice is exhausted by the ability to form linguistic practices that acquire their truth-value only within a fictional domain. This represents a broader difficulty for mental fictionalism: while its motivations stem from features of full-fledged adult folk-psychological practice – features that *do* often involves language, narrative participation, and sophisticated pretense – a successful interpretation of folk psychology must also account for more minimal forms of mindreading in non-linguistic primates and infants. By focusing too narrowly on adult practice, mental fictionalism risks mistaking a slice of folk psychology for its entire structure.

For the mental fictionalist, the most obvious response to these findings would be to claim that the experiments do *not* show that non-linguistic primates and infants are engaging in folk-psychological practice. Rather than being mindreaders, they are merely sophisticated *behavior-readers* – through a domain-general learning process (perhaps a sophisticated kind of associative learning), they utilize rules that correlate a kind of stimulus in other agents with a future sequence of behavior. For example, rather than tracking the perceptual state of the experimenter in the eye-tracking study discussed above, the macaque could be deploying a rule such as "If the experimenter looks last at the left box, then they will subsequently reach towards the left box." In fact, for *any* experiment (including those discussed), it is possible to simply "snip" the relevant mental state out of the mindreading interpretation to generate a "complementary behaviour-reading explanation" (Andrews, 2016, p. 306). Mental fictionalists may then argue that alleged cases of folk-psychological practice in primates and infants are instead cases of mere behavior-reading.

There are strong methodological reasons to be suspect of *ad hoc* behavior-reading hypotheses that are always posited after an experiment has yielded *prima facie* evidence for a mindreading hypothesis. As Fletcher and Carruthers say, this method of positing behavior rules after the fact "is too underspecified to make determinate predictions, hence there is no risk of it turning out to be wrong" and also provides "no clues as to how the animals [or infants] might be expected to react in particular circumstances and suggests no potential lines of experimentation" (Fletcher & Carruthers, 2013, p. 83). In effect, the programme of universal behavior-reading is

unfalsifiable. This point becomes most obvious when you consider just how different these behavior rules are given the range of experiments that have been run. Consider again Hare and colleagues' barriers experiment and suppose that they had not implemented a control to rule out a direction-of-gaze behavioral rule, i.e., "if the chimpanzee looks towards barrier one, then it will go towards barrier one." While this might work for *this* experiment, it is unclear how exactly it can make sense of experiments that involve mindreading of a different modality like hearing (e.g., Melis et al., 2006). Here a different behavior-rule must be posited unconnected to direction-of-gaze. This problem does not arise if what is being tracked are the relevant mental states. In other words, treating non-linguistic creatures as at least minimal mindreaders yields *testable* predictions whereas the behavior-reading approach provides no unified framework for new experiments. The approach is more than just inelegant – it stifles genuine progress in understanding the social cognitive abilities of non-human animals and pre-linguistic infants.[8]

4. The challenge from the late emergence of fictional participation in ontogeny

If these creatures are only engaging in behavior-reading, what is the developmental story of how mindreading emerges? By identifying the capacity to mindread with the capacity to either engage in pretense or participate in shared fictional narratives, the mental fictionalist inherits a substantial evidential burden. The forms of make-believe or narrative participation that the fictionalist appeals to are absent in non-human primates, and where such capacities are observed in children, they are not sufficiently sophisticated to support the kinds of stable, structured attributions characteristic of folk psychology. Crucially, these imaginative practices emerge only after the development of language, and well after the age that most psychologists take children to already exhibit the ability to mindread.[9]

Make-believe must be distinguished from forms of functional play. Following the standard account, pretense can be defined operationally as involving one of three sequences of behavior: (i) object substitution; (ii) attribution of pretend properties; and (iii) use of imaginary objects (Leslie, 1987, p. 414).[10] Here I will focus on just the first kind of pretense: object substitution in infancy. In order for mental fictionalism to be consistent with the evidence, pretense must emerge at essentially the same time that the capacity to mindread does. If to engage in folk-psychological practice (mindread) is to engage in a kind of pretense, then obviously mindreading cannot emerge prior to the ability to make-believe.

Infants first begin engaging in object substitution around nineteen months old (Lillard, 2015, p. 440). However, this is only possible when

presented with a *model* of the action to be performed. For example, children under the age of two might use their finger to act "as if" they are brushing their teeth but only immediately after seeing an adult experimenter perform the same action (Jackowitz & Watson, 1980). Hence, this amounts to a kind of mimicry that falls short of genuine pretense on Toon's construal, which requires that children spontaneously adopt the relevant principles of generation, treating some prop (in this case, a finger) *as if* it were a genuine toothbrush and yielding a fictional truth.[11] Spontaneous, non-model-based object substitution only emerges after the age of three (Lillard, 2015, p. 441). For example, it is at this point that children begin to partake in the cases of make-believe that Toon (drawing on Walton) discusses in his book, such as pretending that a doll is sleeping or that a stick is a sword or magic wand. However, even at this point in development, children would be unable to meet the requirements mental fictionalists think are required to engage in folk-psychological practice.

Recall that for pretense-fictionalism to be plausible, the principles of generation underlying folk psychology must be robust and stable. The same applies to the kind of narratives that underwrite folk-psychological practice on affective-fictionalism. Thus, a further problem emerges for the fictionalist in accommodating the behavioral evidence, for while it is true that children engage in pretend play from an early age, there are well-known issues with their ability to maintain the appearance-reality distinction (e.g., Bourchier & Davis, 2002). For example, in an influential study by Harris and colleagues, children between four and six years old were asked to imagine the existence of a make-believe creature within a transparent box. After affirming to the experimenter that there was no creature in the box and being told the experiment was over, younger children still approached the transparent box with caution. Further, "in the final interview ... [the children] wondered whether there was a creature in the box" (Harris et al., 1999, p. 120). While there remains open debate about what causes these reality-pretense breakdowns, there is consensus that these confusions are common amongst children as old as six years (Lillard, 2015, pp. 443–6).

This evidence suggests that despite being capable of engaging in one-off instances of pretense, children are nonetheless unable to adopt the kind of stable principles of generation required by the pretense-fictionalist. Further, the failure to maintain the distinction between appearance and reality also speaks against the prerequisite ability to partake in collective narrative that affective-fictionalism requires. If children cannot reliably sustain the "story" that demarcates reality from a fictional world, it is reasonable to think they cannot also maintain the kind of conventions required to align and coordinate affective states with other agents via folk-psychological practice. Consequently, both pretense- and affective-fictionalism rely on cognitive

abilities that appear fragile or absent in young children's early folk-psychological practice.

The developmental evidence thus creates a distinct problem for mental fictionalism. Even when toddlers do begin to engage in make-believe, their early play is *ad hoc* and unstable. Young children are unable to maintain a distinction between appearance and reality required to become a participant in the game of folk psychology with its stable, systematic principles of generation or collective narratives. Thus, even if initial pretend play or narrative participation arose simultaneously with mindreading, it would not exhibit the structure demanded by mental fictionalism. The fictionalist must provide a developmental account of how mindreading is supposed to emerge from the sophisticated cognitive mechanisms that are meant to exhaust folk-psychological practice. By contrast, rival accounts of folk psychology begin with the early, more minimal forms of mindreading observed in infants and explain how full-fledged folk-psychological practice is gradually scaffolded upon them as more complex capacities emerge through ontogeny. Mental fictionalism, in starting from the fully developed adult practice, appears to overlook the developmental trajectory that any adequate theory of folk psychology must capture.

5. Conclusion: twin challenges to mental fictionalism and a path forward

I have argued that two major versions of mental fictionalism face twin challenges stemming from both comparative and developmental psychology. On the one hand, evidence from comparative and developmental psychology shows that non-linguistic primates and pre-linguistic infants are capable of engaging in forms of mindreading. Mental fictionalists must provide an explanation of this evidence in a way that is not merely *ad hoc*. On the other hand, given that the sophisticated cognitive abilities required for folk-psychological practice on fictionalist accounts appear to emerge ontogenetically later than the ability to mindread, fictionalists must provide a developmental story about how the ability to mindread emerges. The absence of such a story leaves an explanatory void exactly where rival interpretations of folk-psychology claim empirical traction.

It must be stressed that both pretense- and affective-fictionalists provide *descriptive* (hermeneutic) accounts rather than *revisionary* accounts of folk psychology (Stanley, 2001). That is, they seek to describe what is actually occurring when creatures engage in folk-psychological practice, as opposed to offering a way in which we *should* think about it. However, a genuinely descriptive theory of folk psychology can only be plausible if it provides an explanation of its ontogeny from infancy to childhood. Furthermore, given that both affective- and pretense-fictionalists identify linguistic competence

as a precondition for mindreading, they must provide a phylogenetic story about why and how folk-psychological practice first emerged. That is, why would human beings have evolved to read each others mind only *after* gaining the ability to use language?

If all but language-competent humans are mere behavior-readers, they will be forced to explain away even the most compelling evidence from comparative psychology, whereas alternative conceptions of folk psychology *leverage* this evidence to support their conception of mind-reading (e.g., Andrews 2012; Schlicht, 2023). By adopting the behavior-reading approach, mental fictionalism's construal of folk-psychological practice would fit poorly with the evolutionary evidence of incremental gains in mindreading ability across primates (Lewis & Krupenye, 2022). Alternatively, the fictionalist could argue that non-human primates too are engaging in shared make-believe. However, to put a long story short, there is no evidence of anything close to what is demanded by pretense- or affective-fictionalist accounts of make-believe in even our closest primate relatives, chimpanzees and bonobos. Besides some isolated anecdotal cases that have not been subject to any of the controls required to establish such a sophisticated cognitive ability (Matsuzawa, 2020; Mitchell, 2007), it appears that the capacity to engage in make-believe or participate in collective narratives is a uniquely human ability.

A more robust mental fictionalism will then need to do at least three things. First, it must explain away the evidence of non-linguistic mind-reading in primates and infants. Second, it must do so in a way that does not rely upon *ad hoc* behavior-reading accounts and instead offers novel behavioral predictions that distinguish the fictionalist account from other accounts of folk psychology. Finally, it must provide both an ontogenetic and phylogenetic account of how the ability to mindread is inextricably linked to the capacity to participate in shared fictions, whether that be through the capacity to engage in pretense or in collective narratives. Meeting these demands would transform mental fictionalism from an interesting conjecture to a serious contender amongst other interpretations of folk psychology.

I want to conclude by suggesting a general strategy that the fictionalist should take toward accomplishing these tasks. What generates the twin challenges for mental fictionalism is that both pretense- and affective-fictionalism identify the capacity for folk-psychological practice with an activity that requires a cognitively sophisticated capacity that is uniquely human and only emerges late in ontogeny and phylogeny. As a result, mental fictionalists appear forced to posit a stark dichotomy between *mere* behavior-reading and full-blown mindreading as it is evidenced in a grasp of fictionality. In contrast, the most plausible accounts of mindreading in the

psychological literature provide an empirical story about how the ability to track or attribute mental states to other creatures *gradually* develops both evolutionarily and developmentally. I suggest that fictionalists ought to pursue this same gradualist strategy.

Mindreading is not an all-or-nothing capacity. There are accounts of mindreading that posit a middle-ground ability between mere behavior-reading and full-blown mindreading that the mental fictionalist could appeal to and build upon. For example, Butterfill and Apperly propose a dual-systems view of mindreading, which posits the existence of a lower-level, less-sophisticated cognitive mechanism to rapidly and automatically track mental states such as beliefs. On this account, non-human animals and pre-verbal infants are able to track the mental states of others by employing the representation of "registration" as holding between creatures and objects in the environment. To "register" an object's presence means that once a creature first detects it at a particular location, it continues to be disposed to interact with it appropriately, even after losing direct awareness of it (such as by looking away) (Butterfill & Apperly, 2013, p. 617). This goes beyond behavior-reading in requiring the creature to represent an enduring relation between the agent and the object, rather than merely responding to observable cues. However, it falls short of attributing fully fledged propositional attitudes, such as beliefs or desires, to the agent.[12]

I propose that mental fictionalists build upon these middle-ground accounts of "proto-mindreading," using them to devise an evolutionary and developmental story that explains how rudimentary mental state tracking could mature into participation in shared fictions – and, thus, into full-fledged folk-psychological practice. What is essential is that alleged instances of mindreading in pre-linguistic children and non-human primates be characterized as something *more* than mere behavior-reading and yet less demanding than participation in fictionality. The capacity to "proto-mindread" might then be enhanced by further cognitive abilities through phylogeny and ontogeny to enable the sort of full-fledged mindreading focused upon by mental fictionalists. The task is formidable, yet its completion would afford mental fictionalism a central place in both philosophy and cognitive science.

Notes

1. For some recent accounts of how fictionalism stands in relation to these other middle-ground views, see (Dennett, 2022; Mölder, 2022).
2. For brevity's sake, I focus on these two forms of fictionalism; however, I believe that my twin challenges apply to all forms of mental fictionalism. For other versions my challenges apply to, see (Lycan, 2022; Wallace, 2022).

3. In the literature, this has come to be termed *factive* mindreading (e.g., Westra & Nagel, 2021). This contrasts with *non-factive* mindreading, which consists in a sensitivity to agents' true or false beliefs. This capacity is seen as significantly more cognitively demanding, as it requires an appreciation of the difference between appearance and reality. In other words, rather than simply attributing to another agent something that the mindreader themselves know, they must be capable of attributing to them something that *might not be* true. While there have been studies providing convincing evidence of non-factive mindreading (specifically false-belief attribution) in great apes and monkeys, the literature is significantly sparser. Hence, I choose to focus on well-replicated and uncontroversial instances of factive mindreading. See: (Hayashi et al., 2020; Krupenye et al., 2016). In a similar vein, there has been plenty of experiments allegedly showing that infants are sensitive to false-beliefs as young as nine-months-old (e.g., Kovács et al., 2010). However, given how controversial these experiments have become due to methodological issues and failures of replication, I will not consider them. If it turns out that certain experiments are well-replicated, this will only strengthen my argument against fictionalism. For an overview, see, e.g., (Rakoczy, 2022).

4. For similar competitive experiments providing strong evidence for the presence of folk-psychological practice in non-human primates, see the experiments and reviews in, e.g (Bräuer et al., 2007, 2010; Melis et al., 2006). For equivalent findings of factive mindreading in pre-linguistic infants see (e.g., Caron et al., 2002; Liszkowski et al., 2006; Tomasello & Haberl, 2003).

5. For discussion of equivalent findings in infants, see: (Phillips et al., 2021).

6. For similar findings in great apes, see, e.g., (Call et al., 2004).

7. To date, there is no evidence that non-human primates are sensitive to the subjective desires of other agents. However, there is a plethora of literature on their sensitivity to the stable goals of other creatures and expect them to pursue them efficiently. For example, see (Kano & Call, 2014; Uller, 2004; Yamamoto et al., 2012).

8. Many have raised philosophical objections against the behavior-reading programme on the grounds that it runs afoul of good science in lacking scientific virtues like simplicity, unification, fruitfulness, and so on see, e.g., (Buckner, 2014; Fitzpatrick, 2009; Sober, 2015, 207–43).

9. It is worth briefly mentioning a related anticipation of the pretense-based view from the cognitive science literature. On Alan Leslie's early account of mindreading, the capacity for pretense was the first example in ontogeny of the instantiation of a native mindreading module in humans (Leslie 1987, 1994). Subsequent work in development psychology (see the footnotes in the previous section) has led many to argue that other (simpler) forms of mindreading emerge prior to the capacity for make-believe. My overall argument in the previous section engages with this historical shift: if fictionalism ties folk-psychological practice to pretense or shared narrative participation, it must still accommodate the infant and comparative data that indicates the capacity to mindread is prior to the ability to make-believe. In this section, I argue that even *if* we consider pretense as the earliest form of mindreading, that the structure of early make-believe cannot sustain the kind of folk-psychological practice required on both the pretense- and affective fictionalist views. I thank an anonymous reviewer for suggesting I include this historical context.

10. I will tend to refer to this as "pretense" following the literature; however, it is clear that these same points apply equally well to the affective-fictionalist's requirement of participation in collective narratives, which requires an ability to maintain an (even minimal) understanding of the distinction between appearance and reality.

11. It is worth noting that apes too are capable of this simple model mimicry. See: (Call and Carpenter, 2003).
12. Other middle-ground views include the awareness-relations account (e.g., Horschler et al., 2021) and the script-theory of mindreading (e.g., Taylor et al., 2023).

ORCID

Gavin Robert Foster 🆔 http://orcid.org/0009-0001-4450-5639

References

Andrews, K. (2012). *Do apes read minds? Towards a new folk psychology*. The MIT Press.

Andrews, K. (2016). Snipping or editing? Parsimony in the chimpanzee mindreading debate. *Metascience*, *25*(3), 377–386. https://doi.org/10.1007/s11016-016-0119-0

Bourchier, A., & Davis, A. (2002). Children's understanding of the pretence-reality distinction: A review of current theory and evidence. *Developmental Science*, *5*(4), 397–413. https://doi.org/10.1111/1467-7687.00236_1

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science*, *20*(1), 85–91. https://doi.org/10.1111/j.1467-9280.2008.02246.x

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. https://doi.org/10.1007/s10071-007-0088-1

Buckner, C. (2014). The semantic problem(s) with research on animal mind-reading. *Mind & Language*, *29*(5), 566–589. https://doi.org/10.1111/mila.12066

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Call, J., & Carpenter, M. (2003). On imitation in apes and children. *Journal for the Study of Education & Development*, *26*(3), 325–349. https://doi.org/10.1174/021037003322299070

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). 'Unwilling' versus 'unable': Chimpanzees understanding of human intentional action. *Developmental Science*, *7*(4), 488–498. https://doi.org/10.1111/j.1467-7687.2004.00368.x

Caron, A. J., Kiel, E. J., Dayton, M., & Butler, S. C. (2002). Comprehension of the referential intent of looking and pointing between 12 and 15 months. *Journal of Cognition and Development*, *3*(4), 445–464. https://doi.org/10.1080/15248372.2002.9669677

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315–330. https://doi.org/10.1016/S0163-6383(98)90009-1

Churchland, P. M. (1980). Eliminativism and the propositional attitudes. *The Journal of Philosophy*, *78*(2), 67–90. https://doi.org/10.5840/jphil198178268

Demeter, T. (2022). A Mental Fictionalism Worthy of Its Name. In T. Demeter, T. Parent & A. Toon (Eds.), *Mental fictionalism: Philosophical explorations*, 140–167. Routledge.

Demeter, T., & Sugden, S. J. B. (2013). Mental fictionalism: The very idea. *The Monist*, *96*(4), 483–504. https://doi.org/10.5840/monist201396422

Dennett, D. C. (2022). Am I a fictionalist? In T. Demeter, T. Parent & A. Toon (Eds.), *Mental fictionalism: Philosophical explorations* (pp. 539–554). Routledge.

Drayton, L. A., & Santos, L. R. (2018). What do monkeys know about others' knowledge? *Cognition* 170: 201–208.

Fitzpatrick, S. (2009). The primate mindreading controversy: A case study in simplicity and methodology in animal psychology. In R. W. Lurz (Ed.), *The philosophy of animal minds* (pp. 258–277). Cambridge University Press.

Fletcher, L., & Carruthers, P. (2013). Behavior-reading versus mentalizing in animals. In J. Metcalfe & H. S. Terrace (Eds.), *Agency and joint attention* (pp. 82–99). Oxford University Press.

Fodor, J. (1980). *The language of thought*. Harvard University Press.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785. https://doi.org/10.1006/anbe.1999.1377

Harris, P. L., Brown, E., Marriott, C., Whittall, S., & Harmer, S. (1991). Monsters, ghosts, and witches: Testing the limits of the fantasy-reality distinction in young children. *British Journal of Developmental Psychology*, *9*(1), 105–123. https://doi.org/10.1111/j.2044-835X.1991.tb00865.x

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, *3*(4), 379–400. https://doi.org/10.1080/02699938908412713

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., Iijima, A., Someya, T., & Hasegawa, I. (2020). Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Reports*, *30*(13), 4433–4444.e5. https://doi.org/10.1016/j.celrep.2020.03.013

Horschler, D. J., Santos, L. R., & MacLean, E. L. (2019). Do non-human primates really represent others' ignorance? A test of the awareness relations hypothesis. *Cognition* 190: 72–80.

Horschler, D. J., Santos, L. R., & MacLean, E. L. (2021). How do non-human primates represent others' awareness of where objects are hidden? *Cognition*, *212*, 104658. https://doi.org/10.1016/j.cognition.2021.104658

Jackowitz, E. R., & Watson, M. W. (1980). Development of object transformations in early pretend play. *Developmental Psychology*, *16*(6), 543–549. https://doi.org/10.1037/0012-1649.16.6.543

Kano, F., & Call, J. (2014). Great apes generate goal-based action predictions: An eye-tracking study. *Psychological Science*, *25*(9), 1691–1698. https://doi.org/10.1177/0956797614536402

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Kroon, F., Brock, S., & McKeown-Green, J. (2019). *Critical introduction to fictionalism*. Bloomsbury Academic.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false-belief. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Leslie, A. (1987). Pretence and representation: The origins of 'theory of mind'. *Psychological Review*, *94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Leslie, A. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*(1–3), 211–238. https://doi.org/10.1016/0010-0277(94)90029-9

Lewis, L. S., & Krupenye, C. (2022). Theory of mind in nonhuman primates. In B. L. Schwartz & M. J. Beran (Eds.), *Primate cognitive studies* (pp. 439–482). Cambridge University Press.

Lillard, A. S. (2015). The development of play. In S. Lynn (Ed.), *Handbook of child psychology and developmental science* (7th ed., Vol. 2, pp. 425–468). Liben & Ulrich Müller. Wiley.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2

Lycan, W. G. (2022). A Rylean Mental Fictionalism. In T. Demeter, T. Parent & A. Toon (Eds.), *Mental fictionalism: Philosophical explorations*, 115–139. Routledge.

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science*, *14*(6), 1406–1416. https://doi.org/10.1111/j.1467-7687.2011.01085.x

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition*, *130*(3), 300–308. https://doi.org/10.1016/j.cognition.2013.11.016

Matsuzawa, T. (2020). Pretence in chimpanzees. *Primates*, *61*(4), 543–555. https://doi.org/10.1007/s10329-020-00836-z

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*(2), 154–162. https://doi.org/10.1037/0735-7036.120.2.154

Mitchell, R. W. (2007). Pretence in animals: The continuing relevance of children's pretence. In A. Göncü & S. Gaskins (Eds.), *Play and development: Evolutionary, sociocultural, and functional perspectives* (pp. 51–75). Lawrence Erlbaum Associates Publishers.

Mölder, B. (2022). Interpretivism and Mental Fictionalism. In T. Demeter, T. Parent & A. Toon (Eds.), *Mental fictionalism: Philosophical explorations* (pp. 489–519). Routledge.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences*, *44*, e140. https://doi.org/10.1017/S0140525X20000618

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, *1*(4), 223–235. https://doi.org/10.1038/s44159-022-00037-z

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21. https://doi.org/10.1037/0012-1649.33.1.12

Schlicht, T. (2023). *Philosophy of social cognition*. Palgrave Macmillan.

Sober, E. (2015). *Ockham's razors: A user's manual*. Cambridge University Press.

Spaulding, S. (2018). *How we understand others: Philosophy and social cognition*. Routledge.

Stanley, J. (2001). Hermeneutic fictionalism. In P. French & H. Wettstein (Eds.), *Midwest studies in philosophy, 25: Figurative language* (pp. 36–71). University of Minnesota Press.

Taylor, D., Gönül, G., Alexander, C., Züberbühler, K., Clément, F., & Glock, H.-J. (2023). Reading minds or reading scripts? De-intellectualising theory of mind. *Biological Reviews*, *98*(6), 2028–2048. https://doi.org/10.1111/brv.12994

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, *39*(5), 906–912. https://doi.org/10.1037/0012-1649.39.5.906

Toon, A. (2022). Fictionalism and Intentionality. In T. Demeter, T. Parent & A. Toon (Eds.), *Mental fictionalism: Philosophical explorations* (pp. 87–113). Routledge.

Toon, A. (2023). *Mind as metaphor: A defence of mental fictionalism*. Oxford University Press.

Uller, C. (2004). Disposition to recognize goals in infant chimpanzees. *Animal Cognition*, *7*(3), 154–161. https://doi.org/10.1007/s10071-003-0204-9

Wallace, M. (2022). Mental Fictionalism. In T. Demeter, T. Parent & A. Toon (Eds.), *Mental fictionalism: Philosophical explorations* (pp. 50–86). Routledge.

Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of representational arts*. Harvard University Press.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Westra, E., & Nagel, J. (2021). Mindreading in conversation. *Cognition*, *210*, 104618. https://doi.org/10.1016/j.cognition.2021.104618

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Yamamoto, S., Humle, T., & Tanaka, M. (2012). Chimpanzee's flexible targeted helping based on an understanding of conspecifics' goals. *Proceedings of the National Academy of Sciences*, *109*(9), 3588–3592. https://doi.org/10.1073/pnas.1108517109