The Sellarsian Fate of Mental Fictionalism

László Kocsis and Krisztián Pete

forthcoming in Tamas Demeter, Ted Parent, Adam Toon (eds.)
Mental Fictionalism: Philosophical Explorations (Routledge, 2022)

Abstract

This chapter argues that mental fictionalism can only be a successful account of our ordinary folk-psychological practices if it can in some way preserve its original function, namely its explanatory aspect. A too strong commitment to the explanatory role moves fictionalism unacceptably close to the realist or eliminativist interpretation of folk psychology. To avoid this, fictionalists must degrade or dispense with this explanatory role. This motivation behind the fictionalist movement seems to be rather similar to that of Sellars when he came up with the Myth of Jones, his proto-theory of mental concepts. He was faced with the problem of preserving the explanatory status of mental concepts without turning them into proper theoretical entities. By introducing the Sellarsian proto-theory of concepts related to the mental and outlining its main points, this chapter aims to provide a critique of the two versions of mental fictionalism that are arguably the strongest: Adam Toon's prop-oriented pretence theory and Tamás Demeter's expressive storyism.

0. Introduction

All of us are trained to ascribe beliefs, desires and other inner states to each other and to ourselves to explain and predict their and our own actions and behaviours – and yet, we are left in the dark about the real nature of this folk-psychological belief/desire discourse. What are we doing exactly when we talk about mental states? Do we express truth-apt propositions and state mental facts and then appeal to these facts when we explain and predict actions and behaviours? Or do we just pretend to do all of these things, and in fact are doing something else? Can our familiar folk-psychological discourse be a theory about mental states? If not, what is the theoretical role of folk psychology, if it has any at all?

There are two radically opposite views about our commonly used "mental talk", yet both of them maintain that folk psychology has a robust role to play; in a more or less well-defined sense, it is a *theory* about what is going on inside our heads, that is, how our beliefs, desires and other inner states cause how we act or behave.

On the one hand, according to the folk-psychological realists, our mental talk includes many true descriptions of our inner states, probably more so than false ones. Why? Because our folk-psychological explanations and predictions seem to be very successful, which would not be the case if they were not true descriptions of our mental states and their relations to other mental states or to our behaviours. Even if this realist account of folk-psychological discourse stops short of being a substantive theory of the nature of mental states, it maintains that some form of ontology needs to be matched to our presumably true descriptions of our inner states and their causal relevance.

On the other hand, eliminative materialists have a very specific anti-realist attitude towards folk-psychological discourse, maintaining that it is wrong as a theory of mind. For eliminativists, "folk" means faulty, that is, our ordinary mental talk is seriously defective as a

theory of mental states.[1] Accordingly, appealing to mental states while trying to explain why someone behaves as she does is as erroneous as an appeal to phlogiston when we try to explain why something burns.

Contrary to the above-mentioned realist and anti-realist positions, mental fictionalism is an original philosophical view about the real nature of folk-psychological discourse, arguing that it is not a theory of our mental states but only a useful fiction. As such, mental fictionalism appears to be a sort of conservative alternative to mental-talk realism and eliminativism: a fictionalist retains our folk-psychological talk and tries to preserve its utility without admitting the demand for combining it with any robust ontology. In this essay we want to show that the most promising strategy to attain this mental fictionalist goal would be to reach back to the Sellarsian characterization of our everyday mental talk.

1. What is mental fictionalism?

1.1 The common denominator

Mental fictionalism is an umbrella term: though it has different versions, they all have something in common. Our aim is to discuss three more-or-less interconnected characteristics which help us to find the common ground among the various fictionalist views about mental states. Basically, we focus on how mental fictionalism differs from eliminativism.

Firstly, mental fictionalists deny the possibility of a full-blooded eliminativism about our folk-psychological discourse. Needless to say, if we regard folk psychology as a theory that uses folk-psychological terms to refer to mental states as existing theoretical posits – and if we agree with eliminativists that mental states do not exist at all – then folk psychology can be nothing but a mistaken theory. However, getting rid of the whole discourse seems to be too draconian a step, since forms of talk which are apparently problematic from an ontological point of view can have benefits from other points of view. For this reason, Richard Joyce (2013) distinguishes two kinds of eliminativism, ontological and linguistic, which do not necessarily go hand in hand. The main aim of the mental fictionalist is to grant that folk-psychological discourse does not commit us to the existence of mental states. As Joyce points out, linguistic eliminativism, which seeks to jettison folk psychology altogether, differs from mental fictionalism, in the sense that "the fictionalist holds that all those utterances that one would ordinarily think of as committing the speaker to psychological entities in fact do not (or need not) do so, and thus there is no pressure for their abolition" (Joyce 2013, p. 518). How can our mental talk be ontologically innocent? Even if the discourse seems to be assertive and literally true, when we talk about mental states, we indeed simply tell a story or pretend to make assertions; folk psychology is a fictional or figurative discourse including theoretically innocuous metaphors, and not a fact-stating discourse including ontologically committing descriptions.

[1] Eliminativism is an error theory about our whole folk-psychological discourse. According to eliminativism, the question is how seriously defective our mental talk is. It is worth mentioning here that David Lewis (2005) makes a distinction between two kinds of error-ridden theories. On the one hand, an erroneous theory can be essentially wrong, meaning that its errors are "inextricably involved in the working of the theory" (317), so that it wouldn't be a theory at all without these errors; see, for example, phlogiston theory about combustion or witchcraft theory about harmful events. On the other hand, a theory can be erroneous because it contains some part which needs to be and can be corrected (provided that correction is not tantamount to abandonment). Eliminativism treats folk psychology as an error-ridden theory of the first kind and advocates that it would be best to get rid of it. As we will briefly discuss below, mental fictionalism does not consider folk psychology to be an error-ridden theory of mind because it is not a theory of mind at all.

Secondly, rather than taking a strict position about the (non-)existence of mental states, mental fictionalists are agnostic about the ontology of mind.[2] As we have mentioned earlier, all fictionalist strategies aim to preserve particular regions of discourse without carrying their heavy ontological burden. At the same time, at least recently, mental fictionalists have strongly stressed their explicit detachment not just from linguistic eliminativism – which seems to be evident due to their positive attitude towards the practical benefits of ordinary mental talk – but also from ontological eliminativism, in the sense that they try to treat folk psychology as an ontologically innocent discourse. They explicitly seek to avoid any unnecessary and controversial ontological commitment. For example, Adam Toon (2016) argues that contrary to what eliminativists proclaim about the non-existence of mental states, one of the great advantages of being a fictionalist is that it does not oblige us to take a position in the ontological debate about mental states:

> [T]he fictionalist need not follow the eliminativist in denying the existence of mental states; instead, she might remain agnostic on the matter. The distinctive feature of fictionalism is that it allows us to grant that, even if mental states do not exist, we can nevertheless continue talking as if they do. (Toon 2016, p. 290)

It should be noted here that according to mental fictionalism, our literally untrue ordinary mental talk can never be used as a (proto-)scientific theory of psychological states, for the simple reason that terms such as 'folk', 'ordinary' or 'commonsensical' distinguish it from any scientific discourse, meaning it does not matter at all what the future will bring in cognitive science (see Toon 2016, p. 280). As Tamás Demeter (2022) contends,

> fictionalism is compatible with certain forms of realism about mental entities, provided that they are not committed to the constitutive role of folk-psychological concepts in the analysis of the 'mental'. It does not touch upon the question whether there are mental states if they are understood independently of the conceptual resources of folk psychology. (Demeter 2022, p. xy)

Thus, mental fictionalists tend to set aside ontological questions; they basically deal with the problem of how we organise and use our everyday psychological language and what it is good for.

Thirdly, though mental fictionalists deny that folk-psychological discourse can be literally true, they approve of its utility. While folk psychology is undoubtedly a systematically organised normative discourse, it does not, so the mental fictionalist argues, follow from this systematicity that it is also a theory about what is going on inside our heads. Even if folk psychology is a mistaken (proto- or quasi-) scientific theory, as a peculiar storytelling practice it can open the door to many benefits, at least according to mental fictionalists. At the same time, there is no consensus among fictionalists as to what roles folk psychology plays. It is well-known that according to the standard view, folk psychology is uniquely suitable for understanding, explaining and predicting behaviours. But what is folk psychology good for if we do not treat it as a fact-stating discourse, including about true representations of our inner

[2] It should be noted here that not everyone characterizes mental fictionalism as an agnostic. For example, Ted Parent (2017, Sect. 9.2) defines mental fictionalism by two theses: (1) eliminativism about mental states, asserting that mental states as referred to and described in folk psychology do not exist, and (2) the so called "story prefix semantics", according to which mental fictionalists take folk psychology to be a particular kind of fictional discourse. We characterize mental fictionalism differently claiming that even if eliminativism may be the first step towards becoming a fictionalist about folk psychology, mental fictionalists need not accept the thesis of eliminativism (nor the other one). In this paper, we discuss mental fictionalism as an alternative to eliminativism, not as a supplement to it, and focus on those mental fictionalist views that are explicitly agnostic.

states? Can we understand, explain or predict anything by appealing to mental states if they only exist fictionally? Can we ascribe any epistemic virtues to folk-psychological discourse if it is no more than a fictive narrative? There is a big controversy among mental fictionalists about the tasks of folk psychology if it is treated as a special kind of storytelling practice.

1.2 Varieties of mental fictionalism

First of all, surprising as it may be, it is not uncontroversial that mental fictionalists should treat folk psychology as a sort of fictional discourse. And, indeed, is there really any fiction here? Many authors think that there is. For example, Matti Eklund (2007/2019) generally emphasises that "[f]ictionalism about a region of discourse can provisionally be characterized as the view that claims made within that discourse are not best seen as aiming at literal truth but are better regarded as a sort of 'fiction'". Similarly, Craig and Emily Caddick Bourne (2020) also contend that "[w]hat characterizes a fictionalist approach to subject matter *X* is the suggestion that *X* can be understood by appeal to the notion of fiction. Otherwise, fictionalism does not deserve its name" (Bourne and Caddick Bourne 2020, p. 168). If Eklund and Bourne and Caddick Bourne are right, then a fictionalist approach to folk-psychological discourse cannot be worked out unless it yields a discourse which has the definite characteristics of fiction. Accordingly, most advocates and sympathisers of mental fictionalism presuppose that folk psychology is a fictional discourse:

> In the case of psychological fictionalism, the fiction in question might be called "folk psychology." This is the theory that eliminative materialists think is false. But even if false, the theory presumably has enough content to ground "According to folk psychology…" claims. (Joyce 2013, pp. 521-522)

> [I]f the mental fictionalist asserts, 'There is a belief state …', this should not be understood in an ontologically loaded way. Rather, it should be translated as the fiction relative claim that *according to the mental fiction*, there is a belief state, etc., etc. (Parent 2017, p. 229)

> [I]n the fiction *f*, Joe believes that there is beer in the fridge] "in the fiction *f*" will pick out (…) the *mental fiction*. And the mental fiction, in this case, will be *folk psychology*. (Wallace 2022, p. xy)

Using the "according to the folk psychology/mental fiction" or "in the mental fiction" prefixes is a handy solution, as they transform our ontologically committing talk into an ontologically innocent one. However, fictionalists are divided over whether we really have to introduce or identify a particular fictional discourse, though this would seem to be indispensable if we wanted to use a troubleshooting prefix along the lines of "according to the fiction *f*". Some of them do not consider it crucial that the particular discourse for which they elaborate a fictionalist view should have a very strong connection with a fictional discourse.[3]

Toon (2016) and Demeter (2022) argue that our everyday and often loose mental talk cannot be regarded as a fictional discourse, given that we are unable to present any standard or relevant fiction for folk psychology that could be identified as a particular, more-or-less unified and coherent discourse in which we could distinguish the fictionally true statements from false ones. According to Toon, "it is clear which fiction underpins our talk about Sherlock Holmes: we can point to our copy of Conan Doyle's stories. But there is no text that sets out the principles

[3] For example, the mathematical fictionalist Mark Balaguer (2008/2018, Sect. 2.4) is convinced "that despite the name, fictionalist views do not have to involve any very strong claims about the analogy between mathematics and fiction."

of folk psychology" (2016, pp. 289-290). In a similar manner, Demeter (2022, Sect. 3) emphasises the disanalogy between our everyday psychological discourse and other cases of (literary) fictions. In the case of folk psychology, there appears to be a lack of some literary corpus or theory, which would be indispensable if we wanted to defend a so-called prefix-semantical fictionalism about folk psychology.[4]

A further important difference among mental fictionalists relates to how the practice of our actual folk-psychological discourse should be understood. We are here focusing on two versions of mental fictionalism: Toon's pretence fictionalism and Demeter's affective storyism. As *hermeneutic* fictionalists, neither of them wants to alter the actual practice of ordinary psychological discourse. According to Toon (2016; 2022), folk psychology is not a proto-scientific theory but a metaphorical discourse containing literally false but (at least in part) fictionally true statements. Thus, when we talk about mental entities, we do not seriously assert that someone believes/desires something; we just play a game pretending to make an assertion about someone else's beliefs and desires, meaning we use the words 'believe' and 'desire' only in a metaphorical sense. Demeter (2022) is of the same opinion when he argues that his special affective storyist view provides "an interpretation of the discourse without proposing a revision of its practice – it revises only the manifest image of the discourse" (p. xy).[5]

Whilst both Toon and Demeter are committed to not altering our linguistic practices, they disagree with each other about what we really do when we talk about our mental states. Contrary to Toon's pretence fictionalism, according to Demeter's affective storyism, mental fictionalists should not defend a pretence theory but explicitly state that folk-psychological practice is not a game in which we just pretend that we talk about real mental states:

> Taking part in folk-psychological discourse is not pretending in this sense. We take folk psychology seriously as if its propositions expressed facts, by ascribing mental properties to agents. We take these to be true or false depending on how things stand with the agent, and independently of our discourse about them. We have a firm picture of what we are doing while ascribing these mental states, but the affective storyist would argue that this picture is mistaken: we are doing something else than what we say based on the manifest view of our folk-psychological practices. (Demeter 2022, p. xy)

And last but not least, there is a disagreement among mental fictionalists whether folk psychology is a non-literal discourse about fictional mental states that can be used for epistemic purposes, and, if so, to what extent. While all mental fictionalists are convinced that our mental talk is indispensable practically, some of them, like Toon, maintain that despite its literal falsity, it also has some epistemic virtues – for example, we can give (causal) explanations by using only fictionally or metaphorically true statements. Other mental fictionalists, like Demeter, explicitly deny that folk-psychological discourse has any epistemically valuable aspects, meaning it does not play any explanatory role, especially if we believe that mental states can be explained causally. In Section 3, we will discuss in detail the problem of how a fictionalist could and/or should ascribe an explanatory role to folk-psychological discourse. But before that, we present the main lines of the Sellarsian idea, paying special attention to those features

[4] See Wallace (2022, Sect. 3) for an attempt to solve the lack-of-fiction problem. Similarly to Joyce (2013), she argues that prefix-semantical fictionalism is maintainable, provided that it acknowledges what both realists and eliminativists emphasise: that folk psychology, with all its implicit definitions of psychological terms, causal explanations of behaviours, etc., is a kind of (proto-scientific) theory. It does not matter whether this theory is a good or a bad one – according to Wallace, it should be treated as a fiction.

[5] In arguing that revolutionary fictionalism about folk psychology could be a slippery slope towards eliminativism, Joyce (2013) has the following to say about hermeneutic mental fictionalism: "it is not an error theory, inasmuch as it rejects that psychological language was ever really in the business of describing the mind in the first place, and thus could hardly be erroneously misdescribing it" (p. 520).

and demands that relate it to mental fictionalism, which may help us to understand and evaluate different versions of mental fictionalism.

2. The Sellarsian prelude to mental fictionalism

With the Myth of Jones, Sellars tries to establish an alternative conception of the mental that occupies the middle ground between Cartesian realism and Behaviourism (which was widely considered the other extreme at the time of publication of the Empiricism and the Philosophy of Mind), as he thinks that they both rest on the Myth of the Given. According to the "classical" conception, there really are inner episodes and states such as thoughts, perceptions, beliefs and desires, to which we have direct, privileged access, so that they can provide epistemic support for all our knowledge without being themselves epistemically dependent on anything else. The other extreme, Verbal Behaviourism, denies the existence of such mental states and holds them to be reducible to publicly available behavioural episodes and dispositions. Our knowledge about them is always inferential and based on observable behavioural evidence. But on the other hand, Sellars also seeks to retain the fundamental insights of both conceptions of the mental, since he considers both to be basically correct. He thus aims to preserve the behaviourist insight that the meaning of mental concepts is determined according to empirical criteria, as well as the classical intuition that mental events are real inner episodes and not reducible to complex dispositions or behaviours without residue. The essence of his solution is to conceive of our everyday psychological concepts in terms of a pattern of theoretical concepts.

2.1 Folk psychology as a theoretical explanation

Sellars' Myth of Jones is designed to reveal something of the nature and status of folk-psychological concepts, and to answer the question how we can have privileged access to mental episodes without having to accept them as real. Can we get from a description of the publicly accessible experiential world to mental episodes without having to introduce new kinds of entities? Sellars claims that we can. All that is required to do so is that our ancestors are able to describe their commonly – and directly – accessible intersubjective contexts and possess the logical and semantic concepts which are applicable to their own public behaviours. What must be added to this is the ability to theorise, that is, the ability to transpose familiar concepts as models into an unknown domain, by constructing a theoretical explanation for observed phenomena not yet understood. Accordingly, Jones comes up with a theory to explain certain kinds of behaviour that would otherwise seem to be inexplicable in the Rylean framework, in which thinking is nothing more than "thinking-out-loud". However, it is reasonable to assume that the members of his community behave rationally even when their behaviour is not accompanied by overt verbal utterances. Rylean Verbal Behaviourism can cope with this silent thinking by introducing dispositions to thinking-out-loud, but this correlational model of explanation is more like a heuristic device than a proper explanation. Moreover, the explanatory inadequacy of a purely dispositional analysis in silent cases of behaviour is undeniable. This is where the genius Jones comes in, who "develops a *theory* according to which overt utterances are but the culmination of a process which begins with certain inner episodes" (Sellars 1997, p. 103). And the model for this explanatory theory is none other than overt verbal behaviour itself, of which Jones hypothesises that it occurs internally, maintaining that "the true cause of intelligent nonhabitual behavior is 'inner speech'" (Sellars 1997, p. 103). The theory does not state that a mental episode (a thought in particular) is a form of inner speech. Such so-called "inner speech" provides only a model for the theory of inner thoughts and enlightens some important aspects of the newly postulated mental episodes. It states only that an inner episode works as if the person being interpreted were covertly speaking to himself. Here is an example: when Smith looks at the side of the road and does not step off it, Jones' theory assumes that something is going on inside him, and that the semantic dimensions of that something are

similar to, and would naturally be expressed by, "That car is too close, I wouldn't pass in front of it". According to Jones' theory, the reason for his behaviour, and at the same time its cause, is an occurrent inner episode, a thought which has the same meaning and propositional content as what Smith would say to himself if he were in the mood of "thinking-out-loud". The Jonesian theory explains the appropriateness and intelligence of conduct in contexts where no thinking-out-loud occurs.

The key to this mythical theory of thought is that the episodes (hypothetically) postulated by Jones as covert (mediating) states of persons are introduced by a functional analogy to overt verbal behaviour. Consequently, its properties are also modelled on the semantic properties of speech acts. A thought is a "logico-semantic role player" (to use Rosenberg's (2004) terminology), and its ontological instantiation is left open. The Sellarsian theory explains Smith's behaviour in the following way: something with the propositional content ·That car is too close, I wouldn't pass in front of it·[6] occurs in the "logical space" of Smith as he stands at the pavement turning his head right to left and back. The theory also hypothesises that this inner episode, in conjunction with the covert presence in Smith of the desire ·Would that I get home safely!·, yields a volition that, in the right context, eventually causes a silent but rational staying-put behaviour. This is how Jonesian theory explains silent behaviour, which is basically the same way that folk psychology explains rational behaviour.

Since mental episodes are postulated through an overtly accessible model, their properties are functionally analogous to the model – they are functional kinds that leave open the question of their ontological instantiation. And this is precisely how the fictionalist treatment (not in the eliminativist sense) characterises the mental, granting some significance to the folk-psychological discourse without slipping into realism.

2.2 Why Sellarsian folk psychology is not a proper theoretical explanation

Taken along the lines of the traditional (eliminativist) interpretation, the Jonesian theory substantiates what is implicit in the above paragraphs, i.e. that folk psychology is itself a theory for explaining behaviour. We utilise folk-psychological concepts in theoretical reasoning, just as we do in explaining natural phenomena, by appealing to unobservable subatomic particles, and we accept the existence of the building blocks of the postulational theory solely because of the theory's explanatory power. This is, in fact, the theory-theory, whose only attraction lies in its explanatory potential.

Sellars seems to accept the primacy of scientific theorisation in ontological matters: "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (Sellars 1997, p. 83). Yet, Sellars argues, our folk-psychological concepts are not primarily in the business of describing the world; rather, they are indispensable conditions under which we can act in the world as agents. Folk-psychological concepts are ineliminable components of our everyday pragmatic activities. The Jonesian theory does not give us a fully-fledged scientific theory, as it contains no lawlike psychological generalisations. And for this reason, it clearly falls short of being the kind of theory that the eliminativists want to see in it. However, it is clearly more ambitious than "Verbal Behaviourism", which is based on purely correlational explanations and therefore lies within the manifest image (Sellars 1962). With Sellars, we can call Jones' theory a *proto-theory*. And as Sellars introduces this proto-theory in the Myth of Jones, it offers an explanatory enrichment of the Rylean account, in which thoughts *are* sayings or dispositions to say things of persons; hence this proto-theory takes place within the manifest image and not in the scientific one:

[6] Sellars' dot quotations are meant to classify conceptual role players across languages.

> Perhaps the most important point is that what the theory postulates in the way of new entities are processes and acts rather than *individuals*. In this sense, it remains within the manifest image. Persons remain the basic individuals of the system (Sellars 1975, p. 329)

The processes Sellars speaks about in this passage are the states of the perceivers (1997, p. 110) – not new entities proper (in the scientific sense) but only adverbial modifications of an already available and accepted entity of the manifest image. We will shortly come back to this point.

In what sense then are inner episodes theoretical, and in what sense are they not? One thing is certain: Sellars does not consider the distinction between theoretical and observational entities to be absolute. In fact, the difference between the two is methodological and not ontological in nature. We can only have inferential knowledge of theoretical objects, whereas we can have non-inferential knowledge of observational ones. In Jones' theory of thoughts, we infer internal episodes, but the privileged access that Sellars strives to preserve presupposes that we can give direct reports of these inner episodes, at least in our own case. For Sellars, there is an ambiguity in the notion of observability, which also has consequences for Jones' proto-theory. This ambiguity can be elucidated by borrowing O'Shea's (2012, pp. 193-195) distinction between two kinds of perception. According to O'Shea, Sellars uses the term perception in two senses, a looser one in which the physicist sees/perceives the mu meson (or, as we shall see, perceives the internal happenings of herself and others), and a stricter one in which we "directly" perceive only the kinds and properties of "manifest" objects. Within the latter, we can also distinguish the occurrent sensible properties, what we perceive of the physical objects from the kinds (classificatory concepts), i.e. what we perceive them as. In other words, the conceptual framework of directly perceptible objects, the manifest image, has a proper perceptual component (perception *of*) that remains constant and a conceptual component (perception *as*) that may change naturally, for example, due to the intrusion of the conceptual framework of the scientific image. However, this change is limited, since directly observable kinds must be "operationally definable" in terms of sensible properties. As such, the scientific image cannot fully replace the manifest image, since the mu meson cannot be defined in this way, for example. So far so good, but then what can we say about mental states?

Thoughts, as postulated by Jones, are theoretical in the aforementioned "strict" sense, as they are neither perceptual, sensible properties nor "dispositional or causal kinds" that can be defined by such sensible properties. Inner episodes are imperceptible episodes that are theoretically postulated to take place "in" persons. However, as we have noted earlier, they do not constitute a new domain of reality – they do not enlarge our basic ontology. As Sellars says, "They are primarily 'in' the person as *states* of the person" (Sellars 1975, p. 329). That is, Sellars places mental concepts into the manifest image as ways of being a person, yet he does not think of them as identical to any propensities to overt linguistic behaviour. Jones postulates these occurrent, non-dispositionally defined imperceptible inner episodes in order to better explain the very propensities to think-out-loud with which behaviourist explanations operate. Both explanatory frameworks use the same behavioural evidence in their explanations, which is why Sellars says that "in the case of thoughts, the fact that overt behavior is evidence for these episodes is built into the very logic of these concepts as the fact that the observable behavior of gases is evidence for molecular episodes is built into the very logic of molecule talk" (1997, pp. 115-6). In our everyday "folk" understanding, these episodes would indeed have the status of theoretical entities, but only in the "strict" sense. They are not properly perceivable or definable as dispositional properties; they are imperceptible states of manifest, perceptible persons. And yet at the same time, they can be non-inferentially reported to occur in the "looser" sense. They are 'perceived *as*' occurring in us or in others, but certainly not 'perceived *of*' ourselves or other persons.

Sellars does not say much about the process through which "[w]hat began as a language with a purely theoretical use has gained a reporting role" (1997, p. 107). Basically, what Sellars says is that Jones trains his Rylean fellows, by means of operant conditioning, to apply mental concepts to themselves. First, they learn to use behaviour as evidence for self-ascriptions, like an observer would do. This is an intermediate step, since at this stage there is no privileged access, no reporting, and the trainee must draw explanatory inferences from her own behaviour to inner episodes that she is postulating according to the learned theory. But eventually, this can become a kind of hypothetical reflex. If the training establishes a causal connection between her self-ascriptions and her thoughts (taken in the theoretical sense) which does not involve any mediation of the perception of the behaviour, the subject acquires a new way of accessing inner episodes. There are many subtleties regarding the Sellarsian notion of introspection, which cannot be spelled out here,[7] but Sellars does not consider this empirical, developmental problem to be particularly important. He is more concerned with the epistemic status of folk-psychological concepts and reports. While Sellars does not regard privileged access as essential, but rather as a practical consequence of language use, he does view the intersubjectivity of mental concepts as indispensable; without it, they would not be able to fulfil their primary function, i.e. that of explaining the public behaviour of others and of ourselves in an essentially causal way.

2.3 Lessons from the Sellarsian approach

Folk psychology is therefore not a scientific theory, but a mere proto-theory which could have its own place within the manifest image, as Sellars explicitly states in several places (cf. 1967, pp. 338-9). Although the motivation for its use is epistemic, as it is an attempt to understand, explain and predict the behaviour of others (and ourselves), we should not take its fundamental units, our inner mental episodes, as theoretical entities.

> I am going to argue that the distinction between theoretical and observational discourse is involved in the logic of concepts pertaining to inner episodes. I say "involved in" for it would be paradoxical and, indeed, incorrect, to say that these concepts *are* theoretical concepts. (Sellars 1997, p. 97)

As Sellars puts it, if folk-psychological proto-theory were a real theory, we should be able to indicate its subject-matter solely in observational terms, without having to use the vocabulary of the proto-theory itself. However, in the case of folk psychology, the observational and the theoretical vocabulary coincide (or at least overlap), as folk psychology operates *within* the manifest image. As part of the common-sense conceptual framework, it "has no *external* subject matter and is not, therefore, in the relevant sense a theory *of* anything" (Sellars 1967, p. 339). Regardless of this point, folk psychology exhibits an inalienable epistemic aspect, though it does not (exclusively) describe the world as scientific theories do, but rather helps us to navigate in it as agents. Its utility is necessarily epistemic but not exclusively so.

3. Mental fictionalism on the explanatory role of folk psychology

One of the strongest critiques of the standard view comes from Demeter (2013; 2022), who maintains – in favour of denying that folk psychology has any epistemic virtue – that it does not do any explanatory work. As a hermeneutic fictionalist, Demeter does not want to change the actual linguistic practice about our mental states; what he finds fault with is the (Sellarsian) manifest image of folk psychology, that is, the standard approach to the primary role of our

[7] See Knappik (2020) for further details.

mental discourse that we characterised above. (We will discuss and criticise Demeter's view in more detail in Section 3.2 below.)

Contrary to Demeter, Toon's pretence fictionalism provides a sophisticated conception of the explanatory role of folk-psychological discourse. He turns to Kendall Walton's prop-oriented make-believe view, according to which the geographical shape of Italy looks like a high-heel boot, for example; consequently, if we want to describe to someone where she can find the city of Crotone, we do not mislead her by saying that she should travel to the arch of the Italian boot. In this case, of course, we simple pretend to assert that Crotone is located on the arch of the Italian boot. In Toon's view, we talk about mental states in a similar pretending way as we do about the boot shape of Italy. Nevertheless, this interpretation of folk-psychological discourse does not preclude that it can have explanatory value:

> In this game, we imagine that people have certain states inside their heads, such as beliefs and desires. We also imagine that these states are caused by certain experiences, interact in certain ways, and cause certain sorts of behaviour. We are no more committed to the existence of this inner machinery than we are to the existence of the Italian boot. And yet pretending that this machinery exists serves an important purpose, providing us with an enormously valuable means of explaining and predicting people's behaviour. (Toon 2018, p. 163)

Accordingly, Toon tries to retain what is the primary role of folk psychology, at least according to the standard view, and maintains that in folk-psychological discourse, we simply play a game – but this pretence does not rule out that the game can serve epistemic purposes.

3.1 Towards a serious game: Toon's prop-oriented metaphorical explanations

Toon's fictionalist view tries to preserve the explanatory function of folk psychology, which is what Sellars emphasises in formulating his proto-theoretical conception. At the same time, Toon wraps these explanations in a figurative formulation. Thus, he proposes a fictionalist twist on Sellars' myth of Jones and claims that "what Jones introduces to our Rylean ancestors is not a theory, but a useful game of prop-oriented make believe" (Toon 2016, p. 283).

Let us first take a brief look at the Toonian pretence-fictionalist causal explanation by drawing an apparently adequate analogy: an angry cloud has appeared over the mountains, causing you to postpone your planned climbing expedition. Though it is literally false, this seems to be a genuine causal explanation. The cloud was in a certain state, whatever that may be, and this particular kind of cloud can be described as angry – a description that, of course, cannot be literally true, though it can be true fictionally. The point is that the genuine cause was the particular state of the cloud, which could be characterised by the adjective 'angry', thereby attributing emotions to it. In this prop-oriented make-believe, a celestial phenomenon is the prop for imagining the cloud as a big angry face.

According to Toon, using metaphors is indispensable in some cases because they "can allow us to make claims that we are unable to express in a straightforward literal description" (2016, p. 290). Stephen Yablo (1998, pp. 250-251) calls these metaphors *representationally essential*, and as such, Toon maintains, they cannot be given a literal paraphrase. In this respect, Toon argues that we can and should make a distinction between the Italian boot and the angry cloud. In the former case, we can answer the question about the location of Crotone without using the metaphor of "the arch of the Italian boot". As regards the angry cloud, Toon (2016) argues that it is representationally essential because "there might be no way of capturing what is common to all clouds we call 'angry' (apart from that they each make it fictional that they are angry)" (p. 286). But what about the following literal translation of an angry cloud? "The cloud is angry if it is a very dark grey, almost homogeneous, low-level nimbostratus or

cumulonimbus producing thunder, lightning and strong wind". This might not be a complete, literal paraphrase of an angry cloud, but we can say that it is a partial, literal paraphrase, and as such it can play the same role in our discourse about clouds as less complicated but non-literal descriptions.

Nonetheless, based on Walton's and Yablo's prop-oriented make-believe view, Toon (2016) maintains that "the metaphors of folk psychology are representationally essential" (p. 286) – and in this respect, they are rather more similar to the supposedly essentially metaphorical angry cloud than to the not essentially metaphorical Italian boot. Therefore, he has no objection to arguing that someone's behaviours can be explained by means of mental states, in the same way that we pretend that the appearance of an angry cloud over the mountains is the cause of your staying at home:

> Similarly, if we say "John's desire to go to Madrid caused him to spend all his savings," we are claiming that John is in a certain state *S* – whatever it is – such that, fictionally, we speak the truth, and this state *S* caused him to spend all his savings. Once again, even if there are no desires, it is arguable that this is still a genuine causal explanation. But, of course, it falls short of the idea that folk-psychological explanations pick out discrete inner causes of behaviour, and there is much more to be said here. (Toon 2016, p. 292)

Accordingly, all that is required to provide such a metaphorical causal explanation of John's behaviour is that John is in a certain state, whatever it may be, such that we can pretend that it is a mental state. Thus, the real cause is the state of prop, whatever it may be, not the imagined mental state. Therefore, though our inner mental states cannot be causes, appealing to them in this causal explanation seems to be indispensable. But what do we do when we describe the cause metaphorically? Why is the metaphorical use of 'desire', when an inner state is ascribed to John, similar to the metaphorical use of 'angry' when it is applied to a cloud? Why are these cases representationally essential, and, as such, why do they play an indispensable role in understanding of the nature of props? And can we understand the nature of props at all by using such metaphorical descriptions?

It should be emphasised here, as Bourne and Caddick Bourne (2020) note, that "[i]n prop-oriented make-believe, the imaginings prescribed allow for illuminating reflection back on the actual nature of the props" (p. 171), and that "Toon's account [of folk psychology as fictional discourse] is not explicit about what the props are" (p. 173). According to Bourne and Caddick Bourne, in Toon's fictionalist view, the most plausible candidates for props are overt linguistic and non-linguistic behaviours, but it is controversial how using metaphorical descriptions of folk psychology would help us to understand the nature of these props. We agree with Bourne and Caddick Bourne that

> [t]he possibility of make-believe must alert us to something about the props beyond the trivial information that, in a game where they are props for imagining such-and-such, they are props for imagining such-and-such. It is hard to articulate, in the case of Toon's folk-psychological make-believe, what more than this we learn about props. (Bourne and Caddick Bourne 2020, p. 174)

In addition to the question about (the knowledge of) the actual nature of prop, it is not clear whether a metaphor, particularly if it is representationally essential, can be *explanatorily essential*. Mark Colyvan is the first to have raised this question against Yablo's view about the metaphorical nature of mathematical discourse, in which metaphors are indispensable in genuine explanations. His dilemma is as follows:

> when some piece of language is delivering an explanation, either that piece of language must be interpreted literally or the non-literal reading of the language in question stands proxy for the real explanation. (Colyvan 2010, p. 300)

According to Colyvan, it is not uncontroversial that metaphors should be able to carry any genuine explanatory load if there are not, at least partial, literal translations of them. As Colyvan (2010) claims, all that is required for a genuine explanation is that it can provide "some partial, literal translation of the metaphor" (p. 301). As we have tried to show, we can offer such a literal translation in the case of angry clouds, if we describe the celestial phenomenon in question as a special kind of thunderstorm instead of using some metaphor, even if this description only provides a partial, literal translation of the metaphor rather than a complete one. We do not see any problem with the following non-metaphorical explanation of why you postponed your climbing expedition: the appearance of a thunderstorm, that is, a dark, low-level nimbostratus or cumulonimbus over the mountains, caused you to stay at home. No doubt, the metaphorical explanation is less complicated, but what carries the real explanatory load in this case is the more complicated non-metaphorical explanation. Thus, in this case, using the metaphor "angry cloud" has only some practical virtue but lacks an epistemic one. We agree with Colyvan (2010) when he says that "[i]t seems that metaphors can carry explanations only when the metaphor in question stands proxy for some non-metaphorical explanation. It is hard to see how there could be metaphors essential to explanation. At least, as things stand, there is no reason to believe that there are any such cases" (p. 300).

What about folk-psychological explanations of behaviours? If there are no explanatory essential metaphors, as Colyvan presumes, then a mental fictionalist is faced with the following dilemma: either folk-psychological explanations are the only genuine explanations we can give, and they should be understood literally, or metaphorical explanations should be treated as proxy explanations, and should be substituted for some non-metaphorical explanations carrying the real explanatory load. For Toon, neither option is acceptable.

We have pointed out some troublesome questions concerning the explanatory effectiveness of folk psychology, as conceived by Toon's pretence fictionalism, which need to be answered. There seem to be only two ways out of this worrying situation. On the one hand, a mental fictionalist may say that folk-psychological discourse does not play any explanatory role, and that we have to abandon, once and for all, the idea that folk psychology can serve epistemic purposes; this is Demeter's view, and we will discuss it critically in the next section. On the other hand, and we believe this to be the right way, a fictionalist may accept Sellars' non-scientific or proto-theoretical, model-based view of folk psychology, which can be both ontologically non-committal and explanatory in a non-metaphorical way.

3.2 The no explanation explanation of folk psychology

Demeter's affective storyism (2013, 2022) fundamentally departs from the idea that the content of folk psychology is relevant to its usefulness. It is far from clear whether Demeter's theory should be regarded as fictionalism at all,[8] but the question here is not the correctness of the label; it is rather whether it represents a viable strategy for fictionalists to adopt or not.

For Demeter, "folk psychology is a device of social navigation without being a device of metarepresentation: it does not aim at representing internal states, and it does not serve epistemic purposes" (Demeter 2022, p. xy). And by giving folk-psychological explanations, we are "communicating affects… This communication is successful if the listener understands how the interpreter feels" (Demeter 2013, p. 490). According to his fiction, we aim to use folk psychology to solve coordination problems, and he sees these problems as being non-

[8] Bourne and Caddick Bourne (2020) have already expressed doubts about this point.

conceptual in nature. They cannot be described by the concepts of mental discourse, since "the conceptual resources of folk psychology have no role to play" in "our success in predicting other's behaviour" (Demeter 2022, p. xy). He thinks of coordination problems and the conventions by which they are solved in a Lewisian way, without "the robust mental realism of Lewis's account". However, for Lewis, the appeal to attitudes described in folk-psychological terms is an essential part of these coordination problems. It is difficult to see what common (mutual) interest might be involved in coordination problems that have nothing to do with folk psychology. Demeter seems to think of coordination problems as being simply causal in nature, involving naturalised conventions at most. The coordination problems and their solutions can have only ontological aspects, not epistemic ones. We do not use folk psychology to explain, we simply use it to coordinate real items (subpersonal processes) along non-epistemic lines.

Demeter is talking about purely causal processes when he uses the term "social orientation". If he were not, he could not avoid assuming certain factors and processes which are necessarily intertwined with folk psychology. Moreover, he does not seem to allow for normativity either, so that coordination problems are not problems of agents but, at most, problems of the functioning of some complex systems, given that he regards affections to be subpersonal processes rather than folk-psychological kinds. All of Lewis' examples are about the coordination of mental states through action, just as Hume's rowers share a common goal. Without these goals (intentions), conventions will not only have no epistemic role to play, but we cannot really speak of problems, only of a kind of causal functioning of a system. While Demeter claims that folk psychology has no epistemic function, he interestingly argues that the purpose of its application is "to let others know how we feel about others" (2022, p. xy), which definitely seems to assume the relevance of folk-psychological concepts.

It is also problematic that Demeter, while explicitly granting autonomy to folk psychology, does so only in the negative sense, by arguing that "the acceptance of the statements in the discourse is not determined by descriptive truths belonging to the putative area of discourse – i.e. truths about the agent's internal states, behaviour, the situation, etc." (Demeter 2022, p. xy). This is not the kind of autonomy scientific explanations and theories have with respect to observational language. Demeter's autonomy has nothing to do with any kind of fact – it does not even explain social affective orientation. Even its regulatory role is in doubt, since its rules of generation are not related to the affective states themselves, which are part of the causal world. Folk psychology is simply a code for expressing and decoding affective states. The problem with conceiving it as such is that a code has no added value; it has no autonomy in the positive sense because its internal structure is merely a parasitic copy of the internal logic of the coded system. If folk psychology is a code for expressing affective states, then it is, on the one hand, very much related to the way things really are, and will thus be a good or bad representation of the fact of the matter – either way, its epistemic function is undeniable. On the other hand, if folk psychology, as a code, only copies the system of affections it conveys, then either the structure of folk psychology must also prevail among affections, or else folk psychology must be a causal discourse. Neither is good news for the affective storyist. Not to mention that if folk psychology is just a code, and affections themselves are subpersonal causal processes, then this code of folk psychology is unnecessarily complex and extremely wasteful. Its whole mind-boggling conceptual structure is pointless, since its purpose could be achieved by a kind of causal feedback system that is complex enough for the task (even without conventions). If there is only coordination of primary subpersonal processes, then we most certainly do not need a discourse as complex and autonomous (existentially highly creative) as folk psychology, nor do we need to give any epistemic function to the set of signals that perform the coordination.

An inalienable aspect of folk psychology is that it explains by means of causes. And Demeter is right that real explanations explain by means of causes. Why does he say that folk

psychology does not explain anything? Because "there are no facts independent of psychological fiction from which to derive the truths in the fiction" (Demeter 2022, p. xy):

> Affective storyism takes folk psychology to be a strongly creative discourse whose fictional entities are not connected by principles of generation to how things stand in the world. Agents become persons in the psychological fiction as they are represented by psychological concepts. These representations are not connected systematically to neural, behavioural or any other facts; hence principles of generation cannot be formulated for them. (Demeter 2022, p. xy)

However, this only means that we cannot and should not take folk-psychological explanations as scientific explanations (more precisely as theoretical ones), and that their subject matter cannot be given externally, as in the case of any normal scientific theory. And this would be one of Sellars' lessons, i.e. that although the structure of folk psychology mirrors the structure of theoretical explanations, it is different from them. While its purpose is clearly and unambiguously explanatory, it is not part of the scientific image but of the manifest image. The question what folk psychology is about can, at most, be explained by appealing to behaviour, but Demeter is right in saying that behaviour does not provide us with independent external evidence. However, this does not mean that folk-psychological attributions are not explanatory, only that they are not full-blooded theoretical explanations.

Why does affective storyism appear attractive to Demeter? Probably because it points towards a naturalistic picture of human behaviour. To approach our social existence through affects, we do not need to rely on folk-psychological concepts. While we cannot reduce the conceptually structured manifest image to the non-conceptual scientific one, we can replace the former with the latter. If it is indeed the case that folk psychology has nothing to do with the way the world really is, then it needs to be explained how it can be a successful tool or code for social orientation. And to do so would require that the principles of generation must, after all, be tied to the way the world really is. Folk psychology cannot be "frictionless spinning in a void", but affective storyism seems to claim just that.

4. Conclusion

By denying the explanatory function of mental talk, fictionalists can avoid the eliminativist and realist extremes, but this goes completely against our everyday folk-psychological practices. On the one hand, we have tried to show, through a critical examination of the positions of mental fictionalism and some lessons from Sellars, that the strategy of denying the explanatory function, as Demeter propagates it, suffers from a serious shortcoming: it cannot answer the question of how an epistemic discourse can succeed so well in performing a task that has absolutely nothing to do with either truth or explanation.

On the other hand, while Toon (2016) acknowledges the explanatory role of mental fictionalism, he also makes it vicarious by reducing folk-psychological practice to a game of pretence. While this strategy allows us to avoid realism and eliminativism and explains the normative character of folk psychology (as the rules of the game), it also fails to account for our everyday practices. Hence, it does not answer why a discourse that is merely a game, a pretence, is so successful in predicting and explaining behaviour.

Without treating folk psychology as a Sellarsian explanatory proto-theory – which is not just practically but also epistemically valuable in a sense that fictionalists do not want to acknowledge – we have thus tried to argue that the question of its "unreasonable effectiveness" still remains to be answered.

References:

Balaguer, M. (2008/2018), 'Fictionalism in the Philosophy of Mathematics', in, E.N. Zalta (eds.), *Stanford Encyclopedia of Philosophy,* URL = https://plato.stanford.edu/entries/fictionalism-mathematics/

Bourne, C. and Caddick Bourne, E. (2020), 'Folk Stories: What Has Fiction to do with Mental Fictionalism?' in B. Armour-Garb and F. Kroon (eds.), *Fictionalism in Philosophy*, Oxford: Oxford University Press, pp. 168–186.

Colyvan, M. (2010), 'There is No Easy Road to Nominalism', *Mind* 119 (474): 285–306.

Demeter, T. (2013), 'Mental Fictionalism: The Very Idea', *The Monist* 96 (4): 483–504.

Demeter, T. (2022), 'A Mental Fictionalism Worthy of its Name: A Plea for Affective Storyism', in T. Demeter, T. Parent and A. Toon (eds.), *Mental Fictionalism: Philosophical Explorations.* London and New York: Routledge.

Eklund, M. (2007/2019), 'Fictionalism', in E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = https://plato.stanford.edu/entries/fictionalism/

Knappik, F. (2020), 'Sellars on Self-Knowledge', in S. Brandt and A. Breunig (eds.), *Wilfrid Sellars and Twentieth-Century Philosophy.* New York: Routledge, pp. 221–239.

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

Lewis, D. (2005), 'Quasi-Realism is Fictionalism', in M.E. Kalderon (ed.), *Fictionalism in Metaphysics*. Oxford: Oxford University Press, pp. 314–321.

O'Shea, J.R. (2012), 'The "Theory Theory" of Mind and the Aims of Sellars' Original Myth of Jones', *Phenomenology and the Cognitive Sciences*, 11 (2): 175–204.

Parent, T. (2017), *Self-reflection for the Opaque Mind: An Essay in Neo-Sellarsian Philosophy*. New York: Routledge.

Rosenberg, J.F. (2004), 'Ryleans and Outlookers: Wilfrid Sellars on "Mental States"', *Midwest Studies in Philosophy*, 28 (1): 239–265.

Sellars, W. (1962), 'Philosophy and the Scientific Image of Man', in R.G. Colodny (ed.), *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press, pp. 35–78.

Sellars, W. (1967), 'Scientific Realism or Irenic Instrumentalism', in *Philosophical Perspectives*. Springfield: Charles C Thomas Publisher, pp. 337–369.

Sellars, W. (1975), 'The Structure of Knowledge', in H-N. Castañeda (ed.), *Action, Knowledge, and Reality.* Indianapolis: The Bobbs-Merrill Company, pp. 295–347.

Sellars, W. (1997), *Empiricism and the Philosophy of Mind.* Cambridge, MA.: Harvard University Press

Toon, A. (2016), 'Fictionalism and the Folk', *The Monist* 99 (3): 280–295.

Toon, A. (2018), 'Epistemology as Fiction', in O. Bueno et al. (eds.), *Thinking about Science, Reflecting on Art*. London and New York: Routledge, pp. 155–168.

Toon, A. (2022), 'Fictionalism and Intentionality', in T. Demeter, T. Parent and A. Toon (eds.), *Mental Fictionalism: Philosophical Explorations.* London and New York: Routledge.

Wallace, M. (2022), 'Mental Fictionalism', in T. Demeter, T. Parent and A. Toon (eds.), *Mental Fictionalism: Philosophical Explorations.* London and New York: Routledge.

Walton, K. (1993), 'Metaphor and Prop-Oriented Make-Believe', *European Journal of Philosophy*, 1 (1): 39–56.

Yablo, S. (1998), 'Does Ontology Rest on a Mistake?', *Proceedings of the Aristotelian Society, Supplementary Volume* 72 (1): 229–261.