15 Mental fact and mental fiction

Tim Crane and Katalin Farkas

The stream of our consciousness includes many kinds of episodes. There are perceptual experiences and sensations, images and daydreams, sudden flashes of memories, feelings and emotions. All of this seems very real: as we are going through these experiences, it's hard to doubt that they exist. Of course, it's a further and difficult question what their nature is, how we should grasp their tangible presence to our mind, but that is not our concern here. We will just note that at least *prima facie*, episodes in the stream of consciousness have a manifest character of reality or (as we might say) factuality.

Conscious episodes also include conscious thoughts, for example, musing, reasoning and deliberations—often mixed with other kinds of episodes like emotions. Some philosophers think that the conscious character of thoughts is different from the conscious character of other kinds of episodes, because thought doesn't have a phenomenal character. But that again is not our concern here. We just note that conscious thought, insofar as it is present to the mind, also seems to be manifestly real or factual.

Beyond the stream of consciousness, the subject's mental life also has non-conscious features. There are unconscious occurrent processes involved (for example) in visual processing, which, though unconscious, have a claim to the category of "mental", because of their connection to conscious episodes. But these again are not our topic here. Our topic is rather standing mental states: those mental features like beliefs, desires or intentions, which a subject can have even if she is not conscious, or when her consciousness is occupied with something else. In this paper we present a view that supports the claim that standing states are less real than episodes of consciousness.

In Section 1, we introduce the issue through an example: a case when a desire is attributed to someone. We want to lay out the idea that beliefs and desires are not directly present to the mind, but are posited to account for certain phenomena. In Section 2, we summarise a view of standing states defended in some earlier work, which proposes that ascriptions of these states are attempts to model what we call the subject's "Worldview": the totality of her cognitive, conative and affective dispositions. Modelling involves simplification, and when ascribing standing states, we often significantly simplify the complexity of the Worldview. In Section 3, we introduce another idea,

DOI: 10.4324/9781003091073-20

the "Habitus". The Habitus is the home of dispositions that are associated with character traits. Modelling the Habitus is like modelling the Worldview: a useful simplification of a very complex reality that helps to understand other people. In Section 4, we point out a difference between our view and Eric Schwitzgebel's phenomenal-dispositional account of belief, deriving from our emphasis on the different status of conscious episodes and standing states. In Section 5, we compare our view to the interpretationist account of propositional attitudes defended by Daniel Dennett. We agree with Dennett on the key importance of interpretation, but we do not apply an interpretationist account to conscious episodes, only to standing states. In addition, we attach great significance to the phenomenon of self-interpretation. In this respect, our view has some similarity with Richard Moran's view of self-knowledge, as we clarify in Section 6. In Section 7, we elaborate the analogy with scientific models. In Section 8, we show how our view can be regarded as a version of a minimal fictionalism, in its commitment to the twin claims that (i) a discourse can include falsehoods but (ii) is still useful for some purpose. However, we prefer not to call the theory "fictionalist", since we think that standing state and trait ascriptions are not usefully compared to literary fiction. As we state briefly in Section 9, they should be rather compared to models; and though models share some features with fiction and stories, their function is different, and they therefore serve as a better analogy for standing state ascriptions than fictions.

1 An example and a question

We start with a story that will serve as an illustration. Sofia and Ali are about to move to another country, and they visit their good friend Maja, to say goodbye. In the preceding weeks, Sofia and Ali sometimes wondered if Maja would also like to leave the country. Ali thought she certainly wouldn't, but Sofia disagreed. Now they have a chance to ask her: "Do you ever want to move, or do you want to stay here?". Maja needs to think. When someone asked her the same question ten years before, the answer was easy: she had a good job, a family, friends, she loves her country, she could immediately reply that she didn't want to move. But now the situation is different. "Do I actually want to stay here?", she asks herself. The country's politics have been deteriorating for years, with measures that started to affect her life and her friends' lives. The career prospects in her profession are dreary: she is a doctor and the health system of her country is demoralised and underfunded. She recalls the sense of excitement and slight envy she felt when she heard that Sofia and Ali are moving. She also recalls that she has been browsing job offers for doctors in another country. She put it down to idle curiosity, but now it occurs to her that perhaps there was another reason. As she thinks about all this, she realises that she actually wants to live in another country.

As we emphasised above, the conscious process of reflection seems tangibly real. The subject considers one idea after another, some of which exert

a certain pull, others a certain push. The upshot of this process—to a first approximation—is that she formed or identified a *desire.* After that, she turns her attention to something else, and these issues disappear from the stream of her consciousness. What happens to the desire? How does it survive, if it does?

It's very plausible that something survives. As it's been pointed out countless times, attributing to a subject certain beliefs or desires will help to explain many of her subsequent mental features and behaviour. The desire to live in another country leads Maja to have further thoughts on the viability of such a move. She will consider all the pros and cons, and will make investigations about jobs, schools, real estate and so on. Perhaps in the end no intention to move will be formed: she may decide that moving her children from their school would be too much of a disruption, and she decides to reconsider the question in a few years.

Sometimes philosophers talk about conscious beliefs and desires. We assume they mean by these things like the episode of consciously endorsing a statement or a proposition. Maja's conscious episode of realising she wants to live in another country is an example. We don't think talking in terms of "conscious belief" is helpful, because our focus here is the nature of standing states, that is, states that a subject can have even when she is not conscious, and we want to reserve the terms "belief" and "desire" for these states. Obviously, the standing state of believing that *p* and the conscious episode of endorsing *p* have an intimate relationship (see Crane and Farkas 2022), but they are different kinds of entities. We are directly aware of conscious episodes, but, by definition, we cannot be directly aware of standing states (only of their conscious manifestation). Standing states are present in our life through their effects.

We believe that this raises a question about their status which does not arise for conscious episodes. The existence of a conscious episode is confined to the duration of the event. The event is right there, "in front of our mind". Even if we are uncertain about its nature—is this a feeling of dread or elation?—we can still mentally "point" at it, and assure ourselves that "this experience" exists. In contrast, the reality of standing states seems to be rooted in their possible manifestations. This makes their existence more of a speculation: they are usually posited as the states that are responsible for the manifestations. What happens exactly when we posit these states is the focus of this paper.

2 The Worldview

In what follows, we rely on a view of standing states that was originally put forward by one of us (Crane 2017) and further defended in Crane and Farkas (2022). On this view, standing states do not have the kind of reality that episodes in the stream of consciousness do. We do not deny that a large part of a subject's mentality is unconscious and non-occurrent. We call this aspect "the subject's Worldview": it is her unconscious orientation towards the world, which forms

the ground for all her cognitive, conative and affective dispositions. (We will spell it with capital "W" to indicate that it is a term of art.)

There are some discernible nodes in the Worldview which can be understood using our usual notion of standing states. Beliefs and desires are best understood as the bases of certain dispositions. For example, Maja has the dispositions that are normally associated with believing that she is called "Maja". She will be disposed to have the mental episodes and behaviour that someone with this belief would have: she would answer with "Maja" if she was asked what her name is, she would have a fleeting thought "she has the same name as me" if she heard a child being addressed as "Maja" in the playground, and so on and so forth. So when we attribute her the belief that she is called "Maja", we attempt to make sense of all these dispositions.

Some of Maja's other dispositions can be understood in terms of the usual conception of desires: for example, she clearly wants her children to be happy, and many of her dispositions to act and to have thoughts, emotions, intentions and further desires are precisely those that we would expect someone with this desire to have.

Some theories of beliefs and desires regard these states as the *bases* of the dispositions to act and to have certain other mental features. Other theories just *identify* the states with having these dispositions (see Section 4). In both versions, having a desire or a belief entails having a range of dispositions. These dispositions are multi-track: they cannot be identified by a single conditional. When specifying these dispositions, most theories will agree, we have to make reference to other mental states. This can get complicated very quickly. Maja would mostly answer with "Maja" to a question about her name, but not when she doesn't want to tell her name. What about situations when she thinks her audience will suspect her of not being honest? She might do something else again.

On the simple matter of what she is called, we can trace her dispositions relatively easily, but with more complex issues, this will not work. On more complex issues, the subject's Worldview will not ground dispositions that connect to easily attributable standing states. There will be dispositions that would normally come with wanting that *p*, but also dispositions that come with not wanting *p*. Sometimes there aren't even very stable dispositions: the subject's potential actions or thoughts are sensitive to minute details in the circumstances, and are not easily characterisable as being disposed to do this, or to feel that, without going into endless details about the manifestation conditions.

Attributing beliefs and desires to people—including ourselves—is an effort to describe and interpret, as best as we can, their Worldviews. Attribution is a conscious episode and brings with it the determinacy and reality of consciousness. When Ali says that Maja doesn't want to leave her country, the conscious thought that accompanies his utterance gives a far more determinate character to Maja's state of mind that it might have in reality. It ignores many of Maja's dispositions that rather point towards her not wanting to

move. In previous work, we have characterised the conscious episodes of attribution of standing states as *modelling,* as it is done in scientific theorising. In proposing a model of a phenomenon, scientists ignore certain complexities of a situation in order to arrive at a manageable explanation. In the same way, in standing state attributions, we ignore the complexities of a Worldview and focus only on certain aspects.

3 The Habitus

Alongside the Worldview exists the subject's character, or as we shall call it, her "Habitus". (This is an extension of our previously stated view; it is not mentioned in our earlier work.) The Habitus grounds the subject's dispositions that we normally associate with personality or with character traits. In fact, the Habitus is not separate from the Worldview: this separation is already part of our interpretative, modelling efforts. Character traits, just like standing states, are usually invoked when we try to explain or predict people's actions or thoughts or feelings. They often interact with standing states in this respect. An impulsive person may swiftly form an intention when motivated by a desire; a circumspect person could have the same desire but be much slower to act. Forming the same belief will lead to envy in a small-minded person, and to joy in a magnanimous person. And so on.

Our view of the Habitus can be seen as one possible response to the kind of scepticism about character traits that was influentially put forward by John Doris (2002). Doris argued that character traits are much less consistent and much less stable than it is assumed both in ordinary attributions and in moral theory. People's actions are highly sensitive to a huge variety of factors in our circumstances. For example, people's compassion or generosity, instead of being a robust feature of their personality with a uniform impact on their actions, is surprisingly sensitive to rather trivial features of their situation. Non-moral character traits like shyness are also often relative to a large number of variables: someone could be shy in a work environment, but talkative in a private context; or shy with women, but boisterous with men, or shy with grown-ups but can easily connect with children—and so on.

So some people will have mixed dispositions: inclined to behave shyly in some contexts, and differently in others. This echoes the earlier point we made about mixed dispositions for standing states: some are suggestive of desiring one thing, others are suggestive of desiring the opposite. Not all standing states are like this, but Doris argues that this is a pervasive feature of character traits.

In response to the recalcitrance of dispositions to align according to the usual notion of standing states, one could adopt eliminativism: the view that beliefs and desires do not exist. Similarly, the observation about character traits could lead to eliminativism about character traits. This is not in fact Doris's considered view; his position is only that robust character traits are much less common than normally assumed. It would be indeed difficult to

get rid of character trait attributions. For example, it would be very difficult to manage interpersonal relations without relying on introducing people through presenting robust character traits. Here is a classic example: Jane Austen introducing the two heroines of her *Sense and Sensibility*.

> Elinor... possessed a strength of understanding, and coolness of judgment, which qualified her, though only nineteen, to be the counsellor of her mother, and enabled her frequently to counteract, to the advantage of them all, that eagerness of mind in Mrs Dashwood which must generally have led to imprudence. She had an excellent heart; her disposition was affectionate, and her feelings were strong; but she knew how to govern them: it was a knowledge which her mother had yet to learn; and which one of her sisters had resolved never to be taught.
>
> Marianne's abilities were, in many respects, quite equal to Elinor's. She was sensible and clever; but eager in everything: her sorrows, her joys, could have no moderation. She was generous, amiable, interesting: she was everything but prudent.

This description immediately gives us a sense of what the two sisters are like. Of course, as the story unfolds, their traits will prove less robust than the initial introduction suggests. We learn how Elinor's life is profoundly affected by sensibility, and Marianne at the end acts very sensibly in marrying Colonel Brandon. But we could not understand the human interest in this story if the original characterisation completely missed its mark.

We propose that we should think of character trait attributions as efforts to describe and interpret the Habitus, with the help of a model. Attributing character traits ignores certain complexities of a situation and offers a simplified structure that makes explanation and prediction tractable. Probably no one is robustly shy or robustly generous. But ascribing shyness and generosity to people in certain situations still aids our understanding of them. Models can then be refined endlessly. As we said above, standing states and character traits often invoked jointly in explanation, and hence modelling the Habitus and the Worldview is a joint enterprise.

4 Comparison with the dispositional-phenomenal account of belief

It will be best to further clarify the features of our view by comparing it to some familiar theories of standing states, with a special focus on the reality of different types of mental features. Eric Schwitzgebel has defended what he calls a "phenomenal, dispositionalist" account of belief (Schwitzgebel 2002; what he says about belief presumably applies, *mutatis mutandis*, to other standing states as well). Schwitzgebel holds that having a belief is just to possess certain dispositions to act and to have further mental features, including phenomenal states.

It is worth noting that virtually all account of beliefs (understood as standing states) connect beliefs to having some appropriate dispositions (Crane and Farkas 2022). Schwitzgebel's view differs from some other variants in at least three respects. First, unlike what he calls "representationalist" views, he doesn't put any constraint on the realisation of beliefs, and on the subject's internal cognitive architecture. Second, unlike some traditional views, he connects beliefs not only to dispositions to act or outwardly behave in a certain way, but also to dispositions to undergo some private mental episodes. Third, he notes that many of the subject's dispositions will not align clearly with either believing or not believing that *p*. As a consequence, he urges to recognise that some cases are best characterised as cases of "in-between" beliefs. He also emphasises that the question of whether someone matches the dispositional profile of a belief will be highly context-dependent, and he notes in passing that similar flexibility can be seen in the ascription of character traits.

It should be clear that our account is very similar to Schwitzgebel's. We hope we are offering something novel in two respects: first, in stressing the role of conscious belief ascription, second, in developing the view of ascriptions as modelling. In this respect, our view is also similar to an interpretationist view, and we will include a comparison with interpretationism in a moment. We think beliefs are items in a model, and the ontological status of items in models is comparable to the ontological status of fictional entities. For all that Schwitzgebel says, beliefs could be as real as conscious episodes or the underlying dispositions. In contrast, we propose that the ontological status of standing states is fundamentally different.

Consider Ali and Sofia's discussion of whether Maja wants to leave the country. We propose that Maja's Worldview includes a lot of dispositions that are relevant to this question. Ali tries to make sense of them by attributing Maja the desire to stay in the country, even though she is dissatisfied with several aspects of her life. Sofia proposes a different interpretation: that Maja wants to leave, but she is paralysed by an anxiety about starting new things and the complexity of uprooting her family's life.

Someone who is a straightforward realist about desire will have to say that there is a fact of the matter on whether Ali or Sofia are right, since one of them says that Maja has a certain desire, and the other denies this. We disagree. We think that Ali and Sofia are offering competing models of Maja's Worldview, both of which include significant simplifications. Both models can have success in explaining and predicting events in Maja's life. The models can be refined in different ways to decrease the simplification and to increase their explanatory powers, and at the end, one may prove to be more successful. But it doesn't have to be the case that only one of Ali or Sofia is right.

On Schwitzgebel's account, Maja's state of mind (at least some time prior to her self-enquiry) is likely to qualify as an "in-between" case of desiring: because there aren't enough dispositions to tilt the balance between having and lacking a desire. Consequently, on Schwitzgebel's view, it would be a

mistake both to attribute Maja the desire and to categorically deny that she has it, since she is in an in-between case. We disagree again. Even in the so-called "in-between" cases, we can benefit greatly from attributing a desire, as long as we make sure we present a complex enough picture to preserve consistency. And as long as one can present a plausible model, one can say that the subject has the desire.

5 Comparison with the interpretationist account of belief

The interpretationist view of propositional attitudes was defended by Daniel Dennett (1987). Dennett proposed that the behaviour of certain creatures can be best explained or predicted from what he calls the "intentional stance", which involves using the conceptual machinery of beliefs and desires. To have a belief is nothing more than to have one's behaviour explicable from the intentional stance. We share the view that the purpose of the concept of belief is fundamentally used to interpret others, but we depart from Dennett in a number of respects. First, unlike Dennett, we restrict this view to unconscious standing states, and do not apply it to consciousness. This helps to avoid a problem that Dennett's account potentially faces: in his view, the attributing act itself is a propositional attitude, which requires the same treatment. In our view, the attributing act is conscious, and does not need to be the product of a further act of interpretation.

Dennett and some others who are attracted to an interpretationist approach are partly motivated by the prospect of a naturalist or physicalist reduction of mental features. "Inner episodes", whether conscious episodes or standing states, provide similar challenges to this approach. We have no reductionist agenda, and we are not particularly worried about the ontological status of inner episodes of consciousness. We have tangible evidence of their reality: we are directly aware of them. What seems to us problematic, and in need of a separate account, is the realm of unconscious standing states.

Consider, by comparison, Adam Toon's suggestion that talk of mental states is a useful fiction (Toon 2016; we will further address the issue of fictionalism in Sections 8 and 9). Toon introduces his theory with the help of Wilfrid Sellars's "Myth of Jones", an imaginary account of the origin of our talk about mental states (Sellars 1956). Sellars's Jones lives in the community of our "Rylean ancestors", where people have linguistic resources to talk only about publicly observable events, including verbal and non-verbal behaviours. Jones then develops a theory that posits inner episodes, called "thoughts", which are conceived on the model of overt verbal utterances, but happen silently in the head of other subjects. Toon's twist on Sellars's story is that Jones introduces not a theory, but a useful fictional story of make-believe. This renders talk of the mind fictional. Although Toon talks mainly about beliefs and desires, Sellars's original myth clearly targets conscious inner episodes as well, and Toon signals no departure from the original story in this respect.

As we see things, it would be quite strange for someone to come up with the idea of inner episodes first by trying to interpret others, rather than by relying on their own experience. The most obvious source of our realisation that there are feelings or conscious thoughts is that *we have them*, and we are aware of them. That headaches, longings, being frightened or surprised are part of the furniture of the universe is as obvious as the presence of trees or stones or rivers. The further step, in an imaginary reconstruction of mental state talk, is not the step from observing behaviour to hypothesising inner conscious episodes. It is, rather, *the realisation that we have mental features beyond those that pass through the stream of consciousness*. It is the realisation that we don't lose our beliefs and desires when our mind is occupied with something else. To make sense of this aspect of our mental lives, we ascribe ourselves standing mental states. Jones might start to wonder if he wants another season of hunting and gathering, or instead go on a sabbatical.

A further difference between Dennett's interpretationist view and our view is the important difference between first-person and third-person belief or desire attributions. When Ali and Sofia offer an interpretation of Maja's Worldview, they are doing something like scientific modelling, we claim. When Maja is interpreting her own Worldview, she does the same. But self-attribution has a notable feature that third-person attribution lacks: it changes the Worldview itself. During the process of self-enquiry, the Worldview undergoes all sorts of changes. Maja considers various issues, and deems some of them important, others unimportant. This could give additional strength to some preferences, and weaken others. Once she formulates a desire explicitly, she will have a memory of this, and this is added to her Worldview. Once again, the conscious episode of telling herself "I want to leave the country" is very real. But the desire itself is still merely part of a model: in a lot of cases, the self-ascription still simplifies reality. And sometimes the model is badly off: this is the way to understand self-deception. Perhaps Maja has always been too much impressed by Sofia's life choices; now that Sofia decided to move, Maja can convince herself that she wants to move too. In fact, a model that didn't attribute her the desire would be more explanatory.

6 Comparison with the theory of authoritative self-knowledge

In cases like Maja's, the processes of finding out what one desires, and of formulating a desire do not really come apart. We regard this as a welcome feature of the theory. Self-enquiry could lead to self-knowledge and also to a change of mind. It is instructive to recall here another view that connects self-knowledge to formation of mental states: Richard Moran's theory of self-knowledge (Moran 2001). Moran holds that authoritative self-knowledge is available to agents who make a commitment to have the beliefs and desires they have reason to have. Such agents can then reflect on their reasons, and

find what they have reason to believe. Because of their commitment, this will tell them what they actually believe.

Moran highlights something very important about self-knowledge: that a natural way to find out whether we want something is to reflect on the reasons for wanting that thing. For example, thinking of her dreary career prospects provides Maja a reason for wanting to move. This enables us to state more precisely what is involved in modelling the Worldview. We started off by saying that the reality of standing states consists primarily in the structure of dispositions. So when we model standing states, one thing we look at is their possible manifestations. But that's not the only source of our knowledge of them. Reasons for holding standing states are also important. Reasons are not manifestations: they logically precede, rather than succeed the state. But importantly, reasons are also recoverable from the Worldview.

In our view, reasons—in the sense of the reasons for which we do things, the so-called "motivating reasons"—are best understood as standing states like beliefs and desires. For example, the desire not to spend one's working life in an underfunded and demoralised healthcare system is a reason to move. These are recoverable from the Worldview. Their recovery often may include as much simplification as the recovery of any other standing state. The perception of one standing state being a reason for another is also subject to interpretation. Maja may tell herself that her reason for wanting to move is the desire to provide a better future for her children. This could involve a certain amount of self-deception. Her strongest reason may be a dissatisfaction with her own life circumstances. She clearly has the desire to provide a better future for her children, so she is not deceived on that score. But she may be deceived in thinking that this is her main motivation. In other words, someone else modelling her Worldview by putting stress on the other motive could provide better explanations than she herself can.

Thus finding reasons for holding an attitude helps us to identify that attitude. But that's not the only way. For example, Maja's reflections included the realisation of why she started to look at job adverts for doctors. That's not a reason for moving: it is rather a manifestation of her nascent desire. When we model Worldviews—our own and others'—we rely on the totality of available data. Moran's view has been criticised for demanding too much rationality of self-knowing agents, since he requires that we commit ourselves to believe and desire only what we have a reason to believe or desire. We are also sceptical of this kind of rather lofty commitment. Anyone with a recognisably human psychology will want and believe things at least partly for reasons that they recognise, and hence they will be able to reconstruct or form their beliefs and desires on that basis. But they will also make use of lessons about their own behaviour.

We need to include one more element of our theory: the role of free agency in ascribing standing states to ourselves. We claim that prior to her self-enquiry, there is no hard fact about the Worldview that would make it that case that Maja has or lacks the desire to move. When Maja sets off to find

out what she wants, she surveys facts about the Worldview that are accessible to her, and tries to make sense of them as best as she can. Someone could wonder if the outcome of the enquiry was already fixed by the facts of the Worldview. After all, if the Worldview grounds the totality of the subject's mental dispositions, then it presumably includes her dispositions to judge her own state of mind. And it that case, the outcome of the series of mental actions that comprises the self-enquiry is fixed by the Worldview. In the cases when the resulting model is successful, this was already determined by the Worldview. So it looks as if there already was some fact about the Worldview that is pretty close to the fact that Maja had the desire already. So why aren't desires just determined by the Worldview?

What this objection ignores is the role of free agency in the effort to make sense of ourselves. It depends partly on us how we end up understanding ourselves. We can freely choose to emphasise certain things and suppress others; the result of a self-enquiry is not determined by our dispositions alone.

7 Ascriptions as models

We propose that standing state ascriptions are usefully compared to the use of models in science (for an elaboration of how models are used in mental state ascriptions, see Crane 2015). Think of a model that treats planets as having homogenous mass distribution. The planets themselves are real; the objects with homogenous mass distribution posited in the model are not: there is no such thing. Similarly, the Worldview is real, but many of the standing states and the character traits we posit in our ascriptions are not. The reason they are not real is that people simply don't have the kind of dispositional profile that is entailed by the state or trait in question.

Perhaps one difference between many scientific models and models we use in standing state attribution is that often, the kinds of things posited in scientific models do not exist at all. There are no point-sized physical objects; there are no perfectly rational agents. In contrast, at least at first sight, some beliefs and some desires clearly exist: for example, Maja's belief that she is called "Maja" is a good candidate for existence. An analogous feature of economic modelling would be if there actually were some agents who are perfectly rational: for example, some members of the Economics Department at your university. Clearly, if such agents did exist, this would not threaten the explanatory power of the model. On the contrary.

However, one could go even further and question whether there are really "beliefs" at all. Take, for example, the phenomenon of lingering discarded beliefs. If you are staying in a hotel for a week, and after the third day you are moved from the first floor to the seventh floor, chances are that when entering the elevator, you will still automatically push the button for the first floor. You did not forget that you are no longer staying on the first floor. You firmly and clearly believe that you are staying on the seventh, and that you are no longer staying on the first. But the action is not consistent with

the dispositional profile of the belief. So even on such a simple matter, the state we actually have fails to conform to the perfect dispositional profile of any belief.

It may be objected that the action *is* in fact consistent with the dispositional profile, because human beliefs have exactly the character of allowing for the effect of memory traces. That's fine with us too. As mentioned above, we can allow that according to the notion of belief used in our models, there are some actual beliefs. This will not diminish the explanatory power of the model.

8 Minimal fictionalism and mental fictionalism

We have attempted to establish a contrast between different kinds of mental features. We propose that consciousness, the Worldview and the Habitus are real. But many of what we normally call "beliefs", "desires" or "character traits" are not. Yet, these concepts play an indispensable role in our efforts to understand ourselves and others. When a discourse is found to lack referents for some of its terms, but is still useful for some purpose, it is common for philosophers to claim they adopt *fictionalism* about the discourse (for useful surveys of the idea applied to different areas, see Balaguer 2018; Eklund 2017; Nolan 2016). What we shall call "minimal fictionalism" consists at least of two commitments: first, the admission of the literal falsity of some key statements of a discourse (often because some of its terms lack a referent); and second, an insistence that the discourse should not be eliminated but rather kept, because it serves a certain purpose.

By this definition, we are minimal fictionalists about the discourse involving standing states and character traits. We think that standing states and character traits are best understood as entailing a dispositional profile, but for many of the standing states and traits we usefully attribute, the subject possesses no such profile. Note that our theory avoids one of the main challenges for general mental fictionalism: what has been called "cognitive collapse" (Wallace 2016) or "mental suicide" (Bourne and Caddick Bourne 2020). A theory of a fictionalist discourse needs to explain what makes a discourse fictional, and this is normally done with reference to mental states; namely, the attitudes of people who engage in this discourse. If we are fictionalist about the whole mind, then the attitudes that ground the discourse are also fictional, so a circularity or infinite regress threatens. We have already mentioned a version of this worry when comparing our view to Dennett's interpretationist account, and we noted that it doesn't apply to our theory. We think that the "fiction" of standing states and character traits is firmly grounded in the reality of *conscious* attributions. There is no regress or circularity.

This is certainly not fictionalism about "folk psychology", if "folk psychology" is our everyday mentalistic discourse. That discourse clearly refers to many conscious episodes (like pain), and according to us these are wholly real. However, discussions of "folk psychology" often mention "beliefs" and "desires",

without making it clear whether they are talking about conscious episodes or standing states. We suspect that the "folk" concept involves both these two ideas. However, in a philosophical discussion, they can and should be distinguished.

"Fictionalism" is a philosophical term of art which has been understood in different ways in the philosophical debates of the past few decades (see Kalderon 2005; Armour-Garb and Kroon 2020). The minimal commitments mentioned above—admission of falsehoods, yet insistence on keeping the discourse—can be developed in various directions, and there is no consensus on what a more substantive "fictionalism" should involve. Bradley Armour-Garb and Frederick Kroon (2020) recently pointed out that even the three entries in the Stanford Encyclopedia of Philosophy—a standard reference work—on "Fictionalism", "Modal Fictionalism", and "Fictionalism in the Philosophy of Mathematics" give three mutually incompatible definitions. Since everyone seems to have their favoured sense of "fictionalism", we could also elaborate our own account by giving our own definition of the word.

The problem with the term "fictionalism" is that it can suggest that discourse about standing states is similar to literary fiction, and this is somewhat misleading. Of course, "fictionalism" is a term of art, and on some definitions not even the discourse in literary fiction counts as fictional. For example, in a highly influential paper defending moral fictionalism, Daniel Nolan, Greg Restall and Caroline West introduce the idea of fictionalism as follows:

> The simplest fictionalist approach to a discourse takes certain claims in that discourse to be literally false, but nevertheless worth uttering in certain contexts, since the pretence that such claims are true is worthwhile for *various theoretical purposes*.
>
> (Nolan, Restall, and West 2005, p. 308, emphasis added)

Our own minimal commitment is more minimal than this, since our minimal fictionalist would like to preserve the discourse for some (unspecified) purpose, rather than for a *theoretical purpose*. Literary fiction and children's stories arguably don't serve a theoretical purpose. On the conception adopted by Nolan et al., literary fiction and stories do not employ a fictionalist discourse.

Though this is largely a matter of stipulation and terminology, we still think it's an unwelcome feature of a theory of "fictionalism" that it doesn't apply to actual fiction. In earlier work and in this paper, we have compared attributing standing states to *scientific modelling*. You may think that just leads back us to fictionalism again, since there is a venerable tradition in the philosophy of science to treat scientific models as some sort of fiction (Suárez 2009; Toon 2012). If we were to adopt minimal fictionalism, or a more substantive fictionalism that was *not* designed to account for fiction (like Nolan et al.'s fictionalism above), then we could agree with this. In this case, comparing standing state attributions to modelling, and having a fictionalist account of models themselves, will result again in accepting mental fictionalism, through a more circuitous route.

However, as already indicated, we are not happy with the term "fictional". There are important similarities between models and fiction, but we also think there are important differences, and the differences tilt the balance against treating models as fictions. Furthermore, standing state attributions are more like modelling than fiction. So in the end, we prefer not to say that we defend fictionalism about the mental.

9 Models rather than fiction

We compared the attribution of standing states and character traits to scientific modelling. In attributions, like in modelling, we simplify the target situation (the portion of the world we would like to explain or understand) to make it intelligible. We propose that the comparison of standing state attribution to models is more illuminating than a comparison to works of fiction. Unfortunately, we will have to restrict ourselves to stating our claim without going into the debate much, since we have no room to address this issue in its complexity here. Our aim here is merely to highlight an option for the treatment of mental discourse that lacks referents for some of its terms.

Mental state attributions are like theoretical (rather than mechanical or three-dimensional) models, and we restrict our enquiry to these types of models. Among the different varieties of fiction, theoretical models best resemble literary fiction or fictional stories (like children's stories). We believe that while there are some important similarities between models and stories, there are also important differences.

First, the similarities. Both in models and in stories, we use imagination. This could involve imagining that some existing objects have properties that are different from their actual properties, or we could imagine particular things and kind of things that don't exist or even couldn't possibly exist. When we imagine some actual things to be different from the way they are, it's clear what we are talking about—indeed, we can stipulate what we are talking about. But when we imagine things that don't or cannot exist, we need to explain how positive claims about them can be part of a useful discourse. We also need to account for the difference between the status of "Elinor and Margaret Dashwood are sisters" and "Elinor and Margaret Dashwood are cousins". Neither of them states an actual fact, but one is part of the fiction (the novel), and the other isn't.

As far as we can see, the account of referent-free discourse, and the account of what is part of the fiction/model, can be the same for works of literary fiction and for models. Your favourite account of how referent-free discourse works can be plugged in the theory and used for both kinds of discourses. If this is still called "fictionalism", it is a more substantive version than that advertised in Section 7. And if this is what we mean by fictionalism, then we are still fictionalists about standing states and character traits.

However, it seems to us that this conception of "fictionalism" ignores important differences between literary fiction and stories on the one hand,

and scientific modelling on the other hand. We propose it's more useful to regard both models and literary fiction as subcategories of a more general kind, rather than treating models as a type of fiction. We agree with Ronald Giere that while the ontological status of things invoked in models and in fiction may be the same, the two have different functions (Giere 2009).

Giere emphasises that though scientific models serve many purposes, one seems to have a special importance: they all aim to represent (directly or indirectly) aspects of the world. This point gets complicated by the many differing views on scientific realism and representation, but plausibly, on all views, the point of designing a model is to describe, depict, explain or otherwise understand a specific portion of the world.

Importantly, the model will do this by containing or implying statements that are known to be false, or in any case, not known to be true about that portion of the world. But each of the (possible) falsehoods is introduced out of necessity. The necessity frequently comes from the need to simplify. The target system of the model (the portion of reality we are interested in) is too complex, so we will abstract away from certain complicating features, we will use idealisations, and so on. In some cases, postulating a non-existent thing helps to make a situation tractable. In other cases, the necessity comes from ignorance: there are features of the portion of reality that we do not know, so we build a model using only the known parameters. The introduction of the (possible) falsehoods is tolerated only in the service of depicting the portion of reality as best as we can, with no other purpose being served.

Someone might point out that fiction can also aim to depict reality. There are fictional stories about historical events, historical periods or historical persons. Other stories depict romantic love, the phenomenon of coming of age or redemption by faith, or, at its most general, the "human condition". Furthermore—this line continues—you cannot arbitrarily depart from certain truths in a fictional work either, and fictional works can be criticised if this rule is not observed. If a story portrayed three sisters in 18th-century England who had the ambition of becoming a doctor, a scientist and an explorer, and their family and friends were maximally supportive of these ambitions, one could rightly object to this departure from reality. In addition to historical facts, stories also have to respect, at least in broad outlines, facts about human psychology. A particular version of this point is sometimes discussed under the label "imaginative resistance" (Gendler 2000 and Tuna 2020).

These points are certainly well taken. Even if imagination can seem boundless, a story, if it is to speak to us, cannot be entirely arbitrary. So it seems that stories are bound in a way similar to models: not just any departure from reality is allowed. But there is still a difference. The departure from reality in fiction is not out of necessity, but it's prompted by the aesthetic or entertainment or ethical aim of the fictional work. This seems to have no parallel in modelling. You cannot motivate a falsehood introduced into a model by pointing out its aesthetic function. And in this respect, the attributions of

standing states and character traits are similar. When Maja is trying to find the best model that fits her diverse dispositions, she is driven by the aim of finding the best depiction of her state of mind. The best scenario is if she doesn't need much simplification, and her dispositions align along a simply statable desire. In some cases, like in the case of moving, she will not have this. And then she will have to simplify a lot, but this cannot be done in order to offer an edifying story, or a gripping narrative. If she is driven by those aims, then she could easily end up deceiving herself.

We realise that there is hardly a need for a new term of art in this area, but if we were allowed to coin a new term for our view, it would be the rather ungainly mental "modelism", rather than mental "fictionalism". We may be wrong in thinking that "fictionalism" carries a suggestion of a similarity to literary fiction; perhaps philosophers using the term have become so much accustomed to the philosophical use of the term that they don't even make this association anymore. But someone not well versed in these debates might. When Ronald Giere argues that models should not be treated as fiction, one of the reasons he offers is what he sees as a negative effect on the public perception of science, especially in the US. He thinks that supporters of "creation science" could get the wrong message if they learned that "respected philosophers of science think that science is just a matter of fictions" (Giere 2009, p. 257). We don't know if Giere is right about this, but it seems to us that a similar worry may arise about saying that mental state ascription is simply fiction, especially in the first-person case. Achieving self-knowledge is a hard but important quest. It is often difficult to face what our real motives, our real desires and real character traits are. To be told that "we just make this stuff up" does not set the right tone for this exercise.

References

Armour-Garb, B. and Kroon, F. (eds.) (2020), *Philosophical Fictionalism*. Oxford University Press.

Balaguer, M. (2018), 'Fictionalism in the Philosophy of Mathematics', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition). URL = https://plato.stanford.edu/archives/fall2018/entries/fictionalism-mathematics/.

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

Crane, T. (2015), 'The Mental States of Persons and their Brains', in A. O'Hear (ed.), *Mind, Self and Person*, Royal Institute of Philosophy Supplement 76. Cambridge: Cambridge University Press, pp. 253–270.

———. (2017), 'The Unity of Unconsciousness', *Proceedings of the Aristotelian Society*, 117 (1): 1–21.

Crane, T. and Farkas, K. (2022), 'The Limits of the Doxastic', in Uriah Kriegel (ed.), *Oxford Studies in Philosophy of Mind,* 2: 36–57.

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

Doris, J.M. (2002), *Lack of Character: Personality and Moral Behavior.* Cambridge: Cambridge University Press.

Eklund, M. (2017), 'Fictionalism', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition). URL = https://plato.stanford.edu/archives/win2017/entries/fictionalism/

Gendler, T. (2000), 'The Puzzle of Imaginative Resistance', *The Journal of Philosophy*, 97 (2): 55–81.

Giere, R. (2009), 'Why Scientific Models Should Not Be Regarded as Works of Fiction', in M. Suárez (ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization*. New York and London: Routledge, pp. 248–258.

Kalderon, M.E. (ed.) (2005), *Fictionalism in Metaphysics*. Oxford: Oxford University Press.

Moran, R.A. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Nolan, D. (2016), 'Modal Fictionalism', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). URL = https://plato.stanford.edu/archives/spr2016/entries/fictionalism-modal/.

Nolan, D., Restall, G., and West, C. (2005), 'Moral Fictionalism versus the Rest', *Australasian Journal of Philosophy*, 83 (3): 307–330.

Schwitzgebel, E. (2002), 'A Phenomenal, Dispositional Account of Belief', *Noûs*, 36 (2): 249–275.

Sellars, W. (1956), 'Empiricism and the Philosophy of Mind', *Minnesota Studies in the Philosophy of Science*, 1: 253–329.

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

Toon, A. (2012), *Models as Make-Believe: Imagination, Fiction and Scientific Representation*. Basingstoke: Palgrave Macmillan.

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

Tuna, E.H., 'Imaginative Resistance', in. E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), URL = <https://plato.stanford.edu/archives/sum2020/entries/imaginative-resistance/>.

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