



SIMPOSIO

Defending functionalism and self-reference in memory

Jordi Fernández

University of Adelaide, Adelaide, Australia

E-mail: jorge.fernandez@adelaide.edu.au

Recibido: 9 de marzo de 2021 | Aceptado: 9 de junio de 2021

<https://doi.org/10.17533/udea.ef.n64a12>

Abstract: In recent work, Sarah Robins, Gerardo Viera and Steven James have provided some insightful objections to the ideas offered in my book, *Memory: A Self-Referential Account*. In this paper, I put forward some responses to those objections. Robins challenges the idea that being a memory could be a matter of having a particular functional role within the subject's cognitive economy. Viera challenges the idea that the content of a memory could explain some of its phenomenological properties. And James challenges the idea that our memories could be immune to error through misidentification. All three commentators are targeting, not tangential aspects of, but fundamental assumptions in the account of memory proposed in the book. For that reason, modifying some of those assumptions would amount to proposing a whole different account of memory. I hope to show, however, that such a radical move is not necessary. For there are possible responses to the objections from all three commentators which are available within the constraints of the account proposed in the book.

Keywords: memory, functionalism, phenomenology, immunity to error through misidentification

Cómo citar este artículo

Fernández, J. (2021). Defending functionalism and self-reference in memory. *Estudios de Filosofía*, 64, 223-236. <https://doi.org/10.17533/udea.ef.n64a12>

OPEN  ACCESS





SIMPOSIO

Defendiendo el funcionalismo y la auto-referencia en la memoria

Resumen: En trabajos recientes, Sarah Robins, Gerardo Viera y Steven James han presentado algunas objeciones a las ideas ofrecidas en mi libro *Memory: A Self-Referential Account*. En este texto, presento algunas respuestas a esas objeciones. Robins desafía la definición funcional del recuerdo. Viera desafía la idea de que el contenido de una memoria pueda explicar algunas de sus propiedades fenoménicas. Y James desafía la idea de que nuestros recuerdos puedan ser inmunes al error de identificación errónea. Los tres comentaristas están apuntando, no a aspectos tangenciales, sino a supuestos fundamentales de la propuesta presentada en el libro. Por esa razón, modificar algunos de esos supuestos equivaldría a proponer una explicación de la memoria completamente diferente. Sin embargo, espero demostrar que un paso tan radical no es necesario: hay respuestas a las objeciones de los tres comentaristas que están disponibles dentro de las limitaciones de la teoría presentada en el libro.

Palabras clave: memoria, funcionalismo, fenomenología, inmunidad al error por identificación errónea

Jordi Fernández

is associate professor of philosophy at the University of Adelaide. He is the author of *Transparent Minds: A Study of Self-Knowledge* (2013) and *Memory: A Self-Referential Account* (2019).

ORCID: 0000-0002-4502-1003



First of all, I want to thank Sarah Robins, Gerardo Viera and Steven James for their insightful commentaries on *Memory: A Self-Referential Account*. I am not sure that I will be able to do full justice to all of their commentaries in my replies, but I hope that I have been able to concentrate on their most substantive points regarding the book. I am also very grateful to the editors of *Estudios de Filosofía* for giving us the opportunity to discuss the book *Memory* in this journal.

1. James on immunity to error through misidentification

Steven James (2021) concentrates on the discussion of immunity to error through misidentification (for short, ‘EM’) in the book, and my proposed defence of the view that memory judgments are IEM; a defence which appeals to the content of memories. James begins by motivating the view that memory judgments are not IEM through an intuitive example, and through an explanation of why we have the intuition that the memory judgment concerned is not IEM. Then, James puts forward one main criticism for my defence of the view that memory judgments are IEM. This is the criticism that the proposed defence trivialises the view. According to James (2021), the proposed defence of the view that memory judgments are IEM makes the view uninteresting, for two reasons. For one thing, it delivers IEM, as it were, on the cheap, because it is merely based on a technicality. For another, it deprives the thesis of any significance about self-consciousness.

Let us consider James’s example first. In the example, Gayle recalls getting winked by Delta Burke. It turns out, however, that it was her sister Linda who got winked; an episode that Gayle witnessed, and one that she recalls correctly. In this example, James (2021) tells us Gayle’s judgment that she got winked by Delta Burke is vulnerable to error through misidentification. For it is based on an accurate memory of a winking by Delta Burke. But Gayle wrongly assumes that the person who got winked was her and, as a result, her judgment that she got winked by Delta Burke is wrong. Does this not show that memory judgments are not IEM? It seems to me that this example trades on a certain ambiguity about the kind of mental image involved in Gayle’s remembering. Gayle may be visualising the winking episode either from the first-person point of view (or from the inside), or from the third-person point of view (or from the outside, with her as one of the participants in the episode). In the former case, I would argue that, even if Gayle’s mental image qualifies as a memory, it is not an accurate memory. After all, the remembered episode did not happen as Gayle is visualising it (Delta did not wink while looking at Gayle). In the latter case, I would argue that Gayle’s mental image does not qualify as a memory to begin with.

According to James (2021), if we assume the functionalist account of the metaphysics of remembering proposed in the book, then Gayle’s mental image of the winking episode should qualify as a memory. For it is the kind of thing that tends to be caused by such

an event happening. This may be true if Gayle is visualising the winking episode from the first-person point of view, but not if she is visualising it from the third-person point of view. In the latter case, it is not the kind of mental image which tends to be caused by winking episodes. The reason is that Gayle is never in a position to perceive herself being winked at, at least not in the manner in which a witness of the episode would be in a position to perceive Gayle when somebody winks at her. Arguably, she can never be in such a position, because she cannot be in two places at once.

The second horn of my response reveals a deeper disagreement between James and me. According to James (2021), it is no wonder that, in the Delta Burke example, we have the intuition that Gayle's judgment is vulnerable to error through misidentification. After all, we know that memories are formed, modified over time, and reconstructed later (often repeatedly). The way James sees it, then, it would be surprising if memory processes did give rise to the IEM phenomenon. And I can see how if someone accepts such a liberal conception of what counts as a memory, then they would reach the conclusion that memory judgments are not IEM. I therefore think that, ultimately, the disagreement between James and me, with regards to the IEM phenomenon, stems from a difference in our conceptions of what it is for someone to remember something. On the functionalist construal of remembering, a mental image of some event does not need to have been caused by a perception of that event in order for it to count as a memory of the event (if it did, we would not be able to misremember events that we correctly perceived in the past). Nevertheless, on the functionalist construal of remembering, the mental image does need to be the type of mental image which would normally be caused by such an event. And the reconstruction processes to which James (2021) alludes can interfere with this property of the mental image. In the Delta Burke case, for instance, Gayle has a mental image which has been reconstructed from her original perception of the winking episode (assuming, that is, that Gayle is visualising the winking episode from the third person point of view). And the degree of reconstruction is such that the resulting mental image is no longer the kind of mental image which would normally be caused by winking episodes, because we never perceive ourselves from the outside. In this case, Gayle's memory of the event has not been reconstructed by the processes to which James is referring. The way I see it, Gayle's memory has been eliminated by those processes.

Let us now consider the criticism that my proposed defence of the view that memory judgments are IEM makes the view uninteresting because it delivers IEM on the cheap. The reason why James thinks this is that my definition of what it is for a memory judgment to count as being IEM requires that it is impossible for a misidentification error to occur while the memory on the basis of which the judgment has been made remains fully accurate. This condition is doing the heavy lifting in my defence of the IEM view. However, James (2021) objects, if we include this condition in our definition of what it is for a memory judgment to be IEM, then IEM becomes not only a matter of whether it is impossible for the judgment to be false because of a misidentification

of the person, or object, about whom one made the judgment. And, the implication is, IEM should only be a matter of whether that possibility has been ruled out. Thus, the reason why my proposed defence of the IEM view succeeds, James suggests, is that we have re-defined the IEM notion in a somehow inappropriate, or ad hoc, way.

I disagree with the claim that if we require, for a memory judgment to be IEM, that it is impossible for a misidentification error to occur while the memory on the basis of which the judgment has been made remains fully accurate, then this definition makes the thesis that memory judgments are IEM uninteresting. On the contrary, it seems to me that the thesis becomes uninteresting if we do not include such a requirement. For if we do not include such a requirement, I agree with James (2021), the thesis becomes false. But it becomes trivially false. Suppose that I misremember being in front of a tree a week ago and, on the basis of that incorrect memory, I judge that I was in front of a tree. If the question of whether my judgment is IEM is only the question of whether it is possible for my judgment to be wrong because I was not in front of a tree, even though someone else was, then, naturally, my judgment is not IEM. Someone else could have been in front of the tree that I am misremembering. But, then, the reason why my memory judgment is not IEM is simply that memory is fallible. At any point, any of my memories may falsely present to me an event consisting in my instantiation of some property in the past. If, in the past, that property happened to be instantiated by some other person, then the possibility that IEM is supposed to rule out will obtain. This reading of the claim that memory judgments are not IEM, however, only tells us that memory is fallible, and surely that is not the interest of the IEM thesis with regards to memory. The reason why we need the requirement that, in order for a memory judgment to be IEM, it needs to be impossible for a misidentification error to occur *while the memory on the basis of which the judgment has been made remains accurate* is precisely that we want to rule out the possibility that an error occurs just because memory is fallible.

Let us consider, finally, the criticism that my proposed defence of the view that memory judgments are IEM makes the view uninteresting because it deprives the thesis of any significance with regards to self-consciousness. In James's view (2021), the proposed defence of the IEM view tells us that the reason why memory judgments are IEM is that memory judgments have a certain content, a content which involves the self. And, from this diagnosis, I draw a lesson with regards to self-consciousness, namely, that our first-person conception of ourselves, or self-conception (a conception formed through faculties such as introspection, proprioception, or memory), includes the fact that we are the kinds of beings which are not only extended in space, but also in time. Now, if this is right, James objects, then the view that memory judgments are IEM does not tell us anything interesting about our self-conception. The view that memories have certain contents which involve the self does tell us something interesting about our self-conception, but the view that memory judgments are IEM plays no role in delivering the relevant result.

I confess that I am a little confused about the dialectic here. James (2021) seems to concede that the fact that memory judgments are IEM suggests that memories have certain contents which involve the self. And he also seems to concede that the fact that memories have certain contents which involve the self suggests that our self-conception is that of a being extended in time. Then, why does he dispute the claim that the fact that memory judgments are IEM suggests that our self-conception is that of a being extended in time? If we assume that ‘suggests that’ is transitive, James’s worry cannot really be that the view that memory judgments are IEM plays no role in delivering the result that our self-conception is that of a being extended in time. Instead, his worry must be that this is a result at which we do not need to arrive through the IEM route. If we have other grounds for believing that memories have certain contents which involve the self, then we will obtain the same result about our self-conception, independently of our position with regards to the question of whether memory judgments are IEM. If this is James’s point, then I do not disagree with his point. But it seems to me that there is a considerable leap, from this point, to the point that the IEM phenomenon in memory tells us nothing interesting about our self-conception.

2. Viera on the feeling of pastness

Gerardo Viera (2011) concentrates on the discussion of the feeling of pastness in memory, and my proposed explanation in terms of it; an explanation which appeals to the content of memories. Viera raises two concerns for my explanation, depending on how exactly the explanatory link between the content of memories and the feeling of pastness is conceived. He takes these two concerns to show that the proposed explanation does not succeed. Accordingly, he proposes a different way in which time can be part of the content of memories, in terms of what he calls ‘path-dependent representations’ of time. Viera’s proposal makes use of David Lewis’s distinction between personal time and external time. I think that this distinction is very helpful for the purposes of explaining the feeling of pastness in memory, and I am grateful to Viera for bringing it up. As far as I can see, however, this distinction can actually be deployed to address Viera’s two concerns about the explanation of the feeling of pastness in memory in terms of memories having self-referential contents.

If I understand Viera’s (2021) discussion of my proposed explanation of the feeling of pastness correctly, his criticism of it has the form of a dilemma. My proposal is that, when we are aware of a remembered event as being in the past, what the relevant memory represents is that a perception of that event has caused the memory. What the memory represents, then, is a certain property of the remembered event, namely, the property of being at the causal origin of the memory. The question that Viera seems to be raising, in response, is the question of whether that property is the property of being in the past, or not. And, the way he sees it, significant difficulties are going to

arise for my proposed explanation of the feeling of pastness whether we answer this question in the affirmative, or in the negative.

If we decide that the property of being at the causal origin of a memory is not the property of being in the past, then the proposed explanation is an error theory. The outcome is that the feeling of pastness is not a feeling of pastness, but a feeling of causation. We misdescribe this experience as a feeling of pastness but, due to the close connection between causation and time, our experience of a remembered event always picks up an event which, in fact, is in the past.¹ This, Viera (2021) tells us, is not a satisfactory explanation of the feeling of pastness, and it is not what my proposed explanation of the feeling of pastness promised us. By contrast, if we decide that the property of being at the causal origin of a memory just is the property of being in the past, then my proposed explanation of the feeling of pastness becomes similar to an objectivist approach to colour. The thought is that, by looking at a blue wall, my perceptual experience represents what is in fact a reflectance property of the surface that I am looking at, and colour blue is simply that property. Analogously, by remembering my seeing an apple, my memory experience represents a property of that event, namely, the property of being at the causal origin of my memory. And being in the past is simply that property.² But this, Viera tells us, is highly implausible. The pastness of a remembered event cannot consist in a causal property of it.

In fact, Viera could have pointed out that things seem even worse for the proposed account of the feeling of pastness in terms of memories having self-referential contents. It is not only that the account seems to run into trouble whether the crucial question of whether, in memory, we represent the past is answered in the affirmative or in the negative. It is also that, in different places, I seem to be giving different answers to the same question. So the account, one could argue, is not only uninformative, but also incoherent.

In order to present his own account of the feeling of pastness, Viera raises Lewis's distinction between external time and personal time. Normally, a remembered event is earlier than the relevant memory in both personal time and external time. In Viera's example, he remembers, in 2021, his struggling to swim in 1991. His struggle, then, is earlier than his memory of it in external time (since 1991 is earlier than 2021), and in personal time (since it is part of Viera's history, or his life). The case of the time traveller, however, illustrates how a remembered event can be earlier than the relevant memory in personal time, but not in external time. If Viera travels back to 1989 and he remembers, then, his struggling to swim in 1991, then his struggle is earlier than his memory in personal time (since the event remains a part of Viera's life) but not in external time (since 1991 is not earlier than 1989). Viera's suggestion, then, is that, in

1 See (Fernández, 2019, pp. 108-109) for this view.

2 See (Fernández, 2020, p. 294) for this view.

memory, we are aware of an event as being in the past because our memory of it gives us a variety of cues to locate the event in our personal past.

Viera's account of the feeling of pastness is an interesting proposal. However, I will not be engaging with it as a competing account. Instead, I would like to show that, once we help ourselves to the distinction between personal and external time, we can tackle Viera's dilemma for the account of the feeling of pastness in terms of memories having self-referential contents. The question, let us remind ourselves, is whether the property of a remembered event of being at the causal origin of the relevant memory is the property of being in the past or not. I am inclined to give different answers to this question depending on whether we have external time or personal time in mind.

If we have external time in mind, then the answer is 'no'. In that case, I agree that my account of the feeling of pastness becomes an error theory. We talk of a feeling of pastness but, strictly speaking, what the feeling is a feeling of is a causal relationship holding between the memory that we are having and the remembered event. Viera replies that this is not how the proposed account of the feeling of pastness is presented, but I think I have been quite explicit about what the proposed view, both in the book, and in earlier work.³ Leaving aside issues of presentation, though, Viera's (2021) concern may be that an error theory of the feeling of pastness is not much of an explanation. In that case, I have little to say in response. I can point out that this error theory tells us, not only why it is not appropriate to talk of a 'feeling of pastness' in memory, but it also explains why we are inclined to talk in that way. But if the objection concerns the deflationary nature of error theories as explanations generally, then I concede to Viera that the proposed account is an 'explanation' of the feeling of pastness in a weak sense of the term.

If we have personal time in mind, then I am inclined to answer 'yes' to the question of whether the property of a remembered event of being at the causal origin of the relevant memory is the property of being in the past. In this case, I embrace the analogy with colour objectivism. The feeling of pastness is the way in which we experience something about remembered events that our memories represent. What they represent is not that those events have some position in external time, but the fact that those events are causally related to our memories. And the property of being causally related to our current states, such as our memories, is the property of being in our personal past. Think about it in this way: What is it for a remembered event to be in our personal past, that is, to be part of our life? It is for it to be something that happened to us. And what is it for an event to be something that happened to us? It is for the person to whom it happened to be identical with the person who is having the relevant memory. But what makes the two people identical? One plausible answer, it seems to me, is that the properties of one of the two people are causally responsible for the properties of the

³ See footnote 2, and (Fernández, 2008, p. 349).

other one. And, in that case, it seems reasonable to think that what it is for an event to be in our (personal) past is for it to be causally related to our current properties. If the event in question is a remembered event, then our memory of it seems the natural candidate for our relevant current property.

In response to the colour objectivism approach, Viera objects that this approach commits us to the prediction that, in other instances in which we represent causation by occupying other mental states, then we should find that we experience a similar phenomenology to the phenomenology of pastness in memory. And this prediction, Viera points out, seems to be wrong. It seems to me, however, that the approach only commits us to the prediction that, in other instances in which we represent causation in the way represented by memories by occupying other mental states, then we should find that we experience a similar phenomenology to the phenomenology of pastness in memory. And this prediction seems to be consistent with the empirical evidence provided by Viera. For, as far as I can see, the mental states involved in the cases presented by Viera, are not causally self-referential. If they represent causation, they do not represent it by representing themselves as one of the relata involved in the represented causal relations.⁴

Viera's own account of the feeling of pastness in terms of path-dependent representations of time may be a compelling alternative to the subject-dependent and subject-independent theories of temporal representation that I consider, and ultimately dismiss, in the book. My aim in this discussion has not been to challenge Viera's account, but to suggest that, even if it is successful, my proposed account of the feeling of pastness, in terms of memories having self-referential contents, can overcome the difficulties that Viera raises for it, assuming the same conceptual resources.

3. Robins on Functionalism

Sarah Robins (2021) discusses the functionalist account of remembering proposed in the book, and she raises four worries for the account. Two of those worries are general worries, and two of them are more specific ones. The first general worry is the following: on the functionalist account, having a mental image *i* counts as remembering that some proposition *p* is the case if and only if having *i* has a certain functional role in the subject with regards to the fact that *p*. But having *i*, Robins points out, is an occurrent mental state and, as such, it can have no functional role. It does not tend to be produced by anything, and it does not tend to produce anything. The two specific concerns about the functionalist account of remembering stem from this general worry. Robins (2021)

4 I therefore disagree with the view (Searle, 1983) according to which perceptions are causally self-referential. Our intuitions about veridical hallucination seem to suggest that the kinds of thought experiments employed to motivate the causally self-referential nature of memories would not work in the case of perception.

claims that the functionalist treatment of the Embellishment case, and the functionalist treatment of the Confabulation case, are not satisfactory due to this shortcoming in the functionalist account. Robins considers various readings of the claim that having mental image *i* plays a functional role in the subject with regards to the fact that *p*, and concludes that none of them can deliver a satisfactory account of our intuitions about remembering in both Embellishment and Confabulation. The second general worry is that, since having mental image *i* is an occurrent mental state, no reading of the claim that having *i* has a certain functional role with regards to the fact that *p* can capture our intuitions about what it is to remember that *p*. For functional roles concern causal relations in which types of mental states, such as mental images, are involved. And what we want to know, when we ask what it is for a subject to remember that *p* in virtue of having mental image *i*, is what it is for the subject, in this particular instance of their having *i*, to remember that *p*. Our question, in other words, concerns the token mental state that the subject occupies, and not the type to which it belongs.

Let us take the first general concern first. Robins's concern is well-taken in that, sometimes, I speak of tendencies in which having a mental image *i* is involved, and sometimes I speak of the episode of having mental image *i* satisfying certain counterfactuals. I do waver between these two conceptions of what it is for mental image *i* to have a certain functional role, which Robins has noticed. To simplify matters, then, let us focus on the counterfactual reading of the functionalist account, and see whether that counterfactual reading can give us a successful treatment of both Embellishment and Confabulation. The idea is that a subject *S* having a mental image *i* remembers that *p* just in case, if *S* perceived that *p*, then they would have a mental image of *i*'s type. And if they had a mental image of *i*'s type, then they would believe both that *p* and that they perceived that *p* in the past. There is an appeal to types here.⁵ Which is, then, the relevant type? I am inclined to individuate mental images phenomenologically. Thus, we can think of the relevant type as the phenomenological type to which mental image *i* belongs.

In Embellishment, the intuition is that the subject remembers a black rabbit being shot (even though, in fact, a white rabbit was shot). And it seems to me that the relevant counterfactual is indeed satisfied: If the subject had witnessed a black rabbit being shot, the mental image that they are actually having is the type of mental image that they would have had. To be sure, they remember it incorrectly because, even though the mental image they are having is the type of mental image that they would have had if they had witnessed a black rabbit being shot, they did not in fact witness a black rabbit

5 Robins is right in pointing out that the functionalist account of remembering makes a further appeal to a background of normal circumstances. Thus, the counterfactuals above would need to be qualified by prefacing them with 'normally'. However, I do not believe that this qualification is necessary in order to deal with the cases of Embellishment and Confabulation, which are the cases on which Robins is concentrating. Thus, for the purposes of this discussion, I will leave the qualification implicit in the formulation of those counterfactuals.

being shot. Nevertheless, the counterfactual is satisfied, which captures, I suggest, the intuition that the subject is (incorrectly) remembering that a black rabbit was shot. Robins (2021) objects that the fact that, in Embellishment, the counterfactual is satisfied does not allow us to reach any conclusions about the tendencies that the mental image of the black rabbit being shot has in the subject. This seems right. But I am not offering, here, the counterfactual reading of the functionalist criterion for remembering as a way of specifying what the tendencies of mental images are in remembering. I am offering it as a way of specifying what their functional roles amount to.

In Confabulation, the intuition is that the Korsakoff subject does not remember having a conversation in a train. And it seems to me that the counterfactual is not satisfied: If the subject had experienced the conversation in the train, the mental image that they are actually having is not the type of mental image that they would have had. The reason, I claim, is that the Korsakoff subject has amnesia. Now, Robins objects that, even though Korsakoff patients typically have amnesia for events which happened after the development of their disorder, they do not need to have amnesia for events which happened before the development of their disorder. That is, they do not need to lose memories formed prior to that point. Thus, Robins suggests, we can imagine a case of confabulation which involves the report of a conversation in a train happening before the development of the subject's disorder. And the functionalist account, Robins tells us, will not be able to rule out such a confabulation case as a case of remembering since, in that case, the mental image that the subject is having will satisfy the required functional role. With regards to this objection, my response is, in a way, conciliatory and, in a way, adversarial.

Let us take the adversarial part of the response first. I would like to cast doubt on the idea that, in the confabulation case imagined by Robins, the subject's image does satisfy the required functional role. Is it the case that, had the subject been in a train conversing with someone, they would now have the type of mental image that they are actually having? If the possible conversation happens before the development of the subject's disorder, the answer is 'yes'. But if the possible conversation happens after it, then the answer is 'no'. It seems, then, that the mental image that the subject is having plays the functional role required for the subject to remember that they had a conversation in a train *at such-and-such time* (a time which is, in fact, previous to the development of their disorder, whether they can identify it as such or not), but it may not play the functional role required for the subject to remember, more simply, that they had a conversation in a train.

But isn't this all Robins needs in order to make her point? After all, Robins's point was that a patient with Korsakoff's syndrome, and a particular form of amnesia, could still count as remembering, which seemed to conflict with the functionalist approach to memory. And I am granting that, in the case envisaged by Robins, the patient has a mental image which has the required functional role for the fact that they had a conversation in a train at such-and-such time (a time which is, in fact, previous to the

development of their disorder). One might think, therefore, that this case does present a challenge to the functionalist framework, since we are granting that, depending on the particular details of the patient's amnesia, the case could indeed count as a case of remembering.

The conciliatory part of the response is to concede that not all cases of confabulation need to be ruled out as cases of remembering by the advocate of the functionalist approach. If there are cases in which the subject is confabulating when they have their mental image, but their mental image does satisfy the necessary functional role, then I have trouble seeing why we should rule them out as cases of remembering. To be clear, my reason for ruling out Confabulation as a case of remembering was not that the case involves confabulation ('Confabulation' then, may have been a misnomer), but that the case involves the type of amnesia which interferes with the mental image satisfying the functional role which is required for remembering. If there is a type of amnesia which is consistent both with confabulation and with the mental image that the subject has satisfying the necessary functional role for remembering, then I am willing to accept that some cases of confabulation are cases of remembering. At the very least, we should ask what independent grounds we have for assuming that memory and confabulation need to be incompatible.

Robins's (2021) second general worry is that, since having mental image *i* is an occurrent mental state, the appeal to functional roles cannot capture our intuitions about what it is to remember that *p* in virtue of having *i*. For functional roles concern causal relations in which types of mental states are involved, and what we want to know, when we ask what it is for a subject to remember that *p* in virtue of having mental image *i*, is what it is for the subject, in this particular instance of their having *i*, to remember that *p*. Now, it seems to me that if a specific mental state, which is occupied by a specific subject at a specific time, is a particular instance of a mental type, such as remembering that *p*, then the subject qualifies as remembering that *p* at the relevant time. For that reason, it seems to me that if we want to know, for a specific, occurrent mental state, whether it counts as remembering that *p*, it is a legitimate question to ask whether the mental state is of a certain type, namely, remembering that *p*. The type of mental state will provide us with the conditions that the occurrent mental state needs to satisfy. I do not see any confusion here.

According to Robins, the causal theorist of memory, for example, and myself are talking past each other, since the causal theorist is concerned with token mental states whereas I am concerned with types of mental states. I disagree. Both of us are concerned with token mental states. I am just answering the question of whether the token mental state in question is a state of remembering by considering whether it belongs to a certain type, a type which requires certain conditions. Whether the causal theorist is doing the same thing or not will depend on their approach to causation. If they regard the kind of causation involved in memory as token causation, then they will answer the question in a different way. But if they do not regard it as an instance of token

causation, they will also consider whether the token mental state under consideration belongs to a certain type. It is just that the conditions required to belong to that type will be different from mine (they may involve probabilities, processes, or other factors, but they will be, in any case, backward-looking conditions).

Let us keep in mind that the functionalist account of remembering is not meant to be an account of remembering correctly. It is meant to be an account of remembering simpliciter. Thus, the thought is that, when we ask what it is for a subject to remember that *p* in virtue of having mental image *i*, we are asking a question about the particular mental image *i*, but the answer depends on facts about the type of mental state to which mental image *i* belongs. By contrast, when we ask whether the subject is remembering *p* correctly in virtue of having mental image *i*, we are also asking a question about the particular mental image *i*, and facts about the type of mental state to which mental image *i* belongs will not provide us with an answer to that question. If what the causal theorist and the narrative theorist are after is an account of remembering correctly, then the functionalist approach will not help them to build such an account. I agree with Robins on that point.

4. Concluding remarks

There is an interesting common factor in some of the objections raised by James (2021), Viera (2021) and Robins (2021). The common factor is that all three commentators are targeting, not tangential aspects of, but fundamental assumptions in the self-referential account of memory proposed in the book. Some of James's objections, for example, hinge on whether memories can be reconstructed or not. Likewise, Viera's objections depend on the issue of how intentionalist explanations of phenomenology work, and whether the components of the content of a mental state can account for what it is like for one to be in that mental state. These issues are not discussed in the book. They are mentioned, in the first chapter, but they are only mentioned as assumptions. There is a sense, then, in which all three commentators are focusing on features which are essential to the self-referential account of memory offered in the book. I do not believe that the account can therefore be modified in order to address the relevant concerns. Instead, the account becomes a different account of memory if some of those non-negotiable assumptions are dropped. However, I hope that I have made a cogent enough case for the idea that, on reflection, dropping the relevant assumptions will ultimately not be necessary.

References

- Fernández, J. (2007). Memory and time. *Philosophical Studies*, 141, 333-356. <https://doi.org/10.1007/s11098-007-9177-x>
- Fernández, J. (2019). *Memory: A Self-Referential Account*. Oxford University Press. <https://doi.org/10.1093/oso/9780190073008.001.0001>
- James, S. (2021). Immunity to error through misidentification and the functionalist, self-reflexive account of episodic memory. *Estudios de Filosofía*, 64, 189-200. <https://doi.org/10.17533/udea.ef.n64a10>
- Robins, S. (2021). The failures of functionalism (for memory). *Estudios de Filosofía*, 64, 201-222. <https://doi.org/10.17533/udea.ef.n64a11>
- Searle, J. (1983). *Intentionality*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139173452>
- Viera, G. (2021). Feeling the past: beyond causal content. *Estudios de Filosofía*, 64, 173-188. <https://doi.org/10.17533/udea.ef.n64a09>