



SIMPOSIO

# The failures of functionalism (for memory)\*

*Sarah Robins*

University of Kansas, Kansas, United States  
E-mail: [skrobins@ku.edu](mailto:skrobins@ku.edu)

Recibido: 3 de febrero de 2021 | Aceptado: 24 de mayo de 2021  
<https://doi.org/10.17533/udea.ef.n64a11>

**Abstract:** In *Memory: A Self-Referential Account*, Fernández offers a functionalist account of the metaphysics of memory, which is portrayed as presenting significant advantages over causal and narrative theories of memory. In this paper, I present a series of challenges for Fernández's functionalism. There are issues with both the particulars of the account and the use of functionalism more generally. First, in characterizing the mnemonic role of episodic remembering, Fernández fails to make clear how the mental image type that plays this role should be identified. Second, I argue that a functionalist approach, which appeals to the overall structure of the memory system and tendencies of mental state types, is ill-suited to the metaphysical question about episodic remembering that is of interest to the causal and narrative theorists with which Fernández engages. Fernández's self-referential account of memory has many other virtues, but functionalism is a poor fit for episodic remembering.

**Key words:** memory, functionalism, episodic memory, mental image

\* The research and writing of this paper was supported in part by an award from the General Research Fund at the University of Kansas (award #2145081).

## Cómo citar este artículo

Robins, S. (2021). The failures of functionalism (for memory). *Estudios de Filosofía*, 64, 201-222.  
<https://doi.org/10.17533/udea.ef.n64a11>

OPEN  ACCESS





ARTÍCULO  
DE REFLEXIÓN

# Los fallos del funcionalismo (para la memoria)

**Resumen:** En *Memory: A Self-Referential Account*, Fernández ofrece una explicación funcionalista de la metafísica de la memoria, que se dice tener ventajas significativas sobre las teorías causales y narrativas de la memoria. En este artículo presento una serie de desafíos para el funcionalismo de Fernández. Hay problemas tanto con los detalles del relato como con el uso del funcionalismo en general. En primer lugar, al caracterizar el papel mnemónico del recuerdo episódico, Fernández no aclara cómo debe identificarse el tipo de imagen mental que desempeña este papel. En segundo lugar, sostengo que un enfoque funcionalista, que apela a la estructura general del sistema de memoria y las tendencias de los tipos de estados mentales, no se adapta a la pregunta metafísica sobre el recuerdo episódico que es de interés para los teóricos causales y narrativos con los que Fernández discute. La descripción autorreferencial de la memoria de Fernández tiene muchas otras virtudes, pero el funcionalismo no encaja bien con el recuerdo episódico.

**Palabras claves:** memoria, funcionalismo, memoria episódica, imagen mental

## Sarah Robins

is an Associate Professor of Philosophy at the University of Kansas and an affiliate of the Cognitive and Brain Sciences Ph.D. program in KU's Psychology Department. Her research is focused primarily on memory, with an emphasis on memory traces and the ways they are invoked and investigated in philosophy, psychology, and neuroscience.

**ORCID:** 0000-0003-2439-4633



## 1. Introduction

Fernández's *Memory: A Self-Referential Account* is ambitious, tackling questions about the metaphysics, intentionality, phenomenology, and epistemology of memory simultaneously. The result is a bold and broad theory of remembering, full of arguments and ideas that will no doubt influence the philosophy of memory for years to come. Fernández has built the account so that its central components are relatively free-standing—he claims that one can, for example, accept his proposed view of memory content while rejecting the account of its metaphysics, or vice versa. This design allows me to focus my critique on the metaphysical account of memory that Fernández proposes, specifically the functionalist view he sets out in chapter two.

Fernández's functionalism characterizes experiences of episodic remembering in terms of the mnemonic role played by their mental images. When a mental image fulfills this mnemonic role, the subject is remembering. In this paper, I present a number of challenges to this functionalist proposal. First, I argue that there are a number of difficulties with the particulars of Fernández's functionalism, which cause problems for the mnemonic role as he's characterized it. Second, I argue that there are broader problems for the attempt to apply a functionalist framework to episodic remembering. Functionalism offers a characterization of mental state types and allows for assessments of remembering across individuals, in terms of whether or not they possess that mental state type. While on the surface they may look similar, these concerns are importantly distinct from those at issue for causal and narrative theorists, who are proposing conditions on token states of remembering that will make possible assessments of remembering within individuals. Ultimately, I conclude, functionalism fails to fit the explanatory demands of episodic remembering.

## 2. Fernández's functionalism

Fernández's account of memory is concentrated on the intersection between memory for facts and memory for perceptual experiences. Focusing on memory for facts, or propositional memory, is fairly common amongst philosophers offering accounts of memory (e.g., Bernecker, 2010). Fernández's (2019) approach is unique because he narrows the scope of his account to a subset of propositional memory: "memory for facts involving objects perceivable through sensory modalities" (p. 5). In the book, he focuses specifically on vision and visible objects—a memory of the fact that the keys were left on the counter, or a memory of the fact that Mary attended the party.

Fernández then narrows his focus further. His interest is in how facts about perceptible objects are remembered *episodically*. This involves appeal to the familiar distinction between episodic and semantic memory, but Fernández puts it to a particular use. The distinction between episodic and semantic memory is often cast as

one between memory for experiences and memory for facts, respectively. Fernández is already calling attention to ways that distinction can be blurred with his interest in memory for facts based in perceptual experience. He uses the distinction between episodic semantic to characterize forms of *remembering*, the activity of entertaining a memory. Episodic remembering involves experiences; semantic remembering involves beliefs. Even for memories of facts based in perceptual experience, both forms of remembering are possible. My remembering, now, that they keys are on the counter could be the activation of a belief that I formed when I saw the keys on the counter earlier in the day. My remembering now could also be experiential: having just noticed the keys are not with me, I may visualize the rooms of my house, use my attention to scan various objects lying around, and in so doing, locate my keys on the counter. This latter form is remembering episodically.

Episodic remembering, for Fernández (2019), is experiential: “when you remember [episodically], you are having an experience; an experience that typically originates in a past perceptual experience of the fact” (p. 6). Episodic remembering, like other experiential mental states, has phenomenal and intentional features. Fernández discusses these features later in the book. First, he addresses a complementary metaphysical question about episodic memory: What are the conditions under which an experience qualifies as an episodic memory?<sup>1</sup> The candidate experiences are mental images –the image of keys on the counter or the image of Mary amongst the party attendees. But this alone does not answer the question. Lots of mental states involve mental imagery, not just remembering. What identifies the mental images of remembering uniquely?

Answering this question is difficult, in part, because of the diachronic nature of memory. Episodic remembering is a mental activity *now* but which is about and, in some sense, due to a previous experience. The metaphysical conditions on episodic remembering must therefore include consideration of the past and the present. In other words, the metaphysics of memory involves saying, of a current mental image, something about where it came from *and* what is being done with it now. Fernández’s discussion begins with discussion of two metaphysical views –the causal theory of memory and the narrative theory of memory. Each of these alternative views emphasizes one aspect of memory’s diachronic nature. Causal accounts are focused on the connection to the past event; narrative accounts are focused on the role of the state in one’s present mental life. Fernández argues that these approaches, while distinct, have parallel flaws: each over-emphasizes its selected dimension and neglects the importance of the other. Fernández thus offers functionalism, a theoretical approach that appeals to

1 Fernández (2019) characterizes his view as a hybrid approach that pursues metaphysical and intentional features of memory to be equally fundamental. He of course views the account he develops of the two approaches to be complementary, but acknowledges that the two are meant to function independently (p. 24).

both causes *and* effects, as an account better suited to accommodating both aspects of the diachronic nature of episodic remembering. He then demonstrates its superiority by showing how functionalism can remedy the deficiencies of each alternative, better addressing the cases that were problematic for both causal and narrative theories.

In what follows, I provide a brief summary of Fernández's characterization of each of the alternative theories and the cases that present problems for each. These cases play a critical role in Fernández's own account, as functionalism's key strength is its alleged ability to offer better responses to them. In discussing these cases, I use the same examples as Fernández does in his original presentation, so as to best ensure that the relevant features are preserved.

## 2.1 Causal theories of memory

Causal theories of memory privilege the connection between the episodic remembering experience and the past event (e.g., Bernecker, 2010; Debus, 2010; Martin & Deutscher, 1966). More specifically, they require a causal connection between the event and the subsequent experience. Versions of the causal theory differ over how they identify the particular kind of causal connection that is required. What all versions share, Fernández argues, is too rigid a focus on the past event and its influence on remembering. This leads the causal theory to omit some cases of remembering where the connection to the past is weaker, and to include cases that have this connection even when they lack features that are intuitively essential for remembering. Cases of these sorts Fernández labels *embellishment* and *epistemic irrelevance*, respectively.

*Embellishment cases* are ones where the content involved in the experience of remembering goes beyond what was available in the previous experience. Fernández uses the example of a person hunting with their father who, on one occasion, sees their father shoot a white rabbit. Later, the person visualizes their father having shot a black rabbit. Despite the change in content, the case should still count as a case of remembering, albeit one that involves *misremembering*.<sup>2</sup>

*Epistemic irrelevance cases* involve a person generating a mental image that derives from a past experience, but that they do not recognize as such. Fernández uses the example of a person who is painting, and draws a bird sitting atop a house. Unbeknownst to the painter, this scene is one they saw as a child. Since the painter is unaware of this connection to his past, he does not engage with the mental image (or the painting) in ways characteristic of remembering. He does not, for example,

2 Here it is important to note that this way of classifying states makes sense on Fernández's view because remembering is not a factive state. For Fernández, it is possible to be remembering and do so incorrectly.

believe that the scene depicted is one that he previously experienced. Since the image does have a connection to his past experience, causal theories admit such cases. But such cases lack several important features –the experiential feeling of remembering, an associated belief that the image derives from past experience, etc.– which should lead us to preclude such epistemic irrelevance cases, Fernández argues.

## 2.2 Narrative theories of memory

Narrative theories of memory, in contrast, focus on how experiences of episodic remembering are integrated into a person's sense of their life and its temporally outstretched narrative (e.g., Brockmeier, 2015; Goldie, 2012; Schectman, 1994). The emphasis is on the role the mental image plays in the person's sense of their self and their life, not on the connection to the past. The view is thus well-suited to accommodate the embellishment cases that causal theories neglected, while also requiring epistemic relevance that would block cases like the painter. The view encounters troublesome cases of its own, however, because of its overemphasis on integration and because of its failure to require a connection to the past. Narrative theories are thus susceptible to cases of *isolation* and *confabulation*.

*Isolation cases* are ones where a person has a vivid mental image of a past experience, that they recognize as a memory and believe to be a depiction of something that previously happened to them, but where the person is simultaneously unable to situate the experience at any particular point in the past. Fernández's example involves falling into a pool at some point during childhood. The person in the example vividly recalls the experience, but cannot remember when this occurred, where the pool was, who was there, etc. The person's failure to integrate this experience with his broader life narrative precludes it from the class of rememberings for supporters the narrative theory. Even if such memories are isolated, Fernández argues, they should still be included.

*Confabulation cases* are the inverse of epistemic irrelevance cases. They occur when a person has a mental image that is treated as a memory, and well-integrated into the person's understanding of their prior experience and life narrative –despite the fact that the mental image does not derive from a past experience. Confabulation often occurs in clinical cases of psychiatric disorder, and indeed, Fernández's example of such a case involves a patient with Korsakoff's syndrome who, despite having amnesia and being hospitalized, claims to have had a conversation on the train while traveling over the weekend. Such a case meets the narrative theory's requirements for integration, but lacks any connection to the past experience. Narrative theories thus allow such cases to count as instances of remembering, but they should not.

## 2.3 Memory functionalism

Fernández's presentation of causal and narrative theories of memory, and their respective limitations, provides an elegant setup for generating a new account of the metaphysics of memory. Each account has strengths and weaknesses, which complement one another. The aim is for a middle-ground account that can maintain both sets of strengths while avoiding the accompanying weaknesses. An adequate account of what is required for a mental image to qualify as an experience of remembering episodically must include both a connection to the past and integration into one's present life, while at the same time stopping short of demanding complete fealty to the past or wholesale enmeshing into one's life narrative.

Functionalism, Fernández argues, can provide such an account. Functionalist analyses are common amongst accounts of mental states, providing an important precedent for extending this framework to memory. Functionalism involves a characterization of the mental state in question in terms of its functional role, generally understood as a specification of the state's typical causes and typical effects. By appealing to both causes and effects, functionalism meets the requirement of connecting to both the past and the present. Additionally, the appeal to *typical* causes and effects gives the account flexibility, helping to ensure that neither connection is interpreted rigidly enough to cause the problems that were shown above for causal and narrative accounts.

Episodically remembering a fact about one's past perceptual experience is, for Fernández, a matter of having a mental image that plays the right functional role. He characterizes the requisite mnemonic role as follows:

*S* remembers that *p* just in case *S* has some mental image *i* such that *i* tends to cause in *S* a disposition to believe both that *p* and that *S* experienced that *p*, and *i* tends to be caused in *S* by having experienced that *p* (Fernández, 2019, p. 49).<sup>3</sup>

The account is centered upon the mnemonic role of a mental image, *i*. The account includes a connection to the past and a connection to the present. To play the requisite role, *i* has to have certain causes *and* certain effects. These relations to the past and present are, however, framed as tendencies, allowing minor aberrations in a way that helps in handling trickier cases like embellishment and isolation.

Fernández argues that these features allow functionalism to retain the successful qualities of both causal and narrative accounts, while also addressing the cases that caused trouble for these alternatives. Functionalism can, he claims, accommodate embellishment and isolation, while excluding epistemic irrelevance and confabulation. The required connection to the past prevents confabulation; the required connection to the present prevents epistemic irrelevance. Construing both required connections

<sup>3</sup> Fernández (2018) offers an initial account of functionalism.

as tendencies leaves enough wiggle room to allow embellishment and isolation into the account of remembering. Or so Fernández argues.

Before turning to an evaluation of Fernández's functionalist proposal, it is worth pausing to reflect on its significance. Fernández's functionalism brings a novel perspective to the metaphysics of memory, re-energizing a debate where the number of viable positions has long been limited. By promoting functionalism, Fernández also encourages philosophers interested in the metaphysics of memory to engage with material available in the metaphysics of mind more broadly. Functionalism also offers a few particular advantages. First, by characterizing the mental state of remembering in terms of its mnemonic role, Fernández is able to sidestep murky debates over the mental content of episodic memory. Second, Fernández puts the requirement of connection to the present on equal footing with the requirement of connection to the past. This is rare amongst accounts of memory, which are understandably focused on the past as the source from which remembering derives. In so doing, he helps to highlight the importance of a memory's relevance in a person's cognitive and epistemic activities at the time of remembering that past-directed accounts often overlook.<sup>4</sup>

Having laid out Fernández's functionalism and its alleged advantages over alternative accounts, I now turn to critiquing the view. My critique comes in two forms, addressed in the following two sections. First, I focus on the particulars of Fernández's account –whether the mnemonic role for memory images, as formulated, can do the work he intends for it. Second, I introduce a more general set of concerns about the use of functionalism to provide a metaphysics of memory. I argue that the structure and strengths of functionalism are ill-suited to the concerns at issue for the causal and narrative theorists with which Fernandez is engaged.

### 3. Mental images and the mnemonic role

When looking into the details of Fernández's view, my concerns are focused on the mnemonic role he sets out for mental images –specifically, how to individuate the mental state type *i* that's meant to play this role. Fernández's defense of functionalism is brief: he offers a characterization of the mnemonic role, quoted above, and some remarks about how this approach can improve upon the deficiencies of the causal and narrative approaches. His presentation does not, however, involve walking through any of the examples in detail, explicating how each variable in the mnemonic role is filled and how its tendencies are evaluated. This is understandable, given how many aspects of memory Fernández addresses in this book-length account. Taking the time

<sup>4</sup> Although see Debus (2010) as an example of a causal theorist who has incorporated concerns about relevance.



to go through some examples in detail, as I do below, raises questions that need to be addressed before the account can deliver on its promised advantages.

Let's start with a straightforward case, derived from one of Fernández's own examples: episodically remembering the fact that Mary was at the party last month. In this case, *S* has an experience of a mental image –presumably, an image of the party, with Mary amongst the attendees. What is required for this experience to qualify as one of episodic remembering? To answer the question, Fernández's functionalism asks us to consider *i*, the mental image of Mary at the party, and its role in *S*'s mental life. What tends to cause this image and what effects does it tend to produce?

In order to evaluate *i*'s tendencies, we need to situate this token within its broader mental state type. Fernández does not tell us how this is to be done, but his view contains material that provides some suggestions. It seems clear that the mental state type in question will be *subject-relative*. The mnemonic role is identified through *i*'s tendencies *in S*. This makes sense. These are images of experiences, from a visual perspective occupied by *S*, not anyone else. People differ not only in the perspectives they have on any particular fact in the world, but more broadly in terms of when and where they live, and so, which mental images they have and what they tend to cause and effect. I have a mental image of my first day of college. My son, who was not alive at the time, does not. The mental image types that operate as *i* for me will be different than the ones that operate for my son, which in turn will differ from yours and everyone else's.

Even once the relevant mental image type has been restricted to a given subject *S*, there are still multiple ways it could be characterized. It could be the set of all occurrences of *i* –the collection of all the times the mental image has been tokened in *S*'s experiences. Take the case of remembering that Mary was at the party. Suppose the first time *S* tokened this mental image it was as part of visualizing the party scene to determine whether Mary was in attendance. Subsequently, however, this is the mental image that comes to mind whenever Mary's name comes up in conversation or whenever *S* wonders how Mary is doing. This party may be the last time *S* saw Mary, and as more time passes *S* begins to wonder what has happened to Mary and what has gone wrong with their friendship. The mental image type *i* now has several tokened instances, making it easier to evaluate its tendencies. The problem, at least for *i*'s mnemonic role, is that *i* doesn't have the tendencies required for episodic remembering. In the first instance, when *S* visualized the party to scan the image for Mary, this seemed like an instance of remembering. But *i* does not tend to produce the belief that *S* experienced Mary being at the party. It had this effect once, but now it is more likely to be involved in the production of other mental states –worries that Mary is upset with *S*, or a belief that Mary is avoiding *S*. To put the worry more generally: as the evaluation shifts to *i* as a mental state type, any tendencies that would have supported its mnemonic role are weakened.

Given this problem for this formulation of *i*, we could try another approach – instead of all of the ways *S* has tokened this mental image of Mary at the party, *i* might be all of the ways that *S* could do so. Fernández speaks in this way in some of the cases he considers, where the relevant roles are described as possible situations or what would have happened had *S* seen a particular thing other than what was actually seen. This modification provides the account with more flexibility, but it is unclear how to constrain the realm of possible cases of *S* visualizing *i*. Are we to assume that *S*, and her place in space-time, are fixed? Are we holding constant all of her activities, experiences, and beliefs up to this moment? If too many features change, then *i* becomes difficult to evaluate – *S* could fail to attend the party, or not know Mary, etc. The more of these factors that are open, the more flexibility the account gains.

What is needed is a way of explaining which possible tokenings of *i* are relevant, one that offers more than simply selecting a few cases we would like to consider for a particular evaluation. Fernández does not provide this. Regardless of which way it is specified, the flexibility it provides seems at odds with securing *i*'s mnemonic role. Once we pause to consider all of the possible ways a mental image of Mary at the party could be put to use by *S*, we become aware of just how many effects (and causes) this state could have. The wider the set of options available, the more the tendency toward the generation of any particular belief as an effect becomes increasingly small.

Neither of these candidates for *i* – the set of all of *S*'s experiences of visualizing *i*, the set of all ways *i* could be visualized by *S* – fit into the mnemonic role as Fernández has sketched it. Other alternatives may come from Fernández's discussion of how functionalism can better handle the cases that posed problems for causal and narrative accounts. Let's start, as Fernández does, with embellishment. Embellishment cases are ones where the content involved in the act of remembering goes beyond (i.e., provides more detail than or alternative detail to) what actually occurred. Fernández's example involves an *S* who, in the past, saw their father shoot a white rabbit. *S*'s mental image now, however, is of their father shooting a black rabbit. Fernández believes such cases should be counted as instances of remembering, albeit ones that involve misremembering. A functionalist account can accommodate such cases, Fernández argues, thanks to its reliance on tendencies rather than what actually occurred. We should thus be able to see how the mental image *i* is understood in this embellishment case so that it succeeds in filling the mnemonic role.

Recall Fernández's account of the mnemonic role for *i* involves tendencies in two directions: what tends to cause *i* and what effects *i* tends to produce. When considering the effects *i* tends to produce, Fernández appears to have a very specific image in mind – *S*'s an image of their father shooting a black rabbit. It fulfills this half of the requisite mnemonic role because

my mental image tends to cause in me the belief that I once saw a black rabbit being shot by my father, and it tends to cause in me the belief that my past perceptual experience was veridical; that the shooting did obtain in the past (Fernández, 2019, p. 51).

Producing the right effects requires that *i* involve the kind of details that would support belief about this particular past event as an experience that *S* has had.

When we turn to the second set of tendencies to evaluate for *i*—what tends to cause this mental image to come about—the *i* under consideration appears to change. Here is what Fernández says:

My mental image is the type of image that tends to be produced in me by past perceptual experiences of black rabbits being shot. To be sure, on this particular occasion, my mental image was not actually caused by a perceptual experience of a black rabbit being shot since, in the past, I did not have such an experience. Nevertheless, the fact remains that my faculties of perception and memory are related in such a way that perceptual experiences of black rabbits do produce in me the type of mental image that I am currently having. Had I seen, in other words, a black rabbit being shot in the past, this is the type of mental image I would be having now (Fernández, 2019, p. 51).

To generate a case where *i* has the tendencies necessary for this half of its mnemonic role, Fernández considers *i* as a generic image type: mental images of the type *S* tends to have when *S* sees black rabbits being shot. In the example, we are not given enough detail about *S* to know how many times this has happened for *S*, whether it is a common perceptual experience for *S* or indeed whether *S* has ever had this experience. Suppose that *S* has seen black rabbits being shot several times in the past. We might first wonder what type of mental image this would be: presumably, it is some generic image amalgamated from each of these experiences, with the details about the particular size of the rabbit, location of its wound, nearby ground cover, weather, etc. somehow smoothed out. However this goes, the *i* that emerges from the collection of all the times *S* has seen black rabbits will differ from the *i* that produces, in *S*, the belief that *S* had the past experience of their father shooting a black rabbit. Producing the belief required perceptual details about a particular experience; the image type associated with seeing black rabbits shot would seem to lack both these details and a connection to a particular experience.

The above quotation from Fernández also suggests a different reading of *i* and what tends to cause it. This alternative doesn't rely on *S* having had multiple experiences of seeing black rabbits shot in the past. Instead, it appeals to the general workings of and connections between *S*'s perceptual and memory systems to establish what would have happened if *S* had seen a black rabbit being shot.

Fernández does not elaborate on how this evaluation of perception and memory should go, but we can make a straightforward guess. *S* has a well-functioning visual system, so had she seen a black rabbit being shot, she would be likely to see it correctly. Further, *S*'s memory system is well connected to her visual system; her memory tends to store images of what she saw in this past. So had she seen a black rabbit, she would have been likely not only to see it correctly, but to retain a mental image of that experience. These judgments seem fine enough, and we can stipulate that they are true of the *S* in the example.

What remains unclear, however, is how consideration of this counterfactual gives us an evaluation of *i*'s tendencies that yields the desired conclusion. To secure *i*'s mnemonic role, it must be the case that *i* tends to be produced by *S*'s experience of *p* (here *p* = *S*'s father shooting a black rabbit). In actuality, *i* was produced by *S*'s father shooting a white rabbit, but evaluation of the counterfactual shows another way of producing *i*. In short, *i* came about in one way, but it could have come about another way. Considering these two cases together, what does that tell us about *i*'s tendencies? It is unclear to me how we are supposed to arrive at any conclusion about its tendencies, much less the conclusion that *i* tends to be produced by seeing a black rabbit being shot. And either way, this interpretation of *i* does not help with the problem previously identified about reconciling the senses of *i* used to determine its causes and its effects.

To put the point from embellishment cases succinctly: evaluating *i*'s mnemonic role involves assessing two tendencies –what tends to cause *i* and what *i* tends to cause. The understanding of *i* used to establish one side of its mnemonic role looks different than the understanding of *i* used to establish the other. In order for *i* to produce the right beliefs in *S*, *i* has to be very specific. This is the only way for *i* to tend to cause beliefs that a particular perceptual experience occurred. But in order for *i* to be caused by the right sort of experience, it needs to be generic. There does not appear to be a stable conception of *i* that can play both roles.

The difficulty in providing a stable conception of *i* across both sides of its mnemonic role creates further problems for Fernández's treatment of confabulation cases –and the ability of his functionalism to keep cases of embellishment and confabulation separate.

Confabulation cases are ones where a mental image is integrated into a person's thoughts and actions as if it were a memory, but where the image has no connection to the person's past experience. Fernández uses the example of a person with Korsakoff's syndrome, a memory disorder induced by extreme thiamine-deficiency, often as a result of heavy alcohol use. In the example, the Korsakoff's patient is hospitalized, and has been for some time, but when asked about their activities on the prior day, the patient generates an elaborate account of going on a trip and having an extended conversation on the train. The patient has a mental image *i* of

conversing with a fellow passenger on a train while taking a trip. This image fulfills one half of the requisite mnemonic role: the image tends to produce beliefs in the patient that they were on a train over the weekend, having a conversation with a fellow passenger. This tendency led to confabulation cases being accepted by narrative theories of memory. Fernández argues, as many memory theorists do, that cases of confabulation should be excluded from an account of remembering. In order to exclude such a case, the operative question is whether “the patient’s mental image of the conversation in the train is of a kind which does not tend to be produced, in that patient, by experiences of such conversations” (Fernández, 2019, p. 52).

Fernández believes that the answer to this question is no. The Korsakoff’s patient has amnesia. As such, they lack the kinds of tendencies required. For them, mental images active during perception do not tend to be stored and converted into memories. Without such a connection, *i* fails to fill the mnemonic role, and cases of confabulation fail to count as instances of remembering.

At first glance, this judgment on confabulation cases looks consistent with Fernández’s treatment of embellishment cases, as discussed above. That is, Fernández is using an assessment of *S*’s faculties of perception and memory to determine whether *i* tends to be caused in the right way. Embellishment and confabulation both involve errors: the mental images offer an incorrect depiction of *S*’s past experience in both cases. Despite this similarity, embellishment cases count as remembering while confabulation cases do not because of the broader tendencies that exist in the cognitive systems from which they are generated. Embellishment cases count as (mis)remembering because they emerge from functional perceptual and memory systems; confabulation cases fail to count because they emerge from dysfunctional perceptual and memory systems.

This approach to confabulation is difficult to reconcile with the literature on Korsakoff’s and other forms of amnesia. Korsakoff’s patients, almost universally, have *anterograde* amnesia –the onset of the disorder is marked by the inability to form and store new memories (Fama, Patel & Sullivan, 2012). There are significant differences amongst persons with Korsakoff’s, however, on the extent of additional memory damage. There is a tendency for at least some retrograde amnesia (Kopelman, Thomson, Guerrini & Marshall, 2009), but the extent of a person’s inability to remember past events differs across individuals. Fernandez responds to all persons with Korsakoff’s in the same way, dismissing them from possible remembering because of their lack of properly functioning perceptual and memory systems. Assessment of persons with Korsakoff’s, however, and the determination of how much of their memory systems are or are not functioning proceeds individually. Persons who are diagnosed with Korsakoff’s often present with confabulations about recent events. Spiegel and Lim (2011) describe an individual who reported that he had just arrived on a flight from out of state (when in fact he was in the

emergency room at the hospital, as a result of falling from a balcony) and that Barack Obama, the US President at the time, had died. From initial confabulations such as these, doctors go on to administer additional tests, in order to determine whether or not the patient's retrograde memory systems are also damaged, and if so, how severely.

It is thus consistent with a Korsakoff's diagnosis for a person to retain memories from the past, maintaining some functional connections between perception and memory. Acknowledging this complicates the assessment of whether any particular mental image produced by a person with Korsakoff's is an instance of confabulation. In the case Fernández describes, the patient has a mental image of a recent event, one that would have occurred after the development of the disorder and so would not have the tendency to be the result of a past perceptual experience. But suppose this patient had (prior to developing Korsakoff's) previously traveled a lot, often doing so by train, and on some occasions, talked with fellow passengers. In this case, the patient may very well have a mental image *i* that has the tendencies needed in order to play the mnemonic role. That is, the mental image of talking to other travelers on a train is one that tends to be caused in *S* by having had this experience. In this particular instance, *i* was not caused in this way –but the appeal to tendencies is meant to provide a way to step back from what actually happened when making our evaluation. For many confabulation cases, then, extending the evaluation of *i* and its mnemonic role that was used to assess embellishment cases, will yield the conclusion that confabulations are instances of remembering.

Across all of these cases, what matters for determining whether it counts as remembering is how the mental image *i* is characterized. To secure its role in acquainting the rememberer with a fact about past perception, there is pressure for *i* to be highly specific. To evaluate its broader tendencies, allowing for hiccups in the standard process of remembering, there is pressure in the opposite direction, for *i* to be highly general. Fernández does not offer much in way of elaboration on *i*. What he does provide pulls in both directions, as illustrated above. Is there a way for *i* to retain its connection to a particular past experience while still allowing a way for its tendencies to be evaluated? Moreover, can this formulation of *i* allow in cases of embellishment while keeping out cases of confabulation?

Fernández, no doubt, has more to say about these cases and the resources within functionalism for responding to the challenges presented here. It is possible that there are ways to modify the mnemonic role, or the understanding of *i* to address these issues. Whether functionalism can withstand these challenges and make good on its promises as a metaphysics of memory is thus yet to be determined. For now, I will set these concerns about the details of functionalism aside and turn to a broader critique of the use of a functionalist framework to account for episodic remembering.

## 4. Functionalism's failure of fit

In the previous section, I challenged Fernández's account of the mnemonic role for episodic remembering, arguing that his characterization lacked the detail required to sort through particular cases. Viewed from another angle, this alleged flaw is actually a feature, one that is characteristic of functionalist accounts of mental states. Fernández's functional characterization of episodic remembering is highly similar to functionalist characterizations of other mental states, like belief (e.g., Leitgeb, 2017) and desire (e.g., Alvarez, 2017). These general sketches of the role played by belief and desire do the work asked of them; theorists recognize the full specification of any particular belief or desire will be longer and more complicated.<sup>5</sup> In other words, I have been criticizing Fernández's functionalist view for being functionalist. This could indicate that my objections to the view are off base. Instead, I think this observation opens the door to a distinct way of evaluating Fernández's project, one that highlights the tension between the aims and interests of functionalism and those that have standardly been the focus for other philosophers of memory, particularly those endorsing causal and narrative views.

As an account of the metaphysics of remembering, Fernández's functionalism is an answer to the following question: *what are the conditions under which an experience qualifies as an episodic memory?* On the surface, this looks to be the same question asked (and answered) by causal and narrative theorists. In the details, however, they are importantly different. Each asks the question in a way that invokes a distinct contrast class, placing divergent constraints on what counts as an adequate answer.

When Fernández asks about the conditions on episodic remembering, he is looking for a way to distinguish episodic memory—as a mental state type—from other mental state types. As he says at the outset of the chapter, his aim is “to determine what it is to remember something, as opposed to imagining it, perceiving it, or introspecting it” (Fernández, 2019, p. 32). Causal and narrative theorists, in contrast, are interested in exploring these conditions as a way of distinguishing remembering from borderline and/or degenerate cases. In other words, they are asking: amongst the set of mental states that are candidates for remembering, which ones are successful, genuine, or real? Some theorists approach this question by treating the sought-after category of successful remembering as factive. The aim is then to identify all and only the cases of apparent remembering in which the past event is accurately represented. The criterion need not be accuracy, though, as illustrated by narrative accounts that appeal to integration into one's life story as critical for genuine memory. Regardless of how successful cases are defined, the conditions used to establish them are ones that allow sorting between token mental states.

<sup>5</sup> I am grateful to an anonymous reviewer for raising this point.

There are of course broad similarities between Fernández's question and the question asked by causal and narrative theorists. They're all discussing the conditions on episodic remembering, and moreover, they are all engaged with this question as a metaphysical project. The question is about the conditions that must obtain for remembering, not one of how we could know whether they are operative in a particular case or not. The accounts differ in terms of the scope at which the question is pitched—are the conditions placed on a mental state type or on particular token states? This difference in scope has consequences for how the question is explored and answered; putting the two into direct conversation, as Fernández's does here, leads to confusion and to problems.

The tension is apparent in Fernández's account. Although he frames the chapter's question as one of comparing remembering to other mental state types, the account he provides does not answer that question. Nothing in Chapter 2 addresses how memory differs from perception, imagination, or other imagery-involving mental states. Instead, his account develops out of engagement with causal and narrative approaches, where he uses functionalism to address borderline cases of embellishment, isolation, confabulation, and epistemic irrelevance. Despite focusing on these cases, which for causal and narrative theorists require sorting amongst tokens to determine which cases are successful or genuine, Fernández's maintains his approach based in remembering as a mental state type. As such, the answers he gives are ill-suited to the questions these cases raise, at least as they are standardly understood by causal and narrative theorists. Given this mismatch, Fernández's account fails to answer either form of the question about the conditions on episodic remembering.

To illustrate this, I will start by restating Fernández's characterization of the mnemonic role of episodic remembering:

*S* remembers that *p* just in case *S* has some mental image *i* such that *i* tends to cause in *S* a disposition to believe both that *p* and that *S* experienced that *p*, and *i* tends to be caused in *S* by having experienced that *p* (Fernández, 2019, p. 49).<sup>6</sup>

Engaging with causal and narrative views, as he does in this chapter, requires Fernández's to put this mnemonic role to work evaluating particular cases. Fernández's considers four: embellishment, isolation, confabulation, and epistemic irrelevance. As discussed in Section 2, the goal is to develop an account that includes the first two (embellishment and isolation), while excluding the latter two (confabulation and epistemic irrelevance). When Fernández introduces these cases, he includes specific, detailed examples—a hunting trip where one's father shoots a rabbit (embellishment), falling into a swimming pool as a child (isolation). Generating cases in this way is done to mirror the features involved in cases generated by causal and narrative theorists.

<sup>6</sup> Fernández (2018) offers an initial account of functionalism.



When Fernández's responds to these cases, however, his answers are effectively restatements of the mnemonic role. In cases of embellishment and isolation, the right dispositions are formed and maintained; In cases of confabulation and epistemic irrelevance, they are not. Nothing in his discussion engages with nor hinges on the details of the cases under consideration.

Consider the embellishment case, where *S*'s father shoots a white rabbit, and *S* later (mis)remembers the experience as one where his father shot a black rabbit. In addressing this case, Fernández's is not concerned with what happened during the events in question to lead to the misremembering. There is no discussion of how *S*'s perceptual and memory systems were working during the particular events of this case. We do not know, either during the original shooting or subsequent recollection, whether *S* was paying attention or tired or distracted, whether the event was significant or traumatic for *S*, or anything about what occurred to *S* in the interim between the shooting and the recollection. Fernández's is not concerned with the overall distribution of coloration in rabbits where *S* lives, nor with the father's hunting habits, nor *S*'s range of hunting experience or time with his father. Instead, the question of whether this counts as a case of remembering is pitched as a question about the general operations of *S*'s memory faculty –i.e., whether *S*'s perceptions typically cause mental images that result in episodic remembering.

This shift in perspective, from the particular case at issue to more general tendencies of memory, is intentional. It is meant to be a virtue of Fernández's account. As he characterizes the key benefit of functionalism:

On both versions [role and realizer] of functionalism, what matters for whether a subject is having a mental state of some type is not the causal relations that actually hold between that state and other mental states of the subject, as well as the subjects perceptual inputs and behavioral outputs, but the causal relations that tend to hold between all of those states (Fernández, 2019, p. 48).

The appeal to tendencies is meant to give us a perspective on the capacity as a whole, and to cut off the concern that a memory system must work perfectly in order to work properly. What matters for episodic remembering is not whether a person actually remembers the particular event they take themselves to be remembering now, but whether, in general, they remember events from their past. Appealing to the mnemonic role to characterize episodic remembering allows Fernández's to absorb cases where perceptual experiences mischaracterize the world and cases where the contents of memory degrade or go missing. Fernandez makes use of this leniency to absorb the embellishment cases like the rabbit hunt, as well as isolation cases.

The tension between this appeal to tendencies and the cases under consideration becomes clearer in the treatment of the cases that Fernández's wants to exclude from his account of episodic remembering –cases of confabulation and epistemic irrelevance.

As with the others, he introduces each puzzle with a detailed case. For confabulation, he describes a Korsakoff's patient who creates a false memory of riding on a train. For epistemic irrelevance, he describes a painter who fails to recognize a scene he is painting as being from a previous experience. His responses to these cases do not engage with the event details. Instead, the cases are excluded by stipulating that the person involved in each case has a deficit that renders them unable to produce mental states that play the requisite mnemonic role. The Korsakoff's patient has amnesia that interferes with memory's storage of perceptual images, which leaves him unable to form the needed kind of input tendencies. The painter has a different kind of deficit, which prevents stored mental images from being activated and endorsed as past experiences. This deficit renders the painter incapable of exhibiting the requisite form of output tendencies.

Effectively, Fernández's is claiming that a person who lacks the ability to retain mental images and/or put them to use in subsequent experience cannot remember because they lack a critical feature of the basic capacity. This point is fairly straightforward, and it is not particularly controversial. I suspect that both causal and narrative theorists would agree, but would not consider these cases relevant to their concerns in establishing causal or narrative conditions on remembering. Fernández's treatment of cases draws the distinction between remembering and its absence *across* individuals. A person either possesses a mental state type that plays this mnemonic role or they do not. Causal and narrative theorists, in contrast, look to draw the distinction *within* an individual, wanting to allow that a person could episodically remember in one case and not in another.

I illustrate this with evidence from causal theorists. Martin & Deutscher's (1966) version of the painter case, which serves as inspiration for Fernández's epistemic irrelevance, does not involve the stipulation of a cognitive deficit in the painter. We are left to suppose that, in general, the painter has a working episodic memory and tends not only to retain mental images from past experiences, but to deploy those images subsequently in ways that he recognizes as deriving from past experience. That the painter has a few such mental images which he fails to recognize as memories seems perfectly ordinary. I suspect all of us have at least a few such cases where this has occurred. It is these kinds of cases, which occur within the mental life of an otherwise capable rememberer, that Martin and Deutscher want to include in their account of memory. Viewing epistemic irrelevance in this way, it becomes difficult to see how Fernandez could exclude such cases. This painter's error does not derive from a general deficit; it's merely an image that misfires in a capacity for remembering that otherwise has the requisite tendencies.

A similar point can be made for the confabulation case. Fernández's focuses on a clinical case of confabulation, where the memory error occurs as part of a structural deficit, which in turn is symptomatic of a broader psychiatric disorder. The Korsakoff patient has amnesia that precludes the retention and reactivation of perceptual images. Clinical confabulation cases are interesting, but many philosophers of memory, including causal theorists especially (Bernecker, 2017; Robins, 2019) have been interested in

forms of confabulation that occur in ‘everyday remembering’ –i.e., in persons who do not have a clinical diagnosis, whose memories are functioning well or at least normally. Non-clinical confabulations are demonstrated in experimental contexts; most notably, in paradigms developed by Elizabeth Loftus and colleagues that use suggestive interviews to implant information that participants later interpret as remembered events (e.g., Loftus & Pickrell, 1995). Similar cases of confabulation can also be found in Martin and Deutscher’s (1966) original discussion of the causal theory. Their view involves stipulating and then refining a causal condition on remembering. The causal constraint emerges out of consideration of the possibility of veridical confabulation –a case where one has an accurate representation of a past experience, but where the representation is brought about through some serendipitous string of events, not because of any connection to the past experience.

These are all cases where the person involved has a representation of a past event that has no connection to that past event. Importantly, though, there is no broader deficit suspected on the part of the person involved. Outside of the experimental paradigm, or without the convergence of a set of strangely serendipitous circumstances, the confabulation would not have occurred. More importantly, discussion of these cases proceeds against the backdrop assumption that the person’s memory is otherwise working normally. In fact, Loftus uses these cases as a demonstration of the ways in which those of us with properly functioning memories can be susceptible to confabulation and false memory (Loftus, 2003).

Given that these confabulations occur *within* a properly or normally functioning memory, it seems that Fernández would be compelled to include them. The overall tendencies that govern the mnemonic role are intact. This case of confabulation looks more akin to the case of embellishment. They’re both cases where the memory system goes beyond what it should in a particular case, but otherwise is working well. Similarly, the case of epistemic irrelevance looks much like the case of isolation. Both are cases where the usual tendencies are lacking or weaker, but as anomalous instances are not worrisome. In summary: insofar as confabulation and epistemic irrelevance are understood as causal and narrative theorists interpret them, Fernández’s view does not exclude them.

It’s possible that Fernández’s could develop this functionalist account further, in ways that could address these cases. Doing so, however, pushes against the nature of functionalism. The aim of the approach is to characterize the general role of a particular kind of mental state, and distinguish that role from the role of other mental states. Functionalism is not well suited to identifying or labeling a particular occurrent mental state as belonging to one type or another. It is not clear how functionalism could be used to classify a mental image. This was, in one sense, the criticism developed in the previous section.

To evaluate Fernández’s functionalism as an account of episodic remembering, it would be good to see the view in its most fitting context –i.e., as an answer to the question Fernández’s posed at the beginning of the chapter. How does episodic

remembering differ from perception, imagination, introspection, and the like? Fernández's does not answer this form of the question. The lack of an answer here is particularly worrisome for the view given the recent interest in *continuism* amongst philosophers of memory. Over the last two decades, philosophers and memory scientists have become increasingly interested in the relationship between episodic remembering and episodic imagining. There are now many who endorse the claim that remembering is a form of imagining (Hopkins, 2018; Michaelian, 2016) or that remembering and imagination are both forms of episodic simulation (Addis, 2018; 2020). These views are forms of *continuism*. While there are others who defend *discontinuism*, arguing for memory and imagination to be kept distinct (Perrin, 2016; Robins, 2021), *continuism* is by far the more popular view (see Michaelian, Klein, & Szpunar, 2016). Without a discussion of how his view relates to these mental state proposals, it is difficult to determine whether the mnemonic role has been fully and fairly articulated.

Functionalism offers a popular and prominent account of mental states, and Fernández's attempt to use its tools to address the metaphysics of memory is innovative. Ultimately, however, it appears to be unsuccessful. Functionalism is not suited for answering the metaphysical question that has been of interest to philosophers of memory, so it is wrong to characterize it as a direct competitor to causal and narrative views. The view may still be useful in answering a different question. It may offer a unique way of accounting for the mnemonic role of episodic remembering. Given the extent of recent work on the connections between remembering and imagination, however, the uniqueness of this mnemonic role cannot simply be stipulated.

## Conclusion

In this paper, I have presented a multi-dimensional critique of Fernández's functionalist account of the metaphysics of memory. I have argued that the mnemonic role, as he specifies it, leads to a number of challenges for the account and the judgments he wants to make about particular cases. I have also argued further that even if the mnemonic role is adjusted to address these problems, there are deeper problems with the use of functionalism to assess episodic remembering. Functionalism characterizes mental state types and is suited to assessments of episodic remembering *across* individuals, but the concerns of causal and narrative theorists involve token states of remembering and assessments *within* individuals.

Luckily for Fernández, who has structured his multifaceted account so that its various components are largely independent of one another, even if my assessment of this metaphysical proposal is correct, this does not threaten his account of the intentional and phenomenological aspects of episodic memory. Those aspects of the account, and the book, are exceptionally rich, and I look forward to future debates about how to craft a metaphysics of memory that suits them.

## References

- Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *Journal of the Royal Society of New Zealand*, 48, 64-88. <https://doi.org/10.1080/03036758.2018.1439071>
- Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology*, 11(2), 233-259. <https://doi.org/10.1007/s13164-020-00470-0>
- Alvarez, M. (2017). Desire, dispositions, and the explanation of action. In J. A. Deonna & F. Lauria (Eds.), *The Nature of Desire*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199370962.003.0005>
- Bernecker, S. (2010). *Memory: a philosophical study*. Oxford University Press. <https://doi.org/10.1093/obo/9780195396577-0072>
- Bernecker, S. (2017). A causal theory of mnemonic confabulation. *Frontiers in theoretical and philosophical psychology*, 8, 1207. <https://doi.org/10.3389/fpsyg.2017.01207>
- Brockmeier, J. (2015). *Beyond the archive: memory, Narrative, and the autobiographical process*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199861569.001.0001>
- Debus, D. (2010). Accounting for epistemic relevance: a new problem for the causal theory of memory. *American Philosophical Quarterly*, 47, 17-29.
- Fama, R., Patel, A. & Sullivan, E. V. (2012). Anterograde episodic memory in Korsakoff Syndrome. *Neuropsychology Review*, 22, 93-104. <https://doi.org/10.1007/s11065-012-9207-0>
- Fernández, J. (2018). The functional character of memory. In K. Michaelian, D. Debus & D. Perrin (Eds.), *New Directions in the Philosophy of Memory* (pp. 52-72). Routledge. <https://doi.org/10.4324/9781315159591-4>
- Fernández, J. (2019). *Memory: a self-referential account*. Oxford University Press. <https://doi.org/10.1093/oso/9780190073008.001.0001>
- Goldie, P. (2012). *The mess inside: narrative, emotion, and the mind*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199230730.001.0001>
- Hopkins R. (2018). Imagining the past: on the nature of episodic memory. In F. Macpherson & F. Dorsch (Eds.), *Perceptual Imagination and Perceptual Memory* (pp. 46-71). Oxford University Press
- Kopelman, M. D., Thomson, A. D., Guerrini, I. & Marshall, E. J. (2009). The Korsakoff syndrome: clinical aspects, psychology, and treatment. *Alcohol and Alcoholism*, 44, 148-154. <https://doi.org/10.1093/alcalc/agn118>
- Leitgeb, H. (2017). *The stability of belief*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198732631.001.0001>
- Loftus, E. F. (2003). Our changeable memories: legal and practical implications. *Nature Reviews: Neuroscience*, 4, 231-234. <https://doi.org/10.1038/nrn1054>

- Loftus, E. F. & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25, 720-725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Martin, C. B. & Deutscher, M. (1966). Remembering. *Philosophical Review*, 75, 161-196. <https://doi.org/10.2307/2183082>
- Mazzoni, G., Clark, A., De Bartolo, A., Guerrini, Ch., Nahouli, Z., Duzzi, D., De Marco, M., McGeown, W. & Venneri, A. (2019). Brain activation in highly superior autobiographical memory. *Cortex*, 1-36.
- Michaelian, K. (2016) *Mental time travel: episodic memory and our knowledge of the personal past*. MIT Press. <https://doi.org/10.7551/mitpress/10591.001.0001>
- Michaelian, K., Klein, S. B. & Szpunar, K. K. (Eds). (2016). *Seeing the future: theoretical perspectives on future-oriented mental time travel*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190241537.001.0001>
- Parker, E. S., Cahill, L. & McGaugh, J. L. (2006). A case of unusual autobiographical remembering. *Neurocase*, 12, 35-49. <https://doi.org/10.1080/13554790500473680>
- Perrin, D. (2016). Asymmetries in subjective time. In K. Michaelian, S. B. Klein & K. K. Szpunar (Eds), *Seeing the future: theoretical perspectives on future-oriented mental time travel* (pp. 39-61). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190241537.003.0003>
- Robins, S. K. (2020). Mnemonic confabulation. *Topoi*, 39, 121-132. <https://doi.org/10.1007/s11245-018-9613-x>
- Robins, S. K. (2020). Defending discontinuism, naturally. *Review of Philosophy and Psychology*, 11(2), 469-486. <https://doi.org/10.1007/s13164-020-00462-0>
- Schechtman, M. (1994). The truth about memory. *Philosophical Psychology*, 7, 3-18. <https://doi.org/10.1080/09515089408573107>
- Spiegel, D. R. & Lim, K. (2011). A case of probable Korsakoff's syndrome: a syndrome of frontal lobe and diencephalic structural pathogenesis and a comparison with medial temporal lobe dementias. *Innovative Clinical Neuroscience*, 8, 15-19.