



Episodic memory is not immune to error through misidentification: against Fernández

Kourken Michaelian¹

Received: 14 July 2019 / Accepted: 2 April 2020 / Published online: 20 April 2020
© Springer Nature B.V. 2020

Abstract

The claim that episodic memory is immune to error through misidentification enjoys continuing popularity in philosophy. Psychological research on observer memory—usually defined as occurring when one remembers an event from a point of view other than that from which he originally experienced it—would seem, on the face of it, to undermine the IEM claim. Relying on a certain view of memory content, Fernández (Synthese. <https://doi.org/10.1007/s11229-018-02050-3>, forthcoming), however, has provided an ingenious argument for the view that it does not. This paper reconstructs Fernández’ argument and shows that there is reason to reject the definition of observer memory and the view of memory content on which it relies. Once these are rejected, it turns out that observer memory does indeed imply that the IEM claim is false.

Keywords Episodic memory · Immunity to error through misidentification · Observer memory · Jordi Fernández

1 Introduction

Research on the reconstructive character of remembering has tended to lead psychologists to emphasize the ways in which memory—especially episodic memory, memory for experienced past events—can go wrong. While memory errors are sometimes characterized as byproducts of otherwise adaptive processes (Schacter 2019), the view that memories inevitably involve *some* degree of inaccuracy is widespread in psychology. Philosophers of memory have, to an extent, come to share this focus on memory errors (Bernecker 2017; De Brigard 2014; Fernández 2015; Michaelian 2016b, 2020; Robins 2016, 2019), but they remain markedly more optimistic than their colleagues in psy-

✉ Kourken Michaelian
kourken.michaelian@univ-grenoble-alpes.fr

¹ Centre de Philosophie de la Mémoire, Université Grenoble Alpes, Bâtiment CTL, 7 allée de Palestine, 38610 Gières, France

chology in that they tend to hold that, when remembering goes as well as it can, it is free of error. The most widespread view in philosophy is, indeed, that successful memories are *fully* accurate.¹

This optimistic view is reflected in the continuing popularity of the claim that episodic memory is *immune to error through misidentification* (Bermúdez 2012, 2013; Fernández 2014, 2019; Hamilton 2007; McCarroll 2018; Recanati 2007). Earlier defences of the IEM claim often focused on the challenge represented by quasi-memory (which is meant to be just like the memory, with the difference that it does not presuppose that the rememberer is identical to the subject whose experience is remembered; see Michaelian et al. forthcoming). More recent work has instead focused on the threat posed by *observer memory* (roughly, a form of memory in which one sees oneself in the remembered scene).² In one recent defence, Fernández (forthcoming)³ acknowledges that observer memory, which has been extensively studied in psychology (see Nigro and Neisser 1983; Rice 2010) and which has recently begun to receive attention in philosophy (Debus 2007; Sutton 2010; Eldridge 2014; Fernández 2015; McCarroll 2017, 2018; McCarroll and Sutton 2017), would seem, on the face of it, to undermine the IEM claim but goes on to provide an ingenious argument for the conclusion that ultimately it does not.

This paper argues that Fernández is wrong. The paper is structured as follows. Section 2 introduces the notion of error through misidentification as it applies to episodic memory. Section 3 introduces the notion of observer memory and explains why it seems to undermine the IEM claim. Section 4 begins to reconstruct Fernández' argument for the view that observer memory does not undermine the IEM claim, focusing on his definition of observer memory and suggesting that, whereas Fernández defines it in terms of a difference between remembered position and experienced position, it is better defined in terms of a difference between remembered position and believed position. Section 5 continues the reconstruction of Fernández' argument, focusing on the role in the argument of a principle regarding the relationship between memory and belief and arguing that this principle is problematic. Section 6 concludes the reconstruction of Fernández' argument, focusing on the role in the argument of a principle regarding the content of memory. Section 7 assesses the reconstructed argument, arguing that the content principle is problematic. Section 8 shows that, if we adopt the suggested definition of observer memory and reject the content principle, it follows that observer memory does indeed undermine the IEM claim. Episodic memory, in short, is *not* IEM.

¹ Consider, e.g., Bernecker's (2017) insistence, shared by McCarroll (2018), that successful memories are both "true" (accurate with respect to the remembered event) and "authentic" (accurate with respect to the subject's experience of that event).

² In addition to Fernández (forthcoming), see McCarroll (2018). McCarroll's discussion of observer memory would require a separate treatment and will therefore not be considered here.

³ All references without page numbers are to Fernández (forthcoming), published online ahead of print and unpaginated.

2 Immunity to error through misidentification

The thought that animates Shoemaker's (1968, 1970) influential discussion of IEM is that, where certain sources of knowledge are concerned, it is impossible for one to be right in believing that a certain property is instantiated but wrong in believing that it is oneself who instantiates that property. Consider the example of a pain judgement based on introspection: "[i]n being aware that one feels pain", Shoemaker argues, "one is, tautologically, aware not simply that the attribute *feels pain* is instantiated, but that it is instantiated in *oneself*" (1968: pp. 563–564). The question at issue here is whether memory judgements are similarly IEM.

Applying Shoemaker's definition of IEM to memory judgements, Fernández obtains the following definition of IEM in memory.⁴

For any property P and memory M, if I judge that I had P on the basis of M, then that judgement is IEM relative to M iff it is impossible that there is a subject S such that

- (1) M represents S as having had P
- (2) M is fully accurate
- (3) I mistakenly think that I am identical to S
- (4) My judgement that I had P is false as a result of (iii).

Shoemaker himself suggests that memory might be IEM. If I remember John's involvement in a certain event, I can be right about the features of the event but wrong that it was John who was involved; it could, for example, have been his twin brother. If I remember *my own* involvement in an event, Shoemaker suggests, I cannot, in contrast, be right about the features of the event but wrong that it was me who was involved: "[m]y memory report could of course be mistaken, for one can misremember such incidents, but it could not be the case that I have a full and accurate memory of the past incident but am mistaken in thinking that the person I remember [performing a certain action] was myself" (1970: p. 3).

Shoemaker's suggestion notwithstanding, it is not, at first glance, obvious why memory beliefs should be taken to be IEM. Decades of research on memory error have, after all, demonstrated that memory can get virtually everything else wrong. It would, to put it mildly, be surprising if the one thing that it could *not* get wrong were the subject's own identity. In vicarious memory (Pillemer et al. 2015), for example, subjects have memories for events experienced by others that do not appear to differ qualitatively from memories for events experienced by themselves. In borrowed memory (Brown et al. 2015), subjects recount events experienced by others as if they had been experienced by themselves and, in some cases, are uncertain about the identity of the subject who originally experienced the event; presumably, the resolution of this uncertainty results in some cases in beliefs that are inaccurate with respect to the identity of the subject despite being accurate with respect to the event. Vicarious memory, borrowed memory, and similar phenomena—such as disputed memory in twins (Sheen et al. 2001)—would seem to imply that it is not impossible that, when

⁴ It will matter in what follows that, given how IEM in memory is defined, the IEM claim, strictly speaking, says not that *memories* are IEM but rather that *memory judgements* (or beliefs) are IEM.

one believes on the basis of memory that one instantiated a certain property, one is right in believing that the property was instantiated but wrong in believing that it was oneself who instantiated that property. Indeed, they would seem to imply that we should expect that there are not only *possible* but also *actual* cases of memory beliefs that are erroneous due to this sort of misidentification. Fernández argues, however, that this expectation is not borne out.

3 Observer memory: the threat to the IEM claim

Fernández is concerned not with vicarious, borrowed, or disputed memory but rather with *observer memory*, which he defines, as is standard, in opposition to *field memory*:

[A memory may present a past scene from] the type of perspective from which the subject would have perceptually experienced the scene if the subject had been a part of it, or had gone through it, in the past. By having a memory which presents a past scene from a perspective of this type, the subject visualises the past scene, but they do not visualise themselves as a part of it. Let us call memories which present past scenes from this type of perspective, ... “field” memories. A memory may also present a past scene from the type of perspective that a different observer would have had to occupy in the past in order to witness the remembered scene with the subject as a participant of it. By having a memory which presents the past scene from a perspective of this type, the subject visualises not only the past scene but they also visualise themselves, as it were, from the outside. Let us call memories which present scenes from this type of perspective, ... “observer” memories.

Note that this definition characterizes the field/observer distinction in terms of two distinct features. On the one hand, it takes field memories to be memories in which *remembered position* (the position from which the event is remembered) coincides with *experienced position* (the position from which the event was experienced),⁵ suggesting that any memory in which remembered position deviates from experienced position is an observer memory. On the other hand, it takes observer memories to be memories in which one visualizes oneself in the remembered scene, suggesting that any memory in which one does not visualize oneself in the remembered scene is a field memory. These two features, however, need not coincide, for there are memories in which remembered position deviates from experienced position but in which one does *not* visualize oneself in the remembered scene. One might, for example, experience an event from a certain position and remember it from a position five centimetres to the left of that position. It is not, given Fernández’ definition, clear whether such a memory should be classified as a field memory or rather as an observer memory.

It would be natural to attempt to resolve this difficulty by replacing the field/observer distinction with a trisdistinction: there are field memories, in which remembered position coincides exactly with experienced position; there are observer memories, in which remembered position deviates sufficiently from experienced position to enable one to visualize oneself in the remembered scene; and there are intermediate memories,

⁵ Fernández does not employ these terms, but they will serve as convenient shorthand when summarizing his argument. The same thing goes for the notion of “believed position”, introduced below.

in which remembered position deviates from experienced position but does not do so sufficiently to enable one to visualize oneself in the remembered scene. It will, however, become clear below that the difficulty is better resolved by adopting an alternative definition of the field/observer distinction, one on which it pertains *neither* to the relationship between remembered position and experienced position *nor* to whether one visualizes oneself in the remembered scene.

Definitional difficulties notwithstanding, Fernández is clearly right to note that observer memory intuitively seems to undermine the IEM claim. It is, he remarks, plausible to suppose that, when one has an observer memory, one must identify oneself in the remembered scene by means of an inference and that, when one performs such an inference, one might arrive at a mistaken judgement despite the fact that one's memory is fully accurate. Moreover, though Fernández focuses on inference, it is plausible to suppose that the identity component of a belief based on an observer memory might, rather than resulting from inference, be determined instead by semantic information retrieved along with the episodic information and that here, too, error due to misidentification is possible: the identity component of the belief might be wrong, while the remainder of its content is right. It seems, in other words, that, in cases of both sorts, if I have an observer memory of an event in which I participated and judge that I played a certain role in that event, my judgement might in principle go wrong in the same way that it might go wrong if I were to have a memory of an event in which someone else participated and judge that he played a certain role in the event—it might have been someone else who played the role in question.

4 Defining observer memory

Fernández argues that this intuitive thought is mistaken. His argument is organized around a case

in which I have a mental image wherein I visualise the position of the passengers in a car during a traffic accident in the past. Let us suppose that there is a person sitting directly behind the driver's seat, but I do not visualise the scene from that passenger's perspective. Instead, I visualise the scene from the perspective of the passenger sitting directly behind the front passenger's seat. [...] Suppose, then, that I judge, on the basis of [the mental image that I have when I visualise the scene] that it was me who was sitting directly behind the driver's seat at the time of the accident.




Note that, in addition to the two positions distinguished above—remembered position and experienced position—there is now a third position in play: *believed position*, the position from which the subject takes himself to have experienced the event.

The details of the scenario described by Fernández do not matter; what matters is that it has the features captured by the following two assumptions.

(A1) S1 has an apparent memory m_1 of event e_1 from position p_1 (and not p_2).

(A2) S1 has a belief b_1 that he occupied p_2 (and not p_1) during e_1 .

Fig. 1 Three versions of Fernández' case

case 1		
	remembered experienced	believed
case 2		
	remembered	believed experienced
case 3	experienced	
	remembered	believed

In order for the case to amount to an instance of error due to misidentification, it must be such that the subject has a memory of the event that is both genuine *and* fully accurate *and* at the same time such that the resulting belief is false because it misidentifies the subject. There are thus three conditions that must be met before we have a counterexample to the IEM claim:

(IEM) m1 and b1 provide a counterexample to the IEM claim only if it is possible that m1 is a genuine memory, b1 is false, and m1 is fully accurate.

So far, remembered position and believed position have been fixed but experienced position has not. There are three possibilities with respect to experienced position:

(A3) It is necessary that either S1 occupied p1 during e1 or S1 occupied p2 during e1 or S1 occupied neither p1 nor p2 during e1.

This gives us three versions of the case (see Fig. 1):⁶ in case 1, I was sitting behind the front passenger's seat, i.e., experienced position coincides with remembered position; in case 2, I was sitting behind the driver's seat, i.e., experienced position coincides with believed position; and in case 3, I was sitting in neither of those positions, i.e., experienced position coincides neither with remembered position nor with believed position.

Consider the first version of the case:

(CASE1) S1 occupied p1 during e1.

While Fernández' argument is meant to concern observer memory, case 1, in which remembered position coincides with experienced position, is, given his definition of the field/observer distinction in terms of the existence of a discrepancy between remembered position and experienced position, not—as he himself acknowledges—an

⁶ We are assuming that the accident took place in Australia.

instance of observer memory. It is, however, natural to view case 1 as an instance of observer memory, and it is worth pausing to consider why this might be.

Recall that Fernández' definition of the field/observer distinction refers not only to the relationship between remembered position and experienced position but also to the notion of "visualizing oneself in the remembered scene". It is not immediately obvious what it is to visualize oneself in a remembered scene. We cannot say simply that one visualizes oneself in a remembered scene just in case one's visual representation of the scene includes a representation of oneself, for whether one counts as representing oneself presumably depends on the identities that one assigns to the individuals representations of whom are included in one's representation of the scene. The natural way of spelling out what it is for one to visualize oneself in the remembered scene is thus to say that one does so if one *takes oneself* to visualize oneself in the remembered scene, i.e., that one visualizes oneself in the remembered scene just in case one believes that one is identical to one of the individuals representations of whom are included in one's representation of the scene.

This would explain why it is natural to view case 1 as an instance of observer memory. Note, however, that, if one believes that one is identical to one of the individuals representations of whom are included in one's representation of the scene, then remembered position deviates from believed position. Thus, once we have spelled out the notion of seeing oneself in a remembered scene, it is natural to go one step further and say that observer memory should be defined in terms of the existence of a discrepancy between remembered position and *believed* position, rather than the existence of a discrepancy between remembered position and *experienced* position. Observer memories would then be memories in which remembered position deviates from believed position, while field memories would be memories in which remembered position coincides with believed position. Call this "the belief-divergent definition" of the field/observer distinction.

The belief-divergent definition has two key advantages over Fernández' definition. First, it explains why it is natural to group cases 1–3 together: remembered position deviates from experienced position in only two of the three cases (cases 2 and 3) but deviates from believed position in all three cases, so all three cases are, according to the belief-divergent definition, instances of observer memory. Second, it avoids the difficulty noted above for Fernández' definition of observer memory in terms both of the relationship between remembered position and experienced position and of the notion of seeing oneself in the remembered scene: remembered position and believed position either coincide or they do not, so the alternative characterization enables us to rule out the possibility of intermediate cases.

Of course, we might, in order to rule out intermediate cases, instead drop the reference to visualizing oneself in the remembered scene from Fernández' definition of observer memory, opting to characterize the field/observer distinction simply in terms of the relationship between remembered position and experienced position. Observer memories would then be memories in which remembered position deviates from experienced position, while field memories would be memories in which remembered position coincides with experienced position. Call this "the experience-divergent definition".

The experience-divergent definition does not explain why it is natural to group cases 1–3 together. Moreover, it has awkward consequences. Given the reconstructive character of remembering, we should expect that remembered position always or almost always deviates to some extent from experienced position. On the experience-divergent definition, then, all or almost all memories will count as observer memories. The definition is thus of doubtful utility. On the belief-divergent definition, in contrast, many memories will count as field memories, since, while remembered position may (almost) always deviate to some extent from experienced position, remembered position and believed position often coincide. The belief-divergent definition is thus preferable to the experience-divergent definition.

5 The belief principle

What ultimately matters, as far as the success of Fernández' argument is concerned, is not whether case 1 is an instance of *observer* memory but rather whether it is an instance of *memory* (whether observer or field) that provides a counterexample to the IEM claim. Initially, case 1 would indeed seem to provide a counterexample to the IEM claim: S1 not only occupied p1 but also remembers from p1, so his memory is presumably fully accurate; but he believes that he occupied p2, so his belief is false. But Fernández argues that this appearance is misleading.

In opposition to orthodox versions of the causal theory of memory (e.g., Bernecker 2010), he argues that “one of the conditions that a mental image representing some scene must satisfy for it to qualify as a memory of the scene is that the mental image must *dispose us to believe* that the scene took place in the past” (emphasis added). Call this “the belief principle”. Fernández (2018) defends the belief principle at length, and Debus (2010) has defended a closely related claim. The belief principle may be unorthodox, but it is not implausible, and it will not be challenged here. With respect to the subject's own position in the scene, Fernández' discussion suggests that the belief principle amounts to the following.

(BELIEF) If S has a genuine memory *m* of *e* from *p*, then, if S has no special reason to distrust *m*, S has a belief *b* that he occupied *p* during *e*.

Given this, as long as we assume that

(A4) S1 has no special reason to distrust *m*₁

it follows that

(1) *m*₁ is not a genuine memory.

As Fernández puts it, if the belief principle must be met “for a mental image to qualify as a memory, then the mental image that I have in the traffic accident case, *when we assume that I was sitting behind the front passenger's seat at the time of the accident*, does not qualify as a memory. After all, in virtue of having the mental image, I have no inclination to believe that the passenger behind the driver was, in the past, sitting to my right” (emphasis added), since otherwise I would not now believe that it was me who was sitting behind the driver's seat.

It is crucial to note that, if this argument succeeds, it shows that case 1 does not constitute a counterexample to the IEM claim not because m_1 is not fully accurate and not because b_1 is not true but rather because m_1 *is not a genuine memory*. But this means that the emphasized phrase in the passage just quoted does no argumentative work: the assumption that S_1 occupied p_1 during e_1 plays no role in the derivation of the conclusion that m_1 is not a genuine memory. Moreover, if m_1 is not a genuine memory in case 1, it is likewise not a genuine memory in cases 2 and 3: given the belief principle (and setting aside cases in which the subject has special reason to distrust his memory), no case in which believed position deviates from remembered position will qualify as an instance of genuine memory. The consequence is that Fernández' argument could in principle be very short indeed:

(A1) Suppose that S_1 has an apparent memory m_1 of e_1 from p_1 (and not p_2).

(A2) Suppose that S_1 has a belief b_1 that he occupied p_2 (and not p_1) during e_1 .

(A4) Suppose that S_1 has no special reason to distrust m_1 .

(BELIEF) If S has a genuine memory m of e from p , then, if S has no special reason to distrust m , S has a belief b that he occupied p during e .

(1) Then m_1 is not a genuine memory. (From A1, A2, A4, BELIEF.)

From this, we can immediately infer that cases of the kind singled out by the three assumptions poses no threat to the IEM claim.

The difficulty is that it is then unclear what the relevance of cases 1–3 to the question of IEM in memory might be. One way of putting the point is to note that, if we adopt the belief-divergent definition of observer memory, Fernández is, since he endorses the belief principle, straightforwardly committed to the view that observer memories never qualify as genuine memories. Of course, Fernández himself does not adopt the belief-divergent definition, but the underlying point remains: by his own lights, none of the cases in question qualifies as an instance of genuine memory and thus none of these cases is even a potential counterexample to the IEM claim.⁷

6 The content principle

Putting this difficulty aside for the moment, consider the second version of the case:

(CASE2) S_1 occupied p_2 during e_1 .

Since S_1 believes that he occupied p_2 during e_1 , it follows that

(2) b_1 is not false.

⁷ He concedes this point in Fernández (2019). One might take this difficulty to provide sufficient reason to reject the belief principle, but, as it is not necessary, in order to secure the conclusion of this paper, to reject the belief principle, the principle will not be challenged here. Note that the conclusion that episodic memory is liable to error through misidentification does not depend on the belief principle.

As Fernández points out, if $b1$ is not false, then case 2 cannot be an instance of error through misidentification simply because it is not an instance of error.

Consider, then, the third version of the case:

(CASE3) $S1$ occupied neither $p1$ nor $p2$ during $e1$.

Like case 1, case 3 would initially seem to provide us with a counterexample to the IEM claim. But Fernández argues that, in this case as well, the appearance is misleading.

His argument rests on a view of the content of visual representations on which, when a subject represents an object from a certain position, he thereby represents that he occupied that position relative to the object: “given that I visualise the passenger behind the driver’s seat from a certain perspective, my memory represents a past fact about myself, that is, the fact that I was sitting to the left of the remembered passenger”. Call the view expressed in this passage “the content principle”. What the content principle says is that

(CONTENT) If S has an apparent memory m of e from p , then, if m is fully accurate, S occupied p during e .

We will see below that the content principle is problematic. For now, note that it follows from the content principle and the description of case 3 that

(3) $m1$ is not fully accurate.

At this point, we have considered all three relevant cases, and we can infer that

(CONC1) It is necessary that either $m1$ is not a genuine memory, $b1$ is not false, or $m1$ is not fully accurate.

From this, plus (IEM), we can derive Fernández’ main conclusion:

(CONC2) $m1$ and $b1$ do not provide a counterexample to the IEM claim.

7 Assessing Fernández’ defence of the IEM claim

We have thus obtained the following reconstruction of Fernández’ argument.

(A1) Suppose that $S1$ has an apparent memory $m1$ of $e1$ from $p1$ (and not $p2$).

(A2) Suppose that $S1$ has a belief $b1$ that he occupied $p2$ (and not $p1$) during $e1$.

(A3) It is necessary that either $S1$ occupied $p1$ during $e1$ or $S1$ occupied $p2$ during $e1$ or $S1$ occupied neither $p1$ nor $p2$ during $e1$.

(A4) Suppose that $S1$ has no special reason to distrust $m1$.

(IEM) $m1$ and $b1$ provide a counterexample to the IEM claim only if it is possible that $m1$ is a genuine memory, $b1$ is false, and $m1$ is fully accurate.

(BELIEF) If S has a genuine memory m of e from p , then, if S has no special reason to distrust m , S has a belief b that he occupied p during e .

(CONTENT) If S has an apparent memory *m* of *e* from *p*, then, if *m* is fully accurate, S occupied *p* during *e*.

(CASE1) Suppose that S1 occupied *p*₁ during *e*₁.

(1) Then *m*₁ is not a genuine memory. (From A1, A2, A4, BELIEF.)

(CASE2) Suppose that S1 occupied *p*₂ during *e*₁.

(2) Then *b*₁ is not false. (From A2, CASE2.)

(CASE3) Suppose that S1 occupied neither *p*₁ nor *p*₂ during *e*₁.

(3) Then *m*₁ is not fully accurate. (From A1, CONTENT, CASE3.)

(CONC1) It is necessary that either *m*₁ is not a genuine memory, *b*₁ is not false, or *m*₁ is not fully accurate. (From A3, 1, 2, 3.)

(CONC2) *m*₁ and *b*₁ do not provide a counterexample to the IEM claim. (From IEM, CONC1.)

What should we make of the argument as a whole? (A1)–(A4) simply describe the kind of case in which we are interested, while (CASE1), (CASE2), and (CASE3) describe its possible versions; there is nothing to be challenged here. (IEM) simply tells us what would be required for one of the cases in question to undermine the IEM claim; there is likewise nothing to be challenged here. (BELIEF) and (CONTENT) are thus doing all the work.

What, then, should we make of (BELIEF) and (CONTENT)? We saw above that, because (CASE1) plays no role in the derivation of (1), we could in practice move directly from (A1)–(A3) and (BELIEF) to (1) and thence to the conclusion of the argument. This, again, suggests that (BELIEF) is problematic, at least as formulated, simply because it implies that none of cases 1–3 is even a potential counterexample to the IEM claim. Now, we will see below that the belief principle can be reformulated so that it implies (1) not by itself but only in conjunction with the content principle. In order to block the derivation of (1) from the reformulated belief principle, we might therefore reject the content principle, thus securing the status of cases 1–3 as potential counterexamples. But if we reject the content principle, the argument will no longer go through: without that principle, both case 1 and case 3 undermine the IEM claim.⁸ The situation, then, will turn out to be the following. If we do not reject the content principle, then, if we accept the belief principle, cases of the kind at issue here are irrelevant to the IEM claim. If we do reject the content principle (and we will see below that there is good reason to do so), then, regardless of whether we accept the belief principle, certain cases of this kind do indeed amount to counterexamples to the IEM claim. Moreover, if we reject the experience-divergent definition of observer

⁸ Once the content principle is rejected, the memory might be fully accurate in either case. Thus, each case would potentially be such that the subject has a fully accurate memory but nevertheless ends up with a belief that is erroneous due to misidentification.

memory in favour of the belief-divergent definition, then we arrive at the intuitively expected conclusion that observer memory undermines the IEM claim.⁹

What is required is a version of the belief principle that simply says that

(BELIEF') If S has a genuine memory *m*, and if *m* has content *c*, then, if S has no special reason to distrust *m*, S has a belief with *c*.

Because it refers to content without referring explicitly to position, BELIEF' would seem to be the minimal formulation of Fernández' view on the relationship between memory and belief. It will also be convenient to break the content principle into two:

(CONTENT1) If S has an apparent memory *m* of *e* from *p*, then *m* has the content that S occupied *p* during *e*.¹⁰

(CONTENT2) If *m* has the content that S occupied *p* during *e*, then, if *m* is fully accurate, S occupied *p* during *e*.



The argument can then be reformulated so that (1) follows from (A1), (A2), (A4), (BELIEF'), and (CONTENT1). (A1) tells us that S1 has an apparent memory *m*₁ of *e*₁ from *p*₁. It follows from this and (CONTENT1) that *m*₁ has the content that S1 occupied *p*₁ during *e*₁. It follows from this, (A4), and (BELIEF') that, if *m*₁ is a genuine memory, then S1 has a belief that he occupied *p*₁ during *e*₁. But (A2) tells us that S1 does not have a belief that he occupied *p*₁ during *e*₁. We can conclude that *m*₁ is not a genuine memory. The argument is otherwise the same, except that (CONTENT1) and (CONTENT2) replace (CONTENT) in the derivation of (3).

As before, the argument for the conclusion that *m*₁ is not a genuine memory does not depend on (CASE1). (BELIEF') and (CONTENT1) thus imply that *m*₁ is in all three cases not genuine. It thus remains unclear in what sense Fernández' argument concerns potential counterexamples to the IEM claim, since its assumptions straightforwardly imply that none of cases 1–3 is an instance of genuine memory. One might therefore wonder what was the point of introducing BELIEF'. This strategy has two benefits. First, the motivation that Fernández provides for the belief principle is independent of the motivation that he provides for the content principle. It seems preferable on general grounds to reformulate the belief principle so that it implies that observer memories (understood in line with the belief-divergent definition) are not genuine memories only in conjunction with the content principle. Some may be convinced by Fernández' or Debus' arguments for the belief principle but not by Fernández' argument for the content principle, and reformulating the belief principle will enable them to endorse that principle without thereby committing themselves the conclusion that cases in which believed position deviates from remembered position are never instances of genuine memory. Second, because the reformulated belief principle and the content principle together imply that none of cases 1–3 is an instance of genuine memory, the strategy reveals a first (though ultimately not the most decisive) reason to

⁹ Again, the key question here is not how to define the field/observer distinction. Note, however, that it will turn out that, if we reject the content principle, then there are cases of observer memory that undermine the IEM claim *even if* we adopt the experience-divergent definition.

¹⁰ The idea is, of course, not that information about the subject's position at the time of the remembered event is the *only* content of his memory but rather that such information is *part of* the content of his memory.

Fig. 2 Two more versions of Fernández' case

case 4		
	remembered believed	experienced
case 5		
	remembered believed experienced	

reject the content principle (more precisely, to reject (CONTENT1)—(CONTENT2) is innocuous): if we adopt the belief-divergent definition of observer memory, the content principle implies, together with the belief principle, that observer memory never qualifies as genuine memory.¹¹

If, rather than adopting the belief-divergent definition, we retain the experience-divergent definition, the content principle (in conjunction with the belief principle) no longer has that implication. Given the experience-divergent definition, there are three possible cases of observer memory: first, remembered position deviates from experienced position and believed position coincides with remembered position (case 4 in Fig. 2); second, remembered position deviates from experienced position and believed position coincides with experienced position (this is case 2 again); third, remembered position deviates from experienced position and believed position coincides neither with remembered position nor with experienced position (case 3 again). We already know that Fernández cannot count case 2 or case 3 as an instance of genuine memory. Regarding case 4, his assumptions are compatible with the possibility that it is an instance of genuine memory, so he can treat this case as being an instance of observer memory that is at least a potential counterexample to the IEM claim. Fernández will, of course, appeal to the content principle to argue that the memory in case 4 is not fully accurate, just as the memory in case 3 is not fully accurate, thus ruling it out as a counterexample to the IEM claim: the idea will be that the content of *m1* misrepresents the subject's position at the time of the remembered event. What ultimately matters, again, is not whether case 4 is an instance of observer memory but rather whether it is a counterexample to the IEM claim; we will see below that, if we reject the content principle, it does indeed undermine the IEM claim, along with cases 1 and 3.

We have already noted a first reason to reject the content principle, namely, that it implies—if we accept the belief principle—that cases in which remembered position comes apart from believed position are never instances of *genuine* memory. There is a second reason to reject the content principle, namely, that it implies—regardless of whether we accept the belief principle—that cases in which remembered position

¹¹ If, in order to avoid this implication, we were to reject the belief principle, then, if we were to accept the content principle, case 1 would turn out to be a counterexample to the IEM claim; if we were to reject the content principle, both case 1 and (potentially) case 3 would turn out to be a counterexample to the IEM claim.

comes apart from experienced position (cases 2–4) are never instances of *accurate* memory. Such cases are, we have remarked, widespread. Thus, just as the experience-divergent definition of observer memory entails a concept of observer memory of questionable utility, the content principle entails a concept of memory accuracy of questionable utility.

Fernández might in principle insist that this more stringent concept of memory accuracy is preferable. He links the view of memory content captured by the content principle to a similar view of perceptual content. The core claim of both views is that part of what one represents when one represents a scene is one's own position relative to the objects involved in the scene. This claim is not implausible with respect to perceptual content, and the idea that perceiving a scene involves representing one's own position relative to the objects involved in the scene will not be challenged here. But the claim is, for three reasons, much less plausible with respect to memory content.

Consider, first, the relationship between the content of a perceptual or memory representation, on the one hand, and the content of the corresponding perceptual or memory belief, on the other hand. If one *perceives* a scene from a certain position, then one is normally at least disposed to believe that one occupies the relevant position. If one *remembers* a scene from a certain position, in contrast, one is often *not* disposed to believe that one occupied the relevant position. If one were normally disposed to believe that one occupied the position from which a scene is remembered, then cases 1–3, in which believed position deviates from remembered position, would strike us as odd. But they do not—deviation of believed position from remembered position is utterly familiar. This suggests that memory content is unlike perceptual content with respect to the position from which the scene is represented.

Consider, second, the fact that a single scene may be remembered from multiple positions. Suppose that I first remember the car accident from the position of the passenger behind the driver's seat and that I then remember it from the position of the passenger behind the front passenger's seat. If the content principle were correct, this would entail that I first represent myself as having occupied the position of the passenger behind the driver's seat and then represent myself as having occupied the position of the passenger behind the front passenger's seat. But it is implausible to take me to entertain inconsistent representations of the event simply because I switch perspectives. The implausibility becomes particularly evident when we note that subjects often switch perspectives within a single episode of remembering (see Sutton 2010). Suppose that I first visualize the car accident from the position of the passenger behind the driver's seat and that my point of view then shifts, as the camera in a film might shift, to a new position. Clearly, I do not thereby represent myself as having moved within the scene. When the point of view from which I perceive a scene shifts, in contrast, it is plausible to say that I have represented myself as moving to a new position. Again, memory content appears to be unlike perceptual content with respect to the position from which the scene is represented.

Consider, third, the implausibility of the more general principle of which the content principle is a special case. The content principle in effect says that, if a subject would have had to have occupied a given position in order to have perceived the scene as it is visualized during remembering, then his memory represents him as having occupied that position during the scene. The more general principle is that, if a subject would

have had to have instantiated a given property in order to have perceived a scene as the scene is visualized during remembering, then his memory represents him as having instantiated that property during the scene. The content principle is just the general principle applied to spatial properties. The general principle might also be applied to temporal properties, but it is clearly implausible with respect to such properties. Memories involve not only perspective switching but also “cuts”, as films do: one may remember one part of an event and then a later part without remembering the intervening parts. The parts of an event are sometimes remembered out of order: one may remember first the conclusion of scene and then the earlier stages of the scene. Remembering may also involve “replaying” certain portions of a scene and “fast forwarding” through others. And so on. The point is that, when one remembers in these ways, one does not thereby represent oneself as having moved in time in the corresponding ways. The general principle is implausible when applied to temporal properties. And if it is implausible with respect to temporal properties, it is not clear why it should be taken to be more plausible with respect to spatial properties.

In short, perceptual content and memory content may have much in common, but they differ with respect to whether they impute to the subject occupancy of the position from which the scene is represented. If we therefore reject the content principle—again, what we need to reject is (CONTENT1)—then Fernández’ argument no longer goes through; instead, we obtain the positive conclusion that *memory is liable to error through misidentification*. Note that rejecting the content principle does not commit us to the strong claim that a memory *never* represents the remembering subject as having occupied the position from which the event is remembered but only to the weaker claim that a memory does not *necessarily* represent the subject as having occupied that position. This means that each of cases 1–4 will come in two versions: a version in which m1 does represent S1 as having occupied p1 and a version in which m1 does not represent S1 as having occupied p1. Let us restrict our attention to the latter versions. If we suppose, as before, that m1 is accurate with respect to all features of the scene other than the subject’s position, then m1 is fully accurate in cases 1–3. Having rejected the content principle, we are now free to treat cases 1–3 as instances of genuine memory: since m1 does not represent S1 as having occupied p1 during e1, the belief principle does not imply that m1 is not a genuine memory. As noted above, case 2 does not constitute a counterexample to the IEM claim, because b1 is not false in that case. But cases 1 and 3 do constitute counterexamples to the IEM claim. In both cases, m1 is fully accurate but b1 is nevertheless false due to misidentification: in case 1, S1 occupied p1 but believes that he occupied p2; in case 3, S2 occupied neither p1 nor p2 but believes that he occupied p2. In short, if we reject the content principle, and if we adopt the belief-divergent definition, then observer memory undermines the IEM claim; if we reject the belief-divergent definition, then cases 1 and 3 nevertheless undermine the IEM claim.

This leaves case 4 unaccounted for. The experience-divergent definition of observer memory classifies this as an instance of observer memory; the belief-divergent definition does not. If we adopt the content principle, we will see the case as involving a certain sort of error due to misidentification but not as involving the sort of error due to misidentification that would undermine the IEM claim: since m1 is not fully accurate, the fact that b1 inherits the inaccuracy of m1 does not show that a belief can

be inaccurate despite being based on a fully accurate memory. If we reject the content principle, however, the case might well involve the sort of error due to misidentification that would undermine the IEM claim. Believed position coincides with remembered position, but we are supposing that remembered position is not part of the content of m_1 . Thus m_1 might be fully accurate, in which case the inaccuracy of b_1 would not be inherited from m_1 but produced, for example, by the sort of inferential process described by Fernández. In short, if we reject the content principle, then, even if we adopt the experience-divergent definition of observer memory, observer memory undermines the IEM claim.

8 Conclusion: episodic memory is not IEM

What cases 1, 3, and 4 have in common is that believed position comes apart from experienced position. The upshot is thus that, regardless of which definition of observer memory we adopt, if we reject the content principle, then any case in which believed position deviates from experienced position provides a potential counterexample to the IEM claim. Cases of borrowed memory, for example, in which believed position similarly comes apart from experienced position, will thus provide potential counterexamples to the IEM claim: if we reject the content principle, then borrowed memories can be fully accurate, and, in a case in which a borrowed memory is fully accurate and in which the subject subsequently misjudges his own identity when forming a belief based on the memory, the memory and the belief together will undermine the IEM claim.

One way for remembering to go wrong is for it to output a merely apparent memory. Another way for remembering to go wrong is for it to eventuate in a false memory belief. Call remembering that avoids both of these pitfalls “successful remembering”. Fernández’ approach, encapsulated in the conjunction of the belief principle and the content principle, implies that the only way for remembering to be successful is for remembered position, believed position, and experienced position to coincide, as in case 5 in Fig. 2. This is a *highly* restrictive view of successful remembering. We have already seen that, given the reconstructive character of remembering, there are, on his approach, likely to be very few cases of genuine memory and very few cases of accurate memory. Putting these points together, his approach implies that there are likely to be very few cases of successful remembering indeed.

If, however, we reject the content principle, we end up with a much less restrictive view of successful remembering. Cases in which remembered position, believed position, and experienced position coincide, as in case 5, will of course count as instances of successful remembering. But so will cases in which believed position and experienced position coincide with each other but not with remembered position, as in case 2 (unless they are cases in which remembered position is part of the content of the memory). Fernández can, as noted above, acknowledge that the subject’s belief in case 2 is accurate. But he has to view this as a case in which the belief is accurate despite being based on a memory that is both inaccurate and merely apparent, i.e., he cannot view it as a case of successful remembering.

Rejecting the content principle thus has a significant advantage: observer memory (understood in line with the belief-divergent definition) is a routine occurrence, and we are frequently able to form accurate beliefs on the basis of the memories in question; rejecting the content principle enables us to see the relevant cases as instances of successful remembering. The “disadvantage” of rejecting the content principle is, of course, that we are bound to accept that memory is not IEM; as noted at the outset, however, this is precisely what we should expect.

One might object that, while the argument of this paper may succeed in showing that Fernández does not succeed in showing that observer memory does not undermine the IEM claim, it does not succeed in showing that observer memory does indeed undermine the IEM claim. In particular, one might object that Evans’ (1982) “stipulative” strategy shows that neither error through misidentification nor any other form of memory error can undermine the IEM claim.¹² Evans in effect argues that identity is given by fiat, leaving no room for error: if the subject takes himself to be remembering his own experience, then, regardless of whether he is remembering the experience accurately, it is his own experience that he is remembering; thus, if the memory is fully accurate, a belief based on it cannot be erroneous due to misidentification. In other words: it is impossible for a memory to be fully accurate and for the subject to subsequently misjudge his own identity, because his identity is stipulated from the outset.

This is not the place for a detailed discussion of Evans’ strategy (see Coliva 2006), but note that Perrin (2016) has pointed out that, while the stipulative strategy is plausible with respect to episodic future thought (the future-oriented counterpart to episodic memory; see Perrin and Michaelian 2017), since, in the case of episodic future thought, the subject’s current representation of the event is not caused by his experience of it, so that a causal link is unavailable to determine identity, it is less plausible with respect to episodic memory, since, in the case of episodic memory, the subject’s current representation of the event is caused by his experience of it, so that a causal link is available to determine identity. If Perrin is right, the stipulative strategy can be ruled out, with the consequence that error due to misidentification remains possible.

Whether one accepts Perrin’s argument will depend in part on whether one accepts the causal theory of memory (Martin and Deutscher 1966), according to which, if a subject genuinely remembers an event, then his current representation of it is necessarily caused by his experience of it. If one rejects the causal theory in favour of a post-causal theory such as Fernández’ (2018, 2019) functionalist theory¹³ or Michaelian’s (2016c) simulation theory,¹⁴ then one grants that a subject may in some cases genuinely remember an event despite the fact that his current representation of it is not caused by his experience of it. And in cases of episodic memory in which no causal link is available, it would seem to be natural to advert to something like Evans’ strategy, as Perrin does with respect to episodic future thought.

¹² McCarroll, who maintains that identity “is given immediately and non-inferentially” in both field memory and observer memory (2018: p. 165), arguably adopts a similar strategy, but, again, space does not permit a detailed discussion of his approach to IEM here.

¹³ Note that Fernández’ argument for the IEM claim turns specifically on his view of memory content, not on his functionalist theory of memory.

¹⁴ On the causalist–postcausalist debate, see Michaelian and Robins (2018).

As Michaelian (2016a) has pointed out, however, the simulation theory does not imply that genuine episodic memories are never caused by the events that they are about but only that they are sometimes not caused by the events that they are about; thus, if one rejects the causal theory in favour of the simulation theory, it is most natural to adopt a mixed strategy on which identity is sometimes determined by causation and sometimes determined by stipulation. A similar line of reasoning applies to the functionalist theory. In short, regardless of whether one adopts the causal theory or a postcausal theory, the pure version of the stipulative strategy that would be required to secure immunity to error due to misidentification can be ruled out.

To summarize: Fernández argues that observer memory does not undermine the IEM claim. This paper has argued that Fernández is wrong. His argument relies on a problematic definition of observer memory and a problematic view of memory content. Once these elements are rejected, it becomes clear that observer memory does indeed undermine the IEM claim. Episodic memory, then, is not immune to error through misidentification.

Acknowledgements Thanks to audiences at the Experience of Ownership workshop at the University of Adelaide, the Memory: A Self-Referential Account workshop at the Université Grenoble Alpes, and seminars at the University of Otago and East China Normal University. Thanks also to three referees for *Synthese*. Special thanks to Jordi Fernández for detailed discussion. This work is supported by the French National Research Agency in the framework of the “Investissements d’avenir” program (ANR-15-IDEX-02).

References

- Bermúdez, J. L. (2012). Memory judgements and immunity to error through misidentification. *Grazer Philosophische Studien*, 84(1), 123–142.
- Bermúdez, J. L. (2013). Immunity to error through misidentification and past-tense memory judgements. *Analysis*, 73(2), 211–220.
- Bernecker, S. (2010). *Memory: A philosophical study*. Oxford: Oxford University Press.
- Bernecker, S. (2017). A causal theory of mnemonic confabulation. *Frontiers in Psychology*, 8, 1207.
- Brown, A. S., Croft Caderao, K., Fields, L. M., & Marsh, E. J. (2015). Borrowing personal memories. *Applied Cognitive Psychology*, 29(3), 471–477.
- Coliva, A. (2006). Error through misidentification: Some varieties. *The Journal of Philosophy*, 103(8), 403–425.
- De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), 155–185.
- Debus, D. (2007). Perspectives on the past: A study of the spatial perspectival characteristics of recollective memories. *Mind and Language*, 22(2), 173–206.
- Debus, D. (2010). Accounting for epistemic relevance: A new problem for the causal theory of memory. *American Philosophical Quarterly*, 47(1), 17–29.
- Eldridge, P. (2014). Observer memories and phenomenology. *Phenomenology and Mind*, 7, 160–167.
- Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Fernández, J. (2014). Memory and immunity to error through misidentification. *Review of Philosophy and Psychology*, 5(3), 373–390.
- Fernández, J. (2015). What are the benefits of memory distortion? *Consciousness and Cognition*, 33, 536–547.
- 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). Abingdon: Routledge.
- Fernández, J. (2019). *Memory: A self-referential account*. Oxford: Oxford University Press.

- Fernández, J. (forthcoming). Observer memory and immunity to error through misidentification. *Synthese*. <https://doi.org/10.1007/s11229-018-02050-3>.
- Hamilton, A. (2007). Memory and self-consciousness: Immunity to error through misidentification. *Synthese*, 171, 409–417.
- Martin, C. B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75(2), 161–196.
- McCarroll, C. J. (2017). Looking the past in the eye: Distortion in memory and the costs and benefits of recalling from an observer perspective. *Consciousness and Cognition*, 49, 322–332.
- McCarroll, C. J. (2018). *Remembering from the outside: Personal memory and the perspectival mind*. Oxford: Oxford University Press.
- McCarroll, C. J., & Sutton, J. (2017). Memory and perspective. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory* (pp. 113–126). Abingdon: Routledge.
- Michaelian, K. (2016a). Against discontinuism: Mental time travel and our knowledge of past and future events. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the future: Theoretical perspectives on future-oriented mental time travel* (pp. 63–92). Oxford: Oxford University Press.
- Michaelian, K. (2016b). Confabulating, misremembering, relearning: The simulation theory of memory and unsuccessful remembering. *Frontiers in Psychology*, 7, 1857.
- Michaelian, K. (2016c). *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge: MIT Press.
- Michaelian, K. (2020). Confabulating as unreliable imagining: In defence of the simulationist account of unsuccessful remembering. *Topoi*, 39(1), 133–148.
- Michaelian, K., Perrin, D., & Sant’Anna, A. (forthcoming). Continuities and discontinuities between imagination and memory: The view from philosophy. In A. Abraham (Ed). *The Cambridge handbook of imagination*. Cambridge: Cambridge University Press.
- Michaelian, K., & Robins, S. K. (2018). Beyond the causal theory? Fifty years after Martin and Deutscher. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory* (pp. 13–32). Abingdon: Routledge.
- Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, 15(4), 467–482.
- 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: Oxford University Press.
- Perrin, D., & Michaelian, K. (2017). Memory as mental time travel. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory* (pp. 228–239). Abingdon: Routledge.
- Pillemer, D. B., Steiner, K. L., Kuwabara, K. J., Thomsen, D. K., & Svob, C. (2015). Vicarious memories. *Consciousness and Cognition*, 36, 233–245.
- Recanati, F. (2007). *Perspectival thought: A plea for (moderate) relativism*. Oxford: Oxford University Press.
- Rice, H. J. (2010). Seeing where we’re at: A review of visual perspective and memory retrieval. In J. H. Mace (Ed.), *New perspectives in cognitive psychology the act of remembering: Toward an understanding of how we recall the past* (pp. 228–258). Hoboken: Wiley-Blackwell.
- Robins, S. K. (2016). Misremembering. *Philosophical Psychology*, 29(3), 432–447.
- Robins, S. K. (2019). Confabulation and constructive memory. *Synthese*, 196(6), 2135–2151.
- Schacter, D. L. (2019). Implicit memory, constructive memory, and imagining the future: A career perspective. *Perspectives on Psychological Science*, 14(2), 256–272.
- Sheen, M., Kemp, S., & Rubin, D. (2001). Twins dispute memory ownership: A new false memory phenomenon. *Memory & Cognition*, 29(6), 779–788.
- Shoemaker, S. S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, 65(19), 555–567.
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7(4), 269–285.
- Sutton, J. (2010). Observer perspective and acented memory: Some puzzles about point of view in personal memory. *Philosophical Studies*, 148(1), 27–37.