A Metacognitive Approach to Memory Markers

MATHEUS DIESEL WERBERICH*
Federal University of Santa Maria, Brazil

Abstract Given both the phenomenological and cognitive similarities between episodic memory and imagination, it's difficult to say how we can reliably distinguish them at their moment of retrieval. Several memory markers have thus been proposed, which are characteristics that would reliably indicate to the subject that her mental state is an instance of memory. While the question of what exactly constitutes these memory markers is still an issue to be settled, there is also the more general question of whether they can be reliable at all. In the present paper, I have identified two theses about the latter issue (the reliability and unreliability theses) and have argued that the main cause of disagreement between them lies on their different assumptions on how beliefs about our own mental states are formed, i.e. what are the underlying metacognitive mechanisms responsible for self-attribution of mental states. These different views on metacognition were then further investigated with regards to their use of metarepresentations and their general agreement with recent cognitive psychology research. These analyses corroborate the reliability of certain kinds of memory markers in distinguishing between memory and imagination.¹

1 Introduction

There are many different metaphors used to explain memory in the history of western philosophy, from Plato's wax tablet to Locke's storehouse imagery. What all of these conceptions have in common is that they take memory to be a passive system:

^{*}Matheus Diesel Werberich is currently an undergraduate Philosophy student at the Federal University of Santa Maria (UFSM), Brazil. As a member of the Philosophy of Memory Lab, his research is focused on the nature of episodic memory, metacognition and the phenomenological aspects of mental time travel. He is also interested in metaphysics of time, philosophy of mind and teaching philosophy.

^{1.} Special thanks to Glaupy Fontana Ribas and César Schirmer dos Santos for their help in writing this paper.

information is carried through perception to memory, where it is stored for later retrieval. Although this characterization was used for a few centuries, this framework is no longer widely adopted to account for episodic memory, i.e. the recollection of personal past episodes. This is mainly due to recent findings in neuroscience and cognitive psychology.² Researchers in those areas discovered that a major part of episodic memory's neural substrates are the same as those of episodic imagination: the same areas and processes are active when the subject is remembering a past episode and when she is imagining a future or counterfactual event. Thus, it has been widely accepted that episodic memory, like imagination, is a kind of process that involves the active construction of episodes, not the passive retrieval of them from past experience.

Given its constructive character and relation to episodic imagination, many philosophers have claimed that episodic memory and imagination share the same nature, so much so that there is not a qualitative distinction between them. This proposal, called continuism, is at odds with the more traditional and commonsensical view, named discontinuism, in which episodic memory and imagination are different in nature, even though their phenomenology might be similar.³ Whichever is the true alternative, they both face the same problem: if memory and imagination are so phenomenologically similar, then how can we tell them apart at the moment of retrieval? Is there an element that would be exclusive to, or at least largely more present in memory, that distinguishes it from imagination?⁴ Traditionally, the answer to the latter question has been affirmative, and philosophers of memory call this kind of element a *memory marker*. In what follows, I will analyze this kind of proposal and argue that the discussion about the reliability of these memory markers depends on certain considerations about the functioning of a larger metacognitive system.

2 The Memory Markers Proposal

In general, there are two kinds of memory markers: first person markers and third person markers. The first kind are the ones to which the subject has direct access at the moment of retrieval. One classic example of this kind is the feeling of familiarity: the content of memory representation feels familiar to the subject, giving her the con-

^{2.} For an example of such studies, see Schacter and Donna Addis, "The cognitive neuroscience of constructive memory". Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 362, 1481 (2007). doi: 10.1098/rstb.2007.2087.

^{3.} For a more detailed discussion on continuism and discontinuism, see Kourken Michaelian and Denis Perrin, "Memory as mental time travel" in The Routledge Handbook of the Philosophy of Memory, ed. Sven Bernecker and Kourken Michaelian (New York: Routledge, 2017): 228 – 239.

^{4.} Note that this is a different issue from how the agent can trust that her memory is in fact giving her reliable information about the past. The solution to this problem relies not on memory markers, but rather if the memory process itself can reliably retain, or reconstruct, information from the past.

21

fidence to consider it as a case of episodic memory. Meanwhile, third person memory markers are external to the subject's current mental state because they can only be determined by analyzing certain aspects that concern memory's formation process. For example, an unbroken causal chain between the original experience and its current representation is a kind of third person memory marker because only memory can have it, and it cannot be detected just by checking its resultant mental state. Since philosophers consider that we can tell memory from imagination at the moment of retrieval, their focus will be to search for first person memory markers, which, following Michaelian's classification, can be divided into three groups: formal, content-based, and phenomenological.⁵

Formal memory markers are concerned primarily with the way that memory content is presented to the agent. For instance, memory tends to be more spontaneously retrieved than imagined scenarios, so much so that the speed of memory's retrieval can be an indicator that the subject encountered something similar in the past. Other formal criteria include memory's inflexibility (i.e. memory's susceptibility to change based on other inputs, such as other's testimony), its spontaneity, fluency (i.e. the relative easiness of memory's formation, related to the spontaneity criterion) and the intention to remember.

In turn, content-based memory markers pertain to the idea that certain aspects of memory content can be used as criteria for its classification. Characteristics such as greater level of detail and coherence with other beliefs may play a role in distinguishing between memory and imagination. It is important to note, as it will become relevant in future sections, that these two categories of memory markers, i.e. formal and content-based, aren't a matter of absolute presence or absence, but rather are characteristics that are more present in memory than in imagination.

The third and final type of memory markers is concerned with the phenomenological aspects of episodic memory. In particular, these accounts consider the metacognitive feelings brought by memory that should be absent in cases of imagination. Metacognitive feelings are a way of the metacognitive system to point certain features about other cognitive processes. Such feelings include the feeling of prior belief, pastness, familiarity, and confidence. These criteria are not exactly characteristics of the memory process, but are rather generated by the metacognitive system through its monitoring of the episodic construction process.

^{5.} Kourken Michaelian, Mental Time Travel: episodic memory and our knowledge of the personal past. (Cambridge, MA: The MIT Press, 2016), 181 - 194.

3 The Metacognitive Approach to Memory Markers

Now that we have seen the major proposals for memory markers found in the literature, we can turn our attention to a more general problem: can memory markers generally be reliable in distinguishing between memory and imagination? The answer to this question will largely depend on what counts as "reliable": if the motivation provided by the memory markers comes from a reliable enough system, then the answer would be affirmative; on the other hand, if we define reliability as the capacity to produce justified beliefs based on well-grounded criteria, then memory markers wouldn't be the most appropriate, since most of them can be also found in imagination. Therefore, we have two theses about the overall reliability of memory markers:

- (1) Reliability thesis: the motivation to classify a mental state as memory provided by memory markers is sufficiently reliable, since it comes from an overall reliable system;
- (2) *Unreliability thesis*: the motivation provided by memory markers is not sufficiently reliable, since they don't provide valid justification for metamemory beliefs.

In the recent literature, these two theses are defended mainly by philosophers standing in different camps in philosophy of memory. Simulationists, such as Michaelian⁶, would generally argue for the reliability thesis, since, if episodic memory is defined as reliable imagining, or simulation, of past events, its reliability would be lost if the agent couldn't distinguish between it and other forms of imagination. According to Michaelian, "one normally 'just knows' whether one is remembering or imagining"⁷, thus pointing to the fact that an active and conscious introspection might not be required for determining whether an episodic state is a case of memory or imagination.

On the other hand, causal theorists, like Bernecker⁸, tend to argue for the unreliability thesis, stating that the motivation provided by memory markers to distinguish between memory and imagination does not mean that they provide "an epistemic reason for anything" ⁹. This means that the result of using memory markers should be a justified belief about one's episodic mental state, not just some notion or feeling, as in the reliability thesis.

One thing to note about the two theses is that their considerations on the reliability of memory markers come from different views on how metamemory beliefs are formed. The reliability thesis considers that there is a subconscious background system which closely monitors the memory process and, depending on the characteristics detected, makes the subject more susceptible to treat her mental state as memory. On the other hand, the unreliability thesis takes the metamemory beliefs as forming through

^{6.} Kourken Michaelian, Mental Time Travel: episodic memory and our knowledge of the personal past.

^{7.} Ibid, 194.

^{8.} Sven Bernecker, Memory: a philosophical study (Oxford, Oxford University Press, 2010).

^{9.} Ibid, 33.

a sort of introspection, where the subject will consciously look for memory markers in her mental states and form a judgement based on their presence or absence. Thus, the unreliability thesis requires that the subject has at least some idea of how to classify her mental states, while the reliability thesis does not require such mental capacity, only that a background subsystem must be able to detect certain features of memory.

Therefore, the discussion on the general reliability of memory markers mainly comes from different views about how mental states can be detected and evaluated. These capacities are studied in the field of metacognition, which is still characterized by a general lack of agreement in many issues. In order to organize these problems, Joëlle Proust¹⁰ distinguishes between two general theories of metacognition: the evaluativist view and attributive view. According to an attributive view of metacognition, the subject becomes aware of her own mental states through a conceptual representation of them. Thus, the metacognitive system would generally rely on what is similar to type 2 kind of processing¹¹, which is regarded as the slow, conceptual and propositional mechanism of information processing. In contrast, the evaluativist view, also known as procedural view, considers metacognition to largely rely on what is similar to type 1 processing, which is the fast, subconscious and heuristics-based mechanism, responsible for the evaluation of general features of cognitive processing, such as fluency.

If one considers the evaluativist view to be true and its monitoring system to be reliable enough, then one naturally comes to the conclusion that the reliability thesis is true. Conversely, if one accepts the attributive view, then the memory markers presented in section 2 may not serve as a sufficient justification factor for their metamemory beliefs, thus following the unreliability thesis. In what follows, I will present in more detail how the evaluativist and attributive views work, as well as their implications for the question about the general reliability of memory markers.

3.1 Attributive Views

In the attributive view, any form of metacognition requires a metarepresentation of the target mental state. Metarepresentations are understood as a representation of both the content and the representational vehicle. For example, if a subject has a mental picture of her future birthday party, then her metacognitive system would have to represent the same mental picture *as* a case of imagination – her metarepresentation would have information of both the content and the process involved in making it¹². Following

^{10.} Joëlle Proust, The Philosophy of Metacognition: Mental Agency and Self-awareness (Oxford: Oxford University Press, 2013).

^{11.} For a more detailed account of these kinds of cognitive processing, see Daniel Kahneman, Thinking, Fast and Slow (New York: Farrar, Straus and Giroux, 2011).

^{12.} For a more detailed discussion on metarepresentations, see Josef Perner, "MiniMeta: in search of minimal criteria for metacognition", Foundations of metacognition, ed. Michael Beran, Johannes Brandl

this, many researchers conclude that metacognition employs the same mechanisms of mindreading, which is the capacity to attribute mental states to other people based on perceptual clues, such as facial expressions, tone of voice, bodily movements, among many others. If that is the case, then attributive theorists would have to explain what exactly the relation between mindreading and metacognition is, which would in turn have consequences for their general theory of introspection.

For example, Carruthers claims that metacognition, or introspection, is just self-directed mindreading. His claim is that perceptual and recognition mechanisms, alongside concepts of mental states and the self, would be able to extract clues from the subject's internal states and infer propositional attitudes, just as it would to other people. Such clues include inner-speech, visual imagery, attention shifts, and so on. Of course, some of these clues might not be available for attributing mental states to other people. For instance, mental imagery of other agents is not as easily detected as one's own. However, this does not pose a problem for Carruthers' account, for it still might be the case that the same mechanisms are involved both in mindreading and introspection, just not necessarily the same inputs would be used in evaluating other's or one's own mental states.

One consequence of Carruthers' model is that we only have interpretative access to some of our mental states. If the mechanisms of mindreading are based on the interpretation of clues in another agent, then, when they turn on their own subject, they would have to interpret her behavioral and perceptual clues as well. Thus, much of our self-ascription of mental states has to be interpretative, including metamemory beliefs. In this case, the metarepresentational system would receive an episodic construction as input and ascribe the concept of "remembering" to the subject.

Alternatives to Carruthers' account of metacognition and mindreading generally use the same sort of metarepresentational mechanisms, but put the metacognitive capacity as being more fundamental than mindreading. An example of this is Goldman's account of self-knowledge, in which we only acquire the concepts necessary for metarepresentations through the detection of our own mental states. Based on this, the subject can then perceive behavioral clues in other agents and apply those concepts to them (i.e. the concepts of remembering, reasoning, and so on).

If metamemory beliefs are based on interpretation of clues, then memory markers can be detected for the self-ascription of episodic memory. Considering the formal and content-based markers presented in section 2, the perception of them could be used as input for the metarepresentational system. However, if memory and imagination share a lot of their formal and content-based properties, then it is unlikely that any

et al. (Oxford: Oxford University Press, 2012): 94 - 116.

^{13.} Peter Carruthers, "How we know our own minds: the relationship between mindreading and metacognition", Behavioral and Brain Sciences 32, 2 (2009): 121-182. doi:10.1017/S0140525X09000545.

^{14.} Alvin Goldman, Simulating Minds: the philosophy, psychology and neuroscience of mindreading (Oxford: Oxford University Press, 2006): 246.

single criterion, or set of criteria, can reliably set them apart – any of them would also be present in imagination, undermining the very notion of memory markers, which are characteristics that would characterize only memory states.

One might argue that this is only a problem if we adopt continuism about episodic memory and imagination, which is the view that they only differ in terms of degree, but not of kind. If one adopts discontinuism, on the other hand, they would hold that memory and imagination are different kinds of mental states, thus not being able to share formal and content-based properties in the same way that was described earlier. However, even if they are different kinds of mental states, it does not mean necessarily that they don't have common formal and content-based properties, since they are both mental images of non-present events. Despite the possibility that they are formed by different systems, the fact still remains that they present themselves in a similar way, for they are mental representations of non-present scenarios.

Therefore, notwithstanding the actual differences between memory and imagination, if one assumes an attributive view of metacognition, they would tend to consider formal and content-based memory markers as unreliable criteria for distinguishing memory and imagination. However, what if the metarepresentational system took the phenomenological markers as input? Can feelings of remembering serve as reliable criteria for episodic recollection? The answer to these questions depends on how these epistemic feelings are formed, which is the topic of the following section.

3.2 Evaluativist Views

As was stated earlier, the evaluativist view considers metacognition to be a subconscious, fast process of monitoring and control of cognitive mental states. This largely coincides with what is called type 1 processing, which is a kind of cognitive capacity that aims at achieving fast production of outputs from initial inputs. Since this is not a conscious or deliberate act of the subject, the outputs provided by the evaluativist model of metacognition will generally lack the conceptual and propositional structure that other cognitive states have.

In the evaluativist view, it is useful to distinguish between two levels of processing: the cognitive level, where mental states are primarily constituted of what ordinarily is thought of our mental activities, such as memory, beliefs, imagination and so on, and, the meta level, which is primarily concerned with tracking what is happening at the cognitive level. This kind of interaction between the two levels can happen in two different ways: monitoring, where information flows from the cognitive level to the meta level, and control, which has the opposite flow of information, allowing for the meta level to intervene on the cognitive level.

Since we are concerned with the overall reliability of distinguishing between memory and imagination, our main concern should be how exactly monitoring oc-

curs: how can the metacognitive level acquire relevant information from the cognitive processes to properly categorize them? Since categorization of mental states usually occurs through metacognitive feelings, evaluativist theories mainly focus on how the metacognitive system can monitor the cognitive level and create those subjective experiences, such as the feeling of pastness and of prior belief.

In this framework, we have two questions about memory markers. The first one is concerned with which are the memory markers that the metacognitive system would be able to detect in order to reliably classify a mental state as memory. We can call it *the monitoring question*. The second one is more concerned with later stages of the metacognitive process, asking which sort of epistemic feelings would the system generate to motivate the subject to accept the mental state as memory. We can call it *the feelings question*.

In order to answer these problems, Dokic first considers how the metacognitive system can produce epistemic feelings while trying to explain memory's phenomenology. According to Dokic, episodic memory has a distinct phenomenological feature, namely the episodic feeling of knowing, through which the subject becomes more inclined to think that her mental state comes directly from the past experience. Based on empirical research, Dokic claims that the episodic feeling of knowing is based on the detection of certain features of the episodic construction process, such as fluency and familiarity, which are reliable indicators of episodic memory.

Following this model, we can have some answers to the two questions about memory markers described earlier. The *feelings question* is about which markers are used in order to inform the subject about her mental states. Dokic answers this with the episodic feeling of knowing, since this feeling is a product of heuristics that are indications of the construction process being connected to the original experience. Following the classification presented in section 2, the episodic feeling of knowing would be a phenomenological memory marker.

The *monitoring question* is concerned with which are the memory markers are detected by metacognition that allow the episodic representation to be characterized as memory. Since phenomenological markers are only present in later stages of this kind of metacognitive monitoring, the decision must be between formal and content-based markers. Fortunately, empirical research has been fruitful in this regard, showing that monitoring of content of the object level does not seem to be as useful as tracking their general features.¹⁷ Thus, formal memory markers are the best examples for what exactly the metacognitive system is detecting in order to form the episodic feeling

^{15.} Jérôme Dokic. "Feeling the Past: A Two-Tiered Account of Episodic Memory." Review of Philosophy and Psychology 5, no. 3 (2014) doi: 10.1007/s13164-014-0183-6.

^{16.} Ibid, 422.

^{17.} Asher Koriat, "The subjective confidence in one's knowledge and judgements: some metatheoretical considerations", Foundations of metacognition, ed. Michael Beran, Johannes Brandl et al. (Oxford: Oxford University Press, 2012): 215.

of knowing. Favorable candidates include memory's spontaneity and fluency, since the easiness of memory's formation could be an indicator of the system having encountered the same contents in the past.

While Dokic's metacognitive model, as well as other evaluativists accounts, can ground the overall reliability of memory markers, they have one specific problem to be solved. According to evaluativism, the metacognitive system is relatively simple when compared with the cognitive system, since it does not use concepts or metarepresentations for monitoring and control. If that's the case, then procedural evaluativists would have to explain how the metacognitive level is able to tell, without relying on metarepresentations or concepts, that some elements of the episodic construction process indicate features of memory. This is relevant for the overall reliability of phenomenological memory markers, for their adequacy is dependent on how they are formed.

One possible way of answering this problem is by relying on what Proust calls Test-Operate-Test-Exit (TOTE) units, ¹⁸ which are feedback loops designed to compare a current mental state with the desired one (called the comparator) and, based on that comparison, inform control about what sort of action is necessary. For instance, when monitoring a given memory process, the metacognitive system would compare the episodic process' fluency with the previously stored comparator. If no major discrepancy is detected, then a feeling of remembering would be generated. This mechanism can avoid having conceptual information about the cognitive process, since the comparator does not necessarily need to have a concept of fluency clearly stated in order to operate: if the system is sensitive enough to the difference in speed and easiness between the formation of the current mental state and its comparator, then no concept of fluency is required. Following this, it can also avoid having to deal with metarepresentations: since they are defined as the representation of mental states as such, they necessarily require concepts about those cognitive states. Once the system does not need concepts in order to operate, then it can also do without metarepresentations.

Appealing to empirical research, Proust further shows that procedural metacognition does not need to entail metarepresentations. Research on non-human animals has indicated that procedural metacognition is likely to occur with some species of primates, which shows that "noetic feelings can monitor animals' decisions to act on a memory or on a percept, even though the concepts and percept aren't available to them". Since nonhuman animals can reliably assess their probability in resolving certain cognitive tasks while also not being able to mindread (at least not in the same conceptual way as humans do), this is a strong indication that metacognition does not require metarepresentations or concepts in order to be performed properly.

Furthermore, studies in cognitive psychology show that there is a strong relation

^{18.} Proust, The Philosophy of Metacognition: Mental Agency and Self-awareness: 14. 19. *Ibid*, 98.

between the fluency of the retrieved memory and the accuracy of the subsequent feeling of knowing. Asher Koriat, for example, states that "with regard to the accuracy of metacognitive judgements, the observation that has attracted the attention of researchers in metacognition is that participants are generally accurate in monitoring their knowledge". 20 Since these metacognitive judgements are primarily based on feelings of knowing, we can infer that the same is true of monitoring episodic memory: the cues extracted in the memory process, such as some formal memory markers, are sufficient for the formation of a subsequent feeling of remembering, which reliably prompts the agent to endorse the information as memory, not imagination. Therefore, it appears that empirical research corroborates both the procedural model of metacognition and the reliability thesis about formal and phenomenological memory markers: the former would constitute reliable heuristics to be monitored, while the latter can be seen as a good indicator to the subject about her episodic mental state. With regards to content-based markers, since they cannot be directly monitored by procedural mechanisms, neither serve as reliable input for the attributive system (as was seen in section 3.1), we can conclude that they are not reliable criteria for episodic memory.

4 Summary and Conclusion

Given that episodic memory is so phenomenologically similar to imagination, philosophers have long been debating how we can reliably distinguish them at the moment of retrieval. One of the most popular solutions is to posit certain elements that would characterize the presence of episodic recollection, called memory markers. As we saw in section 2, those can be classified into three groups: formal, content-based and phenomenological markers. Furthermore, philosophers have also debated whether memory markers can be reliable at all. The present paper divided this discussion between two opposing theses, the reliability and the unreliability theses, and showed that their main cause of disagreement is that they make different considerations about the formation of metamemory beliefs, which in turn depend on their views about metacognition in general.

The unreliability thesis considers metacognition to be an attributive and conceptual process. In this view, memory markers would serve as possible criteria for the agent to detect and form a metarepresentational belief. As we saw in section 3.1, if one uses attributive mechanisms for self-attribution of episodic memory, then formal and content-based memory markers should be rejected for they don't provide enough justification or certainty for the subject to reliably form her metamemory beliefs.

On the contrary, the reliability thesis takes metacognition to be a procedural and evaluative system, in which mental states are evaluated through a fast and subcon-

^{20.} Asher Koriat, "The subjective confidence in one's knowledge and judgements: some metatheoretical considerations", 213.

scious process. With regards to memory markers, their overall reliability would depend on the reliability of the metacognitive system, which aims to detect formal markers in the memory process and to generate an epistemic feeling based on their presence. A model of the detection of these markers was presented based on the concept of "Test-Operate-Test-Exit" units. These metacognitive mechanisms would be capable of comparing the detected fluency of the memory system against a certain threshold. If the perceived value meets or exceeds the comparator, then a feeling of recollection is generated. Given that this kind of process may also be present in non-human animals, it is more likely that complex and conceptual metarepresentations aren't present in procedural metacognition. Furthermore, findings in cognitive psychology show that the results of procedural metacognition, i.e. epistemic feelings, are largely reliable in assessing the reliability of memory processes. It was argued that these discoveries corroborate the reliability thesis about formal and phenomenological memory markers.

References

Bernecker, Sven. Memory: a philosophical study. Oxford: Oxford University Press, 2010.

Carruthers, Peter. "How We Know Our Own Minds: *The Relationship between Mindreading and Metacognition.*" Behavioral and Brain Sciences 32, no. 2 (2009): 121–38. https://doi.org/10.1017/s0140525x09000545.

Dokic, Jérôme. "Feeling the Past: A Two-Tiered Account of Episodic Memory." *Review of Philosophy and Psychology 5, no.* 3 (2014): 413–26. https://doi.org/10.1007/s13164-014-0183-6.

Goldman, Alvin. Simulating Minds: the philosophy, psychology and neuroscience of mindreading. Oxford: Oxford University Press, 2006.

Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

Koriat, Asher. "The contentive confidence in one's knowledge and judgements: some metatheoretical considerations". *Foundations of metacognition*, edited by Michael Beran, Johannes Brandl et al., 213 - 233. Oxford: Oxford University Press, 2012.

Michaelian, Kourken. *Mental Time Travel: episodic memory and our knowledge of the personal past*. Cambridge, MA: The MIT Press, 2016.

Michaelian, Kourken; Perrin, Denis. "Memory as mental time travel". *The Routledge Handbook of the Philosophy of Memory*, edited by Sven Bernecker and Kourken Michaelian, 228 – 239, New York: Routledge, 2017.

Perner, Josef. "MiniMeta: in search of minimal criteria for metacognition". *Foundations of metacognition*, edited by Michael Beran, Johannes Brandl et al., 94 – 116, Oxford: Oxford University Press, 2012.

Proust, Joëlle. *The Philosophy of Metacognition: Mental Agency and Self-awareness* Oxford: Oxford University Press, 2013.

Schacter, Daniel; Addis, Donna. "The cognitive neuroscience of constructive memory: remembering the past and imagining the future". *Philosophical Transactions of the Royal Society of London*. Series B, Biological Sciences 362, 1481 (2007): 773 – 786. https://doi.org/10.1098/rstb.2007.2087.