Review of Philosophy and Psychology (2025) 16:1085–1103
https://doi.org/10.1007/s13164-025-00768-x

On Consistently Assessing Alleged Mnemonic Systems (or, why isn't Immune Memory "really" Memory?)

David Colaço[1]

Accepted: 23 January 2025 / Published online: 19 February 2025

Abstract
How should we assess systems whose mnemonic status is contested? There are, for instance, debates over whether immune memory is "really" memory, or akin to memory as ordinarily attributed to human cognition. In this paper, I challenge two arguments often given by detractors in this debate. The first is that the system does not exhibit errors exemplified in human memory. The second is that it can be described and explained in causal terms alone. I argue that our limited knowledge of the causal basis of immunological phenomena and human memory undercuts our ability to conclude that one is merely causal while the other is not. By consequence, it is unclear why we should reject that the immune system exhibits these errors. With my challenges, I question whether we are consistent in assessing alleged mnemonic systems. Though we have schematic causal accounts of both systems, we have differing expectations of what we can conclude about them. Yet, it is unclear what might warrant appeals to these expectations.

1 Introduction[1]

If one searches for 'memory' in scientific repositories, one might be surprised to find that many high-impact papers are from the field of immunology (Roediger 2003). The use of 'memory' and its cognates results from analyses of our immune systems' ability to generate resistance to pathogens that we encountered in the past.[2] This idea has sparked the question of whether immune memory is "really" memory,

[1] This paper won the 2024 Centre for Philosophy of Memory/Review of Philosophy and Psychology essay prize competition.

[2] I focus on the human adaptive immune system. I also do not discuss reported interactions between immune and nervous systems (Marin and Kipnis 2013).

✉ David Colaço
david.colaco@lmu.de

[1] Wissenschaftstheorie und Religionswissenschaft, Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München Fakultät für Philosophie, Geschwister-Scholl-Platz 1, D-80539 München, Germany

or akin to memory as ordinarily attributed to human cognition (Matthen and Levy 1984; Valera 1994; Tauber 1997; Roediger 2003). While previous accounts have endeavored to defend that the immune system is a memory system akin to those underwritten by the brain, I choose to focus on two common arguments detractors give for rejecting this idea. The first is that the immune system does not exhibit the errors exemplified in human memory, which I call the *Error Argument*. The second is that the immune system can be described and explained in causal terms alone, which I call the *Mere Causal Argument*. These reasons are connected, with the latter serving as a reason for the former.

I argue that specific limitations in our ability to causally explain immunological phenomena raise the question of whether we can accept the *Mere Causal Argument*. I do this with appeals to immunological research on what is called "cross-reactivity" (Su et al. 2013; Brodin 2021). Given these limitations, it is unclear why we should accept the *Error Argument*. With my challenges, I question whether we are *consistent* in our assessment of these systems. Though we have schematic causal accounts of both the immune system and uncontroversial human memory systems, we have differing expectations of what we can conclude about them. We appeal, perhaps implicitly, to these expectations when offering reasons to reject that immune memory is "really" memory. Yet, it is unclear what might warrant appeals to these expectations. Thus, I use the immune system as a foil in making a more general point. I challenge the rationale for accepting uncontroversial human systems are "really" memory while simultaneously rejecting that contested systems have the same status.

In *Sect. 2*, I review points of agreement and contention about immune memory. In *Sect. 3*, I discuss immunological phenomena that seem to match human memory errors. I also unpack what we know and do not know about these phenomena. In *Sect. 4*, I present three defenses of the *Mere Causal Argument*, and I challenge these defenses. My challenges are not intended to undercut the defenses outright. Rather, I argue that these defenses can be applied to human memory systems were we to consistently assess them.

2 Debates over Immune Memory

To set up the debate into which I will insert myself, I begin by reviewing what I hope to be points of agreement as well as points of contention.

2.1 Points of Agreement

The human adaptive immune system reacts more effectively to an antigen it has encountered, a secondary response, when compared to its first exposure to an antigen, a primary response (Ratajczak et al. 2018). The immune system responds more effectively by retaining cells that responded to previous infections. These are called "memory cells" (Roediger 2003, p. 9). Roediger notes that "immunologists have recognized this fact for years and have uncovered many of the mechanisms that endow

the immune system with memory, although knowledge remains incomplete in these areas" (2003, 9).

Given the differences between primary and secondary response, it is perhaps unsurprising that immunologists frequently use mnemonic terms. The use of these terms has been common for some time, with work in the 1960s stating:

> There are known to be three types of biological memory: (a) genetic memory, the discovery and unravelling of which has been due to molecular biology; (b) conventional memory, which is a function of the brain; and (c) immunological memory (Mekler 1967, p. 481).

Thus, terms like 'memory' or 'remember' (Roediger 2003, p. 9) have been used in immunology for some time. The presence of these terms is uncontroversial. It is their meaning that is debated.

Not only is the use of mnemonic terms in immunology common, but some memory scientists – that is, scientists who work on "conventional" memory in the vein described by Mekler, which is how I refer to the relevant cases of human memory from this point on – are sympathetic to viewing immune memory as more than metaphorical. For instance, Roediger analyzes immunological phenomena that resemble priming and conscious reflection, querying "why shouldn't the immune system be said to have a memory system" and noting that "perhaps many bodily systems have a kind of memory built into them" (Roediger 2003, p. 11). Thus, not all conventional memory scientists dismiss the idea of immune memory. Further, Roediger notes that "textbook" understandings of the immune system are "somewhat simplistic," as "many mysteries remain in immune system functioning" (2003, p. 10).

At the same time, there are differences between conventional memory systems and the immune system, most notably in how one might causally interact with them. One cannot give the immune system an overt semantic task or ask it to introspect. Additionally, there are differences in their mechanistic underpinnings. Most memory scientists accept that conventional memory is implemented by synapses, while the immune system is not.[3] These differences likely have functional implications, though there are what appear to be non-superficial functional similarities, such as recognition abilities, amongst these systems as well (Kioussis and Pachnis 2009; Pradeu 2020, p. 56).

The immune system is often thought of as a causal system, responding to antigens via the production of antibodies. However, another uncontroversial (I hope) point of agreement is that conventional memory is causal. At minimum, all interlocutors in this debate are materialists. Further, we talk of our memories resulting from perception and affecting action. Finally, whether one is a "preservationist" or a kind of "simulationist" or "contructivist" about memory – broadly, reflecting whether memory involves preserved content-specific traces that are retrieved versus building them at the time of recall (Michaelian and Sutton 2017) – these accounts all express that there is causal power to memory processes (Andonovski 2021). To

[3] See Colaço and Najenson 2024 for non-synaptic accounts of conventional memory. See Pradeu 2020, p. 56 for discussion of the term 'synapse' in immunology.

deny this claim would amount to denying that memory affects behavior. Proponents of these accounts disagree on the nature of these causal relations and their epistemic importance, not whether memory is causal.

2.2 Points of Contention

In this paper, I focus on two points of contention with immune memory being "really" memory, where these scare quotes are intended to capture that there are non-superficial similarities between the immune system and conventional systems underwriting prototypical human declarative memory, which afford shared representation, explanation, and intervention.[4] These points of contention inform two challenges to immune memory being "really" memory.

First, there is the *Error Argument*: immune memory is not "really" memory because the immune system is incapable of generating the sorts of errors exemplified in human memory. In the broad strokes, memory errors, as I use the term, are mnemonic phenomena in which the process of memory occurs, but the thing that is remembered is, in some manner, not correct. This idea has been put in terms of "false memory": "remembering events that never happened, or remembering them quite differently from the way they happened" (Roediger and McDermott 1995, p. 803). One can place these errors into different categories, such as confabulation, which more resembles the first portion of Roediger and McDermott's quote, and misremembering, which more resembles the latter portion (Robins 2016b; Michaelian 2016).

The argument holds that the immune system does not exhibit these sorts of errors and for this reason is not "really" a memory system. Here, a toy example from Adams (2018), though not specifically about memory, is a helpful starting point. Imagine one has a garage door sensor. Either the sensor detects something and keeps the door from closing, or it does not, and the door's operations occur uninhibited. Adams and Garrison claim that for the internal states of a system to rise to the level of cognition, "there must be something that can be falsely tokened, that can misrepresent; that can represent non-actual states of affairs" (2013, p. 350).[5]

Put another way, and specifically in the context of immune memory, the "mechanism that leads from the environment to the internal representation of that environment permits the internal representation to be false, even when the mechanism is undamaged" (Matthen and Levy 1984, p. 366). This idea, paired with Adams' toy example and applied to cases relevant to this paper, is that immune memory more closely matches the operations of the garage door than those of conventional

[4] This reading of "really" brackets types of memory, such as procedural memory, that can have parallel arguments levied against them, noting briefly that this is debatable (see, e.g., Retkoceri 2021). Regardless, I focus on comparing immune memory to declarative memory, as it seems to be the type of memory interlocutors in this debate have in mind. If arguments like these are inconsistently applied even for different types of human memory, it further supports my thesis.

[5] My transition between cognition and memory when presenting variants of these arguments is supported by Adams and Garrison's claim that "everyone agrees and knows that memory… [is a] cognitive process" (2013, p. 340). Like others, they take conventional memory to be cognitive.

memory. Like the door sensor, the argument suggests, the immune system cannot err in the way conventional memory systems can. It cannot err in the manner described by Matthen and Levy.

Second, there is the *Mere Causal Argument*: immune memory is not "really" memory because, unlike conventional memory, immunological phenomena can be described and explained in causal terms alone. Though everyone in this debate agrees that memory processes are causal, this argument presupposes that conventional memory cannot be described and explained in causal terms alone. Robins highlights this point for memory errors: "false memories are *misrepresentations*" (Robins 2016a, p. 1562, my emphasis). This argument in part falls out of a more general variation regarding the idea that cognitive systems are more than merely causal. One must, this argument indicates, invoke internal states or latent variables to these systems, capturing that these systems involve something like the processing of internal representations (see, e.g.s, Adams and Aizawa 2001; Rowlands 2010). These latent variables capture that the system in question manipulates and transforms content.

Many things that might seem cognitive (or specifically mnemonic) can be described and explained in causal terms alone, this argument continues. For instance, Adams and Garrison claim that the behavior of bacteria might initially seem cognitive, but "on closer inspection there are complete explanations (none of which constitute cognitive processing)" (Adams and Garrison 2013, p. 341). They conclude that "there are non-representational explanations of why they do what they do," such as chemical or physical explanations that are causal, while "the explanation of cognitive behavior includes the representational content of the internal states" (Adams and Garrison 2013, p. 346).

This reasoning supports the more specific variation of the argument that unlike conventional memory, the immune system can be described and explained in causal terms alone. As Melander puts it in the context of immune memory, "immunology does not and need not employ intentional explanation or description" (1993, p. 224). Here, 'intentional' captures that one must appeal to content or representation to satisfactorily describe or explain the phenomena that occur in a system. This is consistent with the more general variation of the argument from Adams and Garrison. One could use these terms regardless of whether they are necessary, but nothing about describing or explaining immunological phenomena is lost if one were to omit these terms.[6] One might thus eliminate them or treat them as metaphors, opting instead for more parsimonious causal descriptions and explanations.

These two arguments are related. The latter is often given as a reason for the former. Immunological phenomena that look prima facie like memory errors, when understood causally, are either the outcome of an operative causal system that produces an effect given an input, or this causal system has been perturbed or damaged. Either way, this is a purely causal explanation of the phenomenon, and hence any

[6] Whether something can be described or explained in causal terms alone thus serves as a "null hypothesis," akin to how associative learning is a "null hypothesis" in comparative cognition rather than a proper null hypothesis in statistics (Halina 2022).

discussion of it as an "error" is our imposition on the system. Thus, if the immune system is purely causal, then immune memory is not "really" memory. Conventional memory errs because memory has content that can be incorrect. This is possible because memory systems lose a "tight informational connection with truth" (Adams 2018, p. 28).

In this paper, I needle this argument pair via appeals to what we know and what we do not know about immunological phenomena. My aim is not to argue that the immune system is a memory system. Rather, I use it as a foil to expose the limitations of the selective application of these arguments. Before turning to my case, however, I acknowledge that not all philosophers agree on the necessity of representations in accounting for memory (see, e.g., Caravà 2021). I imagine even fewer agree on what 'representation' means. For those readers who are not sympathetic to these arguments, I hope that you will humor me in my attempts to needle them.

3 Some Immune "Memory" Phenomena

To needle the arguments I have presented, I introduce immunological phenomena that stem from research in the past decade and a half. After introducing them, I discuss how they relate to our understanding of the immune system as a memory system.

3.1 Immune "False Memory" and Cross-Reactivity Studies

During primary response, B and T cells bind to the antigen in question. Following primary response, some active B and T cells are transformed and persist in the body. These are the "memory" cells that will activate in secondary response. This is thought to be achieved via some binding between the molecular and geometric features of the antigen surface, the epitope, with those of the memory cell, the paratope. The persistence of memory cells varies, but some, such as smallpox memory cells, can persist for a human's lifetime (Hammarlund et al. 2003).

One might expect memory cells for an antigen only in immune systems that have undergone primary response to this antigen. This is not the case. For instance, Su and colleagues searched for memory T cell markers in the samples of adult blood-bank donors via exposure to antigens (2013). Though these donors likely have never been exposed to viruses like HIV, the researchers identified memory T cells for them. As a control, they tested newborns, in which they could not find them. Thus, some adults have memory cells for antigens that are not dependent on primary response to these antigens. Though I focus on a single set of studies, they represent a body of research, including more recently in COVID-19 immune responses (Brodin 2021). In a short report on Su and colleagues research, Douglas Heaven provocatively calls these results "false memories" (2013, p. 16).

The researchers aim to explain their results in terms of "cross-reactivity with many antigens in the environment," (Su et al. 2013, p. 374) or the ability of memory cells to respond to more than one peptide binding locus. Su and colleagues claim:

> There is a clear tendency on the part of at least some [T cell receptors] to recognize distinctly different peptide-MHC [major histocompatibility complex] combinations, thereby enabling the same T cell to be activated by different ligands (Su et al. 2013, p. 380).

In other words, memory cells respond to antigen peptide chains that resemble the antigens the immune system has encountered. Thus, it is not just the case that antigens' peptide chains match, leading to one T cell having the same reaction to two distinct antigens. Epitopes and paratopes are not all-or-nothing lock and key mechanisms. Rather, T cells can be activated by different combinations of peptide chains, and antigens might share some of these factors.

To defend their explanation, Su and colleagues performed a controlled intervention. They administered a vaccine for the H1N1 flu to individuals who had not received a flu vaccine for five years. They analyzed the participants' blood before and after the administration. They validated their study by measuring an increased immune response to the H1N1 virus, and they found that the assayed post-administration blood samples, but not pre-administration samples, contained memory T cells that respond to different bacteria. Likewise, Su and colleagues suggest that a similar effect can be induced in humans with algae exposure stimulating HIV-1 reactive clones (378). The authors claim, "arguably the most intriguing finding from this characterization of human T cell precursors is the abundance of memory-phenotype T cells in healthy adults for foreign antigens that those individuals have never encountered" (Su et al. 2013, p. 379).

Su and colleagues speculate upon the advantages of cross-reactivity in the human immune system. They suggest that "cells specific to novel pathogens might confer an advantage for survival against infections" (Su et al. 2013, p. 381). Thus, there might be an etiological explanation for memory T cells that are antigen-specific but to antigens the host has not encountered. However, this suggestion is at best a speculation. Cross-reactivity's immunological importance remains unclear, as does the frequency of cross-reactions in a human's life (Petrova et al. 2012). What seems more supported is cross-reactivity's role in the pathogenesis of autoimmune diseases (Su et al. 2013, p. 380).

3.2 What We Know; what We Don't

Aside from the fact that this case involves the use of mnemonic terms, there are a few things worth pointing out about it and the immunology research that it represents. First, there is a causal story that the researchers present, but they recognize that their causal knowledge is schematic. Cross-reactivity is an attempt at a causal explanation for the phenomenon that researchers have putatively demonstrated. Indeed, appeals to older cross-reactivity research were a point of contention between Matthen and Levy (1984) and Melander (1993) several decades

ago. In this debate, Matthen and Levy appeal to deleterious cross-reactivity cases as a defense for adopting an intentional stance towards the immune system (1984). Melander responds, "purely chemical and hence nonintentional characterizations of the immunological processes that are described in intentional terms are beginning to emerge," noting that a "purely chemical characterization of the mechanism… will be available as soon as the exact chemical composition of the epitopes… have been worked out" (1993, 238). Melander's claim partially bears out: the mechanistic study of the immune system has advanced. Yet, this case and more recent studies (Brodin 2021) indicate that we have yet to arrive at a complete causal account of immunological phenomena. Attempts to bottom out cross-reactivity cases in terms of similarities between epitopes are not unreasonable. However, determining a coherent similarity metric in terms of molecular and geometric features – that is, determining *how* epitopes are similar – remains challenging (Petrova et al. 2012; Peters et al. 2020).

Perhaps we will never develop a "complete" causal story of the immune system, or of any system for that matter, and the causal story we tell will always be "schematic." This ties into my second point: a secondary response stemming from the interaction between memory cells and antigens seems to be complicated by a myriad of interacting factors, creating a situation with variability that looks like the immune system has multiple degrees of freedom. As cross-reactivity shows, there is not a one-to-one relation between epitope and paratope. The degree to which the relation between epitope and paratope will result in a secondary reaction seems to be sensitive to, amongst other things, the strength of the chemical bond between them, the number of binding sites each has and the variable relations between these sites, and the states of assemblages of other immune cells and antigens internal to the system both before and at the time of the reaction in question. The system, when operating reliably, results in matches and what seem to be mismatches between the system and the antigens that are (at least currently) difficult to explain in molecular terms. In this case, straightforward attempts to intervene on molecular details in cross-reactivity cases are also limited. This limitation highlights how schematic the causal story remains some 30 years after Melander's publication. I also encourage the reader to remember that a mechanistic account of conventional memory is present in neurobiology research (Kandel et al. 2014). This account, I hasten to add, is also schematic.

This case highlights a third point about the causal story: immunologists pursue increasingly complex causal explanations for the phenomena that they study. Though they readily use mnemonic terms to account for the phenomena that they investigate, they are nonetheless interested in developing causal accounts of the immune system. They are interested in improving their causal schemata, though this interest does not seem to come paired with any attempts to eliminate their use of mnemonic terms. Again, I add that this is not unique to immunologists. Researchers who study conventional memory also wish to improve their causal explanations, yet they do not try to jettison mnemonic terms from their lexicon.

This example of so-called "false memory" in the immune system is thus fascinating to me not because I think it is a straightforward case of a memory error. Rather, it is fascinating to me because analyzing what we know and what we do

not (yet) know renders the case much more akin to past and current states of conventional memory research. This raises the question of whether our assessment of one requires us to reassess the other.

4 Three Defenses for the *Mere Causal Argument*

With some immune cases introduced for context, I ask: given that our causal understanding of the immune system is limited, what defense can be given for the *Mere Causal Argument*? In this section, I review three defenses that might be given for this argument. My challenges to these defenses are not that they are intrinsically deficient. Rather, they revolve around inconsistences between how we judge the immune system and conventional memory. Thus, I phrase these defenses as why they support the argument for the immune system but not for conventional memory systems. I argue that, if these defenses hold water, they have parallel implications for conventional memory systems.

Before I address these defenses, I note three that I will not address at length. First, I will not address the idea that our causal knowledge of the immune system is complete in any naïve sense. This would involve quibbling over the exact details of cases, which I suspect is not what motivates selective philosophical sympathy to the *Mere Causal Argument*. Second, I will not address any phenomenological dimension to human memory that one might argue is absent from the immune system, as I am uncertain what it is like to be an immune system. Personal correspondences tell me that *people* who suffer from an autoimmune disease have phenomenological experiences of it, suggesting that there is some argument to be made that these experiences are not entirely absent. Third, I will not address any argument related to whether memory is a natural kind. Though a worthy topic, it informs a debate about "unconventional" memory that is distinct from the *Mere Causal Argument*.[7] With these potential defenses put aside, let me turn to those that I suspect will interest those sympathetic to this argument.

4.1 The Plausible Explanation Defense

The first defense for the *Mere Causal Argument* is that, while our causal knowledge of the immune system is schematic, we can nonetheless plausibly explain these phenomena with nothing but causal factors. To make this defense concrete, I appeal to Craver's notion of a "how-plausibly model." These are mechanistic schemata that are "more or less consistent" with what is known, but they may include incorrect parts or "filler terms" (Craver 2007, pp. 112–113). How-plausibly models are more developed than sketches, but they are not "how-actually models," which describe "real components, activities, and organizational features of the mechanism that in fact produces the phenomenon" (Craver 2007, p. 112). Applying this framework to

[7] See Colaço 2022 for discussion of this debate.

the defense, we might "black box" factors in our current explanations, but in doing so, we can account for the immune system in causal terms alone.

This is not the case for conventional memory, this defense continues. We cannot give a how-plausibly model of conventional memory phenomena in causal terms alone, and it is not because the causal details are currently poorly understood. Thus, the defense does not suggest that we have a how-actually model for the immune cases. The science has yet to deliver one. Nonetheless, all that we lack for immune cases is causal detail, this defense suggests. What we have in causal detail is sufficient for the aims of immunologists in formulating a how-plausibly model in causal terms alone.

In response to this potential defense, I remind the reader that the schematic character of our causal knowledge of the immune system does not seem to be one in which we currently lack causal detail but otherwise can provide a how-plausibly model. The details of the case in *Sect. 3* suggest that the limitations of our knowledge are more than just lacking causal detail for two reasons. First, our limited causal knowledge makes it difficult to predict in molecular terms when scientists will achieve a cross-reactivity effect, which is one reason the research is of interest to immunologists. This reappraisal of expectations based on new evidence suggests that the causal processes are not well-understood even at a coarse-grained level, except in the sense that there is a cross-reaction between epitopes and paratopes. Difference makers seem to lie beyond our causal knowledge, raising the question of the viability of anything but a very coarse sketch of the causal operations of the immune system during the occurrence of these phenomena. Second, our limited causal knowledge makes it difficult to intervene upon molecular causal mechanisms to induce, inhibit, or modulate the occurrence of these phenomena, raising concerns about the adequacy of our causal knowledge of the system.

Perhaps it is poetic in this paper that is in part about the *Error Argument* that my response to this defense is akin to an argument from error. Currently, our knowledge of the immune system is schematic, but it is also the case that the immune system resists being explained in terms of straightforward causal factors. Over the decades, what has sometimes been presented as a molecular story of immune memory has required increasing sophistication for its modeling, explanation, and control, as the details in *Sect. 3.2.* indicate and the Roediger quotes in *Sect. 2.1.* suggest. My appeal to wanting causal accounts of the immune system also helps me block a defense from the conceivability of a causal explanation for immunological phenomena. The immune system's properties have proven to be unintuitive, and they resist elucidation in terms of our concepts alone. It might feel like we can conceive of a purely causal system, but this conception runs counter to properties we have learned about the immune system, including the immunological phenomena I have reviewed in this paper. Thus, the idea that we feel as though we can conceive of a purely causal immune system does not seem a reliable indicator of whether it *is* a purely causal system.

My response is not intended to cast a pessimistic outlook on the state of immunology. Nonetheless, the schematic character of our causal knowledge of what underwrites these immunological phenomena, paired with challenges scientists have faced

when trying to molecularly model the system for purposes of prediction and intervention, results in this defense being far less weighty than it might initially appear to be. A schematic causal story can be given for immune phenomena. I do not deny this. However, the idea that a plausible causal explanation we can give for immune phenomena is somehow different in kind from a plausible causal explanation we can give for conventional memory phenomena is far less clear, and it is this conclusion that is needed to warrant selectively accepting the *Mere Causal Argument*.

4.2 The "Nothing Lost" Defense

The second defense for the *Mere Causal Argument* starts with the idea that there is value in describing and explaining conventional memory in mnemonic terms. If one were to describe or explain these phenomena in causal terms alone, something would be lost. By contrast, the defense continues, there is nothing lost when immunological phenomena are described or explained in causal terms alone.

Let me set up what is ostensibly lost in conventional cases and its relation to the *Error Argument*. To do this, let me review the Deese-Roediger-McDermott (DRM) experimental paradigm, which is used to induce a characteristic memory error (Roediger and McDermott 1995). In DRM, participants receive a list of semantically related words, such as "bed, rest, awake." The list items relate to a lure not on the list: these items might have the lure "sleep." The participant is asked whether items, including the lure, were on this list. The aim of the task is to determine participants' tendency to answer that the lure was on the list. If the tendency is high, researchers conclude that they have a false memory of this lure being on the list.

DRM is a popular method, as it "is one of the best-established memory errors in the empirical literature" (Robins 2016b, p. 433), and the "errors produced in DRM… are not easily dismissed as contrived laboratory tasks" (Robins 2016a, p. 1565). The phenomenon that is elicited in the DRM paradigm, regardless of whether it is properly 'false' memory, putatively cannot be satisfactorily explained in causal terms alone.[8] This is because of the perceived need for us to appeal to stored content or internal information processing to make sense of the semantic character of the error. What seems at fault in these cases relates to what these terms *mean*, rather than any similarity between their strings of letters. In this sense, it is different from direct priming (Schacter and Buckner 1998, p. 186). Their differences can be illustrated via the DRM list items. 'Sleep' is a lure for list items because of their semantic similarities. In a direct priming task, we might present items like 'sl—p.' Direct priming involves the perceptual properties of the stimulus, rather than their content. Thus, the idea behind this defense is that something is lost when one tries to eliminate content from explanations for DRM task behavior, while nothing comparable is lost when one eliminates it from explanations for direct priming.

Following on this distinction, a defender might be willing to cede that immune cases match or at least resemble direct priming, but even if they do, nothing is lost

[8] Lures are not wholly unrelated to list items, suggesting that "the memories produced are not entirely false" (Robins 2016a, p. 1576).

in eliminating content from the explanation for these immunological phenomena. Extrapolating, the idea is that if conventional memory phenomena involve content, we can eliminate mnemonic terms from our descriptions or explanations of immunological phenomena because we need not appeal to content. To be as charitable as possible, this defense raises the challenge that it is at least unclear what is lost if we do not appeal to content.

Before responding to the defense, I note that if we aim to account for immunological phenomena in terms of priming, it seems unlikely that this account can be in terms of direct priming. Cross-reactivity studies show that the stimuli need not be the same. Instead, they need only be similar, where the exact sense in which they are similar is not yet known. Perhaps the objector can defend the idea that the H1N1 case more closely matches indirect or "conceptual" priming, where direct priming is modality specific and indirect priming can involve changes in the modality of the stimulus (Schacter and Buckner 1998, p. 186). Yet, this does not aid the objector, as indirect priming is seen as a memory phenomenon and may involve content (Tulving and Schacter 1990; Schacter and Buckner 1998).

With this point out of the way, my first blush response to this potential defense is that, in the immune case, we have phenomena that are characterizable as errors, which motivates providing mnemonic explanations for them. However, this first blush response is a nonstarter in this debate, as the *Error Argument* denies that the immune system is capable of memory errors in part because of the *Mere Causal Argument*. There is disagreement about how we ought to explain immunological phenomena, and there is disagreement about how we ought to describe them. While this disagreement rephrases the central debate that I discuss in this paper, I include it because it shows that the would-be defender seems to conclude something about how we ought to characterize immunological phenomena based on how they think they ought to be explained. This seems to invert the typical order of operations. According to accounts of explanation, one aims to explain after one has characterized the explanandum phenomenon (Craver 2007, p. 128). The order of operations is followed in conventional memory cases, where we (say) perform a DRM study, elicit a phenomenon from it, characterize it, and find that we cannot explain it in causal terms alone. Inverting this order creates problems, as an explanation's adequacy is determined by its fit with a phenomenon's characterization. Arguing that a phenomenon must be recharacterized because of how it is explained thus raises the question of how one determined that this explanation is adequate for this phenomenon in the first place (see Colaço 2020).

A more reflective response to this potential defense begins by highlighting that whether something is lost in explaining immunological phenomena in causal terms alone seems highly dependent on our knowledge of the causal factors that underwrite these phenomena. Since, I wager, our knowledge of these factors is schematic and past attempts to explain in causal terms alone are perceived to be inadequate, it is difficult to make sense of what might be gained or lost by explaining in causal terms alone. This response raises the question of what is needed for us to know whether something is lost via a purely causal description or explanation.

At the same time, it highlights a benefit of the use of mnemonic terms for describing and explaining immunological phenomena in the face of this limited knowledge.

Using these terms gains us at least one thing, which is legibility. By 'legibility,' I have in mind the idea that thinking of immunological phenomena as memory phenomena allows us to account for these phenomena productively and coherently, even if we lack a satisfactory causal explanation for them. While legibility might seem at best a mere shorthand for talking about immunological phenomena, it can play a productive role in the formulation and test of claims about these phenomena. It can, for instance, inform predictions and interventions about the reaction of the immune system to antigens it has and has not been exposed to in the past.[9] These predictions and interventions can be informed by treating immune memories as internal states whose content are related to antigens, which are correct when it has been exposed to the antigen and incorrect when it has not. Legibility might not be ideally reliable, but it is nonetheless deployable when a causal explanation is not (yet) satisfactory.

This legibility parallels how we approach conventional memory. We might feel that we need not unpack the mechanistic details of memory, such as synaptic plasticity, when we want to explain the difference between veridical and false memory. Rather, we can appeal to the mnemonic relation between the source of the memory and the content that is retrieved or recalled in these cases. If we were to eliminate this mnemonic framing from conventional cases, we would lose our ability to account for them. Our current understanding of the mechanistic underpinnings is schematic, despite advances in the past half century. We can make the same point for the immune system, I wager. Eliminating any mnemonic stance towards the immune system would lose us legibility, and it would leave us with an incomplete causal story that limits our ability to predict and intervene.

The reader might recognize that my discussion of the benefits of legibility for memory systems resembles David Marr's motivations for his levels of explanation when accounting for vision (1982). At the time Marr developed his account, he perceived vision science to have hit a wall despite advances in causal explanations of vision phenomena.[10] His levels of explanation provide framework for different analyses of this system, and some of these analyses are not in causal terms alone. I imagine that scholars sympathetic to Marr do not need me to tell them that these analyses have been profitable. Though it might not be the case that a parallel wall has been hit in immunology, a comparison can be made between the two. Mnemonic framing of immunological phenomena likewise provides a way to think about different analyses one could perform on the immune system, and these analyses likewise can be profitable.

Productive legibility is something that will be lost were we to describe and explain immunological phenomena in causal terms alone. It is thus unfair to say that *nothing* would be lost. At minimum, it is unclear why nothing would be lost for the immune system, yet something would be lost for conventional memory. Any materialist must cede that there is a causal story to tell about every conventional memory phenomenon, even though we cannot give this story at this time. Mnemonic framing

[9] Representational modeling of artificial immune systems also exists (McEwan and Hart 2009).

[10] A similar point can be made for George Miller, who found that he needed to appeal to information theoretic terms to make sense of human limitations in number recall (1956).

adds something to this story that is needed for us to explain these phenomena, this defense states. Correspondingly, mnemonic framing adds something to the causal story in describing and explaining immunological phenomena in the face of our limited causal knowledge. Someone sympathetic to this defense must provide reason to accept the conclusion that a mnemonic framing of conventional memory will be required regardless of the causal story that is developed, but a mnemonic framing of the immune system will not be required.

4.3 The Progress Defense

The third defense for the *Mere Causal Argument* starts with the idea that the aim of immunology is to generate knowledge of the immune system, which will help us in the prediction, explanation, and control of this system. We should, the defense continues, adopt a stance towards the immune system that most benefits and least hinders our ability to generate this knowledge. The field has made advancements via causally modeling the immune system, regardless of the current schematic character of our explanations. We should therefore approach this system as a mere causal one, as scientific advancement can be made through observation and experimentation based on our causal models of it. We can adopt this approach as a working hypothesis that orients the research, keeping whatever mnemonic terms we use as a metaphor. If we do otherwise, attempting to build in insights from conventional memory, we might undercut research on the causal basis of the immune system. Thus, this defense concludes, we should strictly adopt a "nonintentional" physical stance towards the immune system. This physical stance would match Dennett's idea that we sometimes focus on the physical, causal states of the system (Dennett 1971), to which I add that those who adopt the stance make an active effort to disregard any implications of the intentional or specifically mnemonic aspects of this system.

This defense is unlike the others because it does not defend the correctness of the *Mere Causal Argument* per se. Instead, it defends the appropriateness of adopting a physical stance as a working hypothesis for best orienting immunological research. While the epistemology of the situation is different, it has the same practical consequences. We should, according to this defense, not treat immune memory as "really" memory because doing so might undercut the advancements we can make by investigating it from the stance that it can be explained in causal terms alone. Correspondingly, we should investigate immunological phenomena as causal interactions and not as errors, thereby providing a defense for the *Error Argument*. This defense indicates that adopting a unified, physical stance that is nonintentional is the best option for making progress in the study of the immune system. Unlike Marr's take on the study of vision, immunologists do not indicate that they think that they have hit a wall in their research, so proceeding in the productive manner of determining causal interactions might also make this defense appealing to practicing immunologists.

In response to this potential defense, I start by noting that it is appealing because it focuses on the positive dimension of immunology. It highlights that approaching the system in causal terms is a reason the field has made advancements. These

advancements are tangible: they afford us new vaccinations and treatments for diseases. It also highlights that something is gained when the causal details of the immune system are investigated. Moreover, it fits into a broader narrative, where advances in many biological systems have resulted from a causal and specifically mechanistic investigative approach. I do not deny these positive contributions to immunology, so my concerns with the defense do not lie here.

My response instead focuses on the question of why adopting a nonintentional physical stance towards the immune system is warranted but adopting this stance towards conventional memory is not. If the *Mere Causal Argument* is ultimately supported by the manifest progress that has been and can be made, then it is fair for me to point out that our knowledge of the neural basis of memory has also advanced via investigating this system in causal terms. For just one example, tools like optogenetics have advanced our ability to study mnemonic mechanisms of encoding, storage, and retrieval, though how we ought to incorporate these insights into our explanations of memory remains up for debate (Colaço and Robins 2023). More generally, it seems difficult for one to deny that our causal knowledge of conventional memory has increased through pursuit of these systems in causal terms. If this progress is not indication that we should construe human memory systems in causal terms alone, some additional argument must be given for why progress entails that we should construe the immune system in this way. This argument cannot amount to either defense I have already discussed in this section, as both defenses seem to apply as well to the immune system as they do to conventional memory.

The fact that we have made progress in studying the immune system in causal terms, akin to how we have made progress studying conventional memory in causal terms, reflects that, returning to *Sect. 2.2.*, both systems are thought to be causal. However, if this defense is one for the idea that we can explain the immune system in causal terms alone, then it is hard to motivate this defense without it being directly applicable to other causal systems. This includes conventional memory systems.

To break this symmetry, one might argue that we can make much more progress adopting this stance towards the immune system versus adopting it towards conventional memory systems. Yet, this argument would reiterate the very point under debate in this paper. This argument would thus demand a reason we can make more progress on some systems rather than others. Further, it would be very hard to evaluate, given our schematic causal knowledge of both systems. One cannot take causal success to reflect some fundamental ontological character of the immune system – that, for instance, it really is just a causal system – and simultaneously deny that causal success reflects the same ontological character of conventional memory systems. For this reason, this defense does not support the *Mere Causal Argument* without having parallel implications for conventional memory.

4.4 The Defenses and the Selective Application of the Arguments

The three defenses that I have addressed are not intrinsically deficient. However, they all seem to apply just as well to arguing that we can explain conventional memory in causal terms alone. If we want to consistently assess these systems, it seems

that we need to apply these defenses in equal measure. By consequence, it seems that we need to apply the *Mere Causal Argument* and the *Error Argument* in equal measure as well. This consistency of assessment stems, at its source, from the fact that we have schematic causal accounts of both systems. The specific limitations in our causal knowledge ultimately complicate a selective conclusion of the *Mere Causal Argument* for the immune system but not for conventional memory systems.

While other defenses for the *Mere Causal Argument* are possible, my needling of the *Mere Causal Argument* and its defenses leads me to suspect that a set of intuitions drives expectations for selectively accounting for the immune system in causal terms alone. These four intuitions are as follows. First, philosophers might not realize how complex the immune system is, how schematic our causal knowledge of it is, and how simplified textbook understandings of it can be. This lack of realization is akin to what Roediger's quotes identify in *Sect. 2.1.*, albeit directed at philosophers rather than at memory scientists. A complex system might be explained in causal terms alone, but if philosophers construe the immune system as simple compared to the brain and mind, they might think that our ability to causally explain the former is fundamentally more viable. Second, there are causal differences between the immune system and conventional memory systems. As I mentioned in *Sect. 2.1.*, both the studies researchers perform and the mechanisms they schematize are different. These differences no doubt have some implications for the functions of the respective systems. Yet, it is unclear why this alone should warrant that one but not the other is properly mnemonic, unless one holds a hard reductionist position about memory. It would, by consequence, run afoul of enduring commitments to multiple realizability, as one could argue that both the immune and nervous systems realize memory systems. Third, the immune system does not remember as "we" do, in the sense that we have lived experience of conventional memory but not of immune memory (Nemati 2024). However, the fact that our experiences of memory are of conventional memory does not, on its own, give us reason to deny immune memory any more than it gives us reason to deny that non-human animals have any sort of memory. Fourth, there might remain some implicit leanings towards mind-brain dualism, which has no analog in the study of the immune system. It is, for instance, unclear to what a personal versus sub-personal distinction in the immune system would amount. However, one would need to adopt an ontological reading of this distinction to defend this intuition. Moreover, any dualist leaning runs afoul of materialist commitments.

Perhaps from these intuitions one could derive an epistemically cogent defense for selectively arguing that the immune system can be explained in causal terms alone. Nonetheless, I am skeptical that such a defense can be mustered until much more is known about the causal bases of the respective systems. As they stand, these intuitions account for why we might have different expectations about what we can causally conclude about these systems. They do not, however, provide warrant for these expectations.

5 Conclusion

In this paper, I hope that I make clear that I do not argue that the immune system is "really" a memory system, nor do I argue that human memory systems can be described and explained in causal terms alone. If the reader takes my challenges to key arguments against immune memory as an indirect case for these positions, so be it. What matters to me is that we do not miss the forest for the trees. I have taken as my target the argument that some alleged mnemonic systems, such as the immune system, are not "really" mnemonic because they do not exhibit errors and can be described and explained in causal terms alone. I have investigated the difficulties in providing warrant for these arguments when applied selectively to the immune system, given the specific limitations of our causal knowledge of it as well as of conventional memory systems. My point is that we cannot argue that the immune system can be described and explained in causal terms alone, but conventional memory cannot be, unless we give up on being consistent in how we assess these systems. Either we ought to be consistent, or we ought to show why inconsistency is warranted. No defense I have addressed warrants this inconsistency, and any expectation that they are inconsistent seems unwarranted or at least premature.

In my estimation, an inconsistent appeal to the *Mere Causal Argument* flounders because any memory system is a causal system on a materialist account of the mind. They can therefore be described and explained in causal terms. The trouble comes when we try to move from this uncontroversial point to one about whether a system can be described and explained in causal terms *alone*. If one is sympathetic to construing memory as something that cannot be explained in causal terms alone, it seems that it is also profitable to think of the immune system in mnemonic terms, just as it is profitable to think of it in causal terms. One might even reconsider whether there is a dichotomy between mnemonic and purely causal explanations, where having the latter entails that the former is unnecessary.[11] Correspondingly, it seems that it is profitable to think of some immunological phenomena as memory errors. If we approach the use of these terms as positions that we might reconsider based on new evidence for or against a causal story for an alleged memory system, then we need not make a realist commitment regarding whether (say) immune memory is "really" memory. I suggest that we adopt an open mind to these kinds of positions. Correspondingly, I suggest that we explore the idea that the alleged memory systems exhibit errors akin to human memory. This, if nothing else, is consistent.

Acknowledgements Thanks to Jonathan Najenson and two anonymous reviewers for feedback on an earlier draft of this paper.

Funding Open Access funding enabled and organized by Projekt DEAL. This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 528370787.

Data Availability This research involved no collection of data, and there are no materials or code to be made available.

[11] See Hochstein (2013) for a parallel argument with intentional and statistical models.

Ethical Approval Ethics approval is not applicable to the content of this research.

Conflict of Interest The author has no relevant financial or non-financial interests to disclose. There are no conflicts of interest to declare.

References

Adams, Fred. 2018. Cognition wars. *Stud Hist Philos Scien Part A* 68: 20–30.
Adams, Fred, and Ken Aizawa. 2001. The bounds of cognition. *Philosophical Psychology* 14 (1): 43–64.
Adams, Fred, and Rebecca Garrison. 2013. The mark of the cognitive. *Minds and Machines* 23 (3): 339–352.
Andonovski, Nikola. 2021. Causation in memory: Necessity, reliability and probability. *Acta Scientiarum. Human and Social Sciences* 43 (3): e61493.
Brodin, Petter. 2021. Immune determinants of COVID-19 disease presentation and severity. *Nature Medicine* 27 (1): 28–33.
Caravà, Marta. 2021. An exploration into enactive forms of forgetting. *Phenomenology and the Cognitive Sciences* 20 (4): 703–722.
Colaço, David. 2020. Recharacterizing scientific phenomena. *European Journal for Philosophy of Science* 10: 1–19.
Colaço, David. 2022. What counts as a memory? Definitions, hypotheses, and kinding in progress. *Philosophy of Science* 89 (1): 89–106.
Colaço, David. 2024. Where memory resides: Is there a rivalry between molecular and synaptic models of memory? *Philosophy of Science* 91 (5): 1382–1392.
Colaço, David, and Sarah Robins. 2023. Why have revolutionary tools found purchase in memory science? *Philosophy and the Mind Sciences* 4.
Craver, Carl. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press
Dennett, Daniel. 1971. Intentional systems. *The Journal of Philosophy* 68 (4): 87–106.
Halina, Marta. 2022. Unlimited associative learning as a null hypothesis. *Philosophy of Science* 89 (5): 1186–1195.
Hammarlund, Erika, Matthew W. Lewis, Scott G. Hansen, Lisa I. Strelow, Jay A. Nelson, Gary J. Sexton, Jon M. Hanifin, and Mark K. Slifka. 2003. Duration of antiviral immunity after smallpox vaccination. *Nature Medicine* 9 (9): 1131–1137.
Heaven, Douglas. 2013. False memories prime immune system for future attacks. *New Scientist* 217 (2904): 16.
Hochstein, Eric. 2013. Intentional Models as Essential Scientific Tools. *International Studies in the Philosophy of Science* 27 (2): 199–217.
Kandel, Eric R., Yadin Dudai, and Mark R. Mayford. 2014. The molecular and systems biology of memory. *Cell* 157 (1): 163–186.
Kioussis, Dimitris, Vassilis Pachnis. 2009. Immune and nervous systems: More than just a superficial similarity? *Immunity* 31(5): 705–710.
Marin, Ioana, Jonathan Kipnis. 2013. Learning and memory… and the immune system. *Learning & Memory* 20(10): 601–606.

Marr, David. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman and Company.
Matthen, Mohan, Edwin Levy. 1984. Teleology, error, and the human immune system. *The Journal of Philosophy* 81(7): 351–372.
McEwan, Chris, Emma Hart. 2009. Representation in the (artificial) immune system. *Journal of Mathematical Modelling and Algorithms* 8 (2): 125–149.
Mekler, L.B. 1967. Mechanism of biological memory. *Nature* 215 (5100): 481–484.
Melander, Peter. 1993. How not to explain the errors of the immune system. *Philosophy of Science* 60 (2): 223–241.
Michaelian, Kourken. 2016. Confabulating, misremembering, relearning: The simulation theory of memory and unsuccessful remembering. *Frontiers in Psychology* 7: 1857.
Michaelian, Kourken and John Sutton. 2017. Memory The Stanford Encyclopedia of Philosophy. Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2017/entries/memory/. Accessed 27 May 2024.
Miller, George. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63 (2): 81.
Nemati, Nedah. 2024. Rethinking neuroscientific methodology: Lived experience in behavioral studies. *Biological Theory* 19: 184–197.
Peters, Bjoern, Morten Nielsen, and Alessandro Sette. 2020. T cell epitope predictions. *Annual Review of Immunology* 38: 123–145.
Petrova, Galina, and Andrea Ferrante and Jack Gorski. 2012. Cross-reactivity of T cells and its role in the immune system. *Critical Reviews in Immunology* 32 (4): 349-372.
Pradeu, Thomas. 2020. *Philosophy of immunology*. Cambridge University Press.
Ratajczak, Weronika, Paulina Niedźwiedzka-Rystwej, Beata Tokarz-Deptuła, and Wiesław Deptuła. 2018. Immunological memory cells. *Central European Journal of Immunology* 43 (2): 194–203.
Retkoceri, Urim. 2021. False procedural memory. *Philosophical Psychology* 34 (3): 397–423.
Robins, Sarah K. 2016. Optogenetics and the mechanism of false memory. *Synthese* 193 (5): 1561–1583.
Robins, S.K. 2016. Misremembering. *Philosophical Psychology* 29 (3): 432–447.
Roediger, Henry L. 2003. Reconsidering implicit memory. In: Bowers JS, Marsolek CJ, eds. *Rethinking Implicit Memory*. 3–18. Oxford: Oxford University Press.
Roediger, Henry L., and Kathleen B. McDermott. 1995. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning Memory and Cognition* 21 (4): 803.
Rowlands, Mark. 2010. *The new science of the mind: From extended mind to embodied phenomenology*. Cambridge: MIT Press.
Schacter, Daniel L., and Randy L. Buckner. 1998. Priming and the brain. *Neuron* 20 (2): 185–195.
Su, Laura F., Brian A. Kidd, Arnold Han, Jonathan J. Kotzin, and Mark M. Davis. 2013. Virus-specific CD4 + memory-phenotype T cells are abundant in unexposed adults. *Immunity* 38 (2): 373–383.
Tauber, Alfred I. 1997. Historical and philosophical perspectives concerning immune cognition. *Journal of the History of Biology* 30 (3): 419–440.
Tulving, Endel, and Daniel L. Schacter. 1990. Priming and human memory systems. *Science* 247 (4940): 301–306.
Varela, Francisco J. 1994. A cognitive view of the immune system. *World Futures: Journal of General Evolution* 42 (1–2): 31–40.