



# LTP Revisited: Reconsidering the Explanatory Power of Synaptic Efficacy

Jonathan Najenson<sup>1</sup>

Accepted: 30 July 2023

© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

Changes in synaptic strength are described as a unifying hypothesis for memory formation and storage, leading philosophers to consider the ‘synaptic efficacy hypothesis’ as a paradigmatic explanation in neuroscience. Craver’s mosaic view has been influential in understanding synaptic efficacy by presenting long-term potentiation as a multi-level mechanism nested within a multi-level structure. This paper argues that the mosaic view fails to fully capture the explanatory power of the synaptic efficacy hypothesis due to assumptions about multi-level mechanisms. I present an alternative approach that emphasizes the explanatory function of unification, accounting for the widespread consensus in neuroscience regarding synaptic efficacy by highlighting the stability of synaptic causal variables across different multi-level mechanisms.

**Keywords** Memory · Mechanism · Explanation · Unification · LTP · Synaptic Efficacy

## 1 Introduction

Often, we encounter statements like this in textbooks and research publications:

“There is now general consensus that persistent modification of the synaptic strength via LTP and LTD of pre-existing connections represents a primary mechanism for the formation of memory.” (Poo et al., 2016, p.1)

How should we understand these assertions? The neuroscientific consensus on these mechanisms’ explanatory role in memory formation has led to several philosophical

---

✉ Jonathan Najenson  
jonathan.najenson@gmail.com

<sup>1</sup> Department of Humanities and Art, Technion - Israel Institute of Technology, Haifa, Israel

views that consider the synaptic efficacy hypothesis as a paradigmatic case of explanation in neuroscience (Bickle 2003; Colaço, 2020; Craver 2007; Sullivan 2016). The status of synaptic efficacy as a paradigmatic case makes it a target system for opposing philosophical models of explanation in neuroscience, fueling a significant debate over which philosophical model best captures the hypothesis' explanatory power. The way the model is described serves as an example of how a successful neuroscientific explanation should look like.

A central view in the discussion of synaptic efficacy is Carl Craver's *mosaic* view (Craver 2002, 2007). According to Craver, long-term potentiation (LTP), a specific type of synaptic efficacy, is an exemplar of mechanistic explanations in neuroscience. Mechanistic explanations describe a mechanism as interconnected components and activities that produce the phenomenon. These mechanisms are nested in multi-level hierarchies, with each level referring to a different mechanism. LTP is situated in such a hierarchy, being both a component in a higher-level mechanism and a mechanism in itself.

The mosaic view provides a particular account of unity. Results from different levels constrain a multilevel mechanistic explanation, thereby integrating diverse fields, with each adding another piece to a multilevel mechanistic structure. Because LTP is situated in a multi-level mechanistic hierarchy, the synaptic efficacy hypothesis unifies research as it is at the center of these integrative efforts. The ability to capture the explanatory structure of LTP research has established the mosaic view as a leading philosophical account of neuroscientific integration.

It is no coincidence that philosophers have sought to understand how memory research can be unified. Memory is one of the central constructs studied in cognitive science. A first step in deciphering cognition involves understanding how memory works. Decision-making, motor control, and categorization all rely, in their own way, on the ability to encode, store, and retrieve information. While the pervasiveness of memory does not necessarily entail a requirement for explanatory unification, its central role in various cognitive functions renders a unified explanation of memory *desirable*. This unified explanation enables collaboration among diverse disciplines by offering a shared understanding of memory, which in turn paves the way for a more integrated model of the cognitive architecture as a whole. If appealing to how a mechanism produces a phenomenon is a mark of successful scientific explanations (Bechtel and Richardson 2010), then understanding how mechanisms provide unity is key to determining whether the aim of integrating mnemonic phenomena is attainable.

Gaining clarity on the explanatory role of synaptic efficacy is also relevant for addressing broader questions of how neurocognitive mechanisms guide the classification of cognitive phenomena. Classification of phenomena is often based on their underlying mechanisms (Boyd 1999; cf. Craver 2009). However, in cases where there is a shared mechanism at one neuroscientific level but not in another, it becomes ambiguous how to classify phenomena as sharing the same mechanism. For example, pain perception shares a common cellular mechanism involving the activation of nociceptors in response to noxious stimuli. However, the neural circuits involved in processing noxious stimuli differ across species, making it challenging to determine the mechanisms underlying pain perception. Thus, even if neuroscientists are mistaken about the unifying character of synaptic efficacy, gaining a better understanding

of the relationship between synaptic efficacy and mnemonic phenomena can serve as a test case for how mechanisms can succeed or fail in guiding the classification of cognitive phenomena despite having different components, activities, and organizational features.

In this paper, I argue that Craver's mosaic view fails to account for the explanatory power of the synaptic efficacy hypothesis. Specifically, I claim that the mosaic view mischaracterizes the explanatory target of synaptic mechanistic models due to assumptions about how synaptic mechanisms are situated in multi-level mechanistic hierarchies. I offer an alternative view of the unifying character of synaptic efficacy that captures its explanatory force by showing how invariance considerations regarding generality and the significance of horizontal evidential constraints, rather than hierarchical constraints, provide a more relevant picture of unity than the one presented by the mosaic view.

The paper is organized as follows: In Sect. 2, I briefly present the synaptic efficacy hypothesis and describe how the philosophical debate surrounding it revolves around levels and unity. Section 3 examines how Craver's mosaic view applies to synaptic mechanisms, while in Sect. 4, I present an alternative approach that emphasizes the explanatory role of unification. I unpack this approach by considering two aspects of unification: generality and similarity. Finally, in Sect. 5, I evaluate the extent to which the mosaic and unificationist views can be reconciled.

## 2 A Synaptic Explanation

### 2.1 The Synaptic Efficacy Hypothesis

The hypothesis that changes in synaptic strength underlie memory formation and storage is considered, by neuroscientists, to be a unifying hypothesis:

“...there is a general hypothesis for memory storage that is available and broadly accepted. This hypothesis is the following: Memories are stored as alterations in the strength of synaptic connections between neurons in the central nervous system. The significance of this general hypothesis should be emphasized – this is one of the few areas of contemporary cognitive research for which there is a unifying hypothesis.” (Hawkins et al. 2017)

But what specifically does this hypothesis propose, and how does it serve to unify and integrate the diverse findings surrounding memory formation and storage? The hypothesis, which refers to a widely accepted family of models describing memory formation and storage in the animal nervous system, posits that memories are encoded through activation of appropriate synapses and then retained through changes in the strength of synaptic connections between neurons (Morris 2003). This biological process is behaviorally manifested as a change in an organism's response to a relevant event, due to reactivation of modified synaptic connections. According to this hypothesis, the process of memory formation involves the conversion of newly acquired information, initially sensitive to interference, into a more stable

format dependent on changes to synaptic strength. The mechanisms of long-term potentiation (LTP) and long-term depression (LTD) underlie this process, altering the strength of synaptic links based on previous inputs (Hawkins et al. 2017a, b).

In what follows, I will use the term ‘synaptic efficacy hypothesis’ to refer to the widely accepted hypothesis that memories are formed and stored as alterations in the strength of synaptic connections between neurons, with a particular focus on the efficacy of synaptic links, which refers to *the capacity of a presynaptic input to influence postsynaptic output* (i.e., how likely an action potential in one neuron is to induce another neuron to fire). Accordingly, I distinguish modifications to synaptic efficacy from the more general category of synaptic plasticity, which includes all changes to synapses, not solely those that are experience-dependent.

See that although the synaptic efficacy hypothesis is the dominant one, it is not without dissenters. Critics of the hypothesis raise several empirical and conceptual challenges, proposing alternatively that memories are stored in a molecular form instead (Gallistel and King 2009; Gold & Glanzman, 2021). It’s important to note that challenges to synaptic efficacy mainly target its ability to explain engrams, the stage in which memories are stored, and rarely extend to its role in memory formation. Furthermore, it remains an open question whether synaptic and molecular models of memory are genuine rivals, and there is growing interest in the potential for integrating these models (Gershman, 2023; Colaço & Najenson, [in preparation](#)).

Nevertheless, the hypothesis’ status stems from the extensive experimental and theoretical evidence supporting it. Although critics challenge its status as a comprehensive explanation of memory, they generally do not deny the substantial evidence supporting it. A primary objective is to clarify that the unifying character of the synaptic efficacy hypothesis does not hinge on it being viewed as a global hypothesis guiding all memory sciences.

## 2.2 Unity in Neuroscience

How should we understand unification in the context of the synaptic efficacy hypothesis? The philosophical discussion about synaptic efficacy revolves around two key elements. The first element has to do with the role of *levels* of investigation in neuroscientific explanation. Neuroscience encompasses various disciplines and methodologies that cluster around phenomena that often share similar temporal and spatial characteristics. On what basis to distinguish between levels has no easy answers, but there is an underlying assumption that models describe phenomena at a given level (Levy 2016). The debate surrounding synaptic efficacy and levels of investigation revolves around the question of which target phenomenon synaptic models aim to explain. We can better understand Craver’s mosaic view by contrasting it with Bickle’s ruthless reductive view, as the idea that reduction guided the explanation of LTP is one of Craver’s main critical targets.

According to reductive views, most notably John Bickle’s, the synaptic efficacy model amounts to a direct psycho-physical link between psychological categories and molecular entities (Bickle 2003). On Bickle’s view, intermediate levels of investigation between synaptic efficacy and psychological phenomena are redundant and play a merely heuristic role. By tracking how changes in behavior are the result of

interventions at a molecular level, it is possible, Bickle argues, to discover direct links that effectively reduce a psychological-behavioral variable to a molecular-physical variable.

On multi-level views, such as Craver's, it is untenable that a description of a fundamental level exhausts the explanation of a phenomenon. The explanation of mnemonic phenomena involves the integration of different fields working at specific levels of investigation. According to Craver (2002), a multi-level mechanistic account explains mnemonic phenomena through the integration of several different scientific fields, where each field works at a different level of investigation, e.g., psychology at the behavioral level, neurobiology at the physiological level, etc. The explanandum of synaptic efficacy is the level of the mechanism of which this activity is a component. Such an account leaves little room for direct reduction, as memory cannot be explained at any one given level.

The second element concerns how the synaptic efficacy hypothesis figures in a *unifying* framework. There is disagreement among philosophers of neuroscience about how the synaptic efficacy hypothesis achieves the goal of explanatory unification. Specifically, the debate concerns the implications of the synaptic efficacy hypothesis for explanatory relations between the different disciplines included in neuroscience.

On Bickle's ruthless reductive account, the relation between synaptic efficacy and mnemonic phenomena fosters unification by directly reducing psychological phenomena to a systematic explanation in the molecular domain that accounts for all the data that higher-level mechanisms explain. This echoes the highly influential view that the unity of science involves reduction to more fundamental levels (Theurer and Bickle 2013).

According to Craver's 'mosaic' view, synaptic efficacy is considered a component in a multi-level mechanism. Unification is achieved by the integration of different fields in neuroscience joining efforts in elaborating on a single multi-level mechanism. As Levy puts it, "the mechanistic approach suggests a view of unification as the synergistic study of mechanisms by different disciplines" (2016, p.7). Thus, unity in this view results from how different fields converge on the same multi-level mechanism by applying different tools and concepts.

### 3 3. Mosaics and Synapses

I now turn to elaborate on Craver's mosaic view. This view aims to capture how synaptic efficacy is a relevant mechanism in the explanation of memory formation and storage by situating it in multi-level mechanisms. I begin by introducing Craver's mosaic view and explain how it leads to the argument that synaptic efficacy is disunified due to hierarchical evidential constraints.

Mechanistic explanations describe the underlying mechanisms that give rise to an explanandum phenomenon. This type of explanation treats the explanans as a mechanism, a composite of interconnected parts that are organized in a way that produces the phenomenon. A mechanistic explanation accounts for the behavior of the system in terms of roles performed by its underlying parts and the interactions between these parts (Bechtel and Richardson 2010).

Mechanisms are said to have a *hierarchical* structure. On the one hand, a part of a mechanism can be considered in itself a mechanism to which a mechanistic explanation can be provided. On the other hand, a mechanism can itself be a component in a mechanism. A component delineates a level in a mechanistic hierarchy for the causal role it plays. As Craver puts it - “The interlevel relationship is as follows: X’s  $\phi$ -ing is at a lower mechanistic level than S’s  $\psi$ -ing if and only if X’s  $\phi$ -ing is a component in the mechanism for S’s  $\psi$ -ing” (2007, p.189). A mechanistic level involves the entities and the ways in which they are organized to support the function of a higher-level mechanism, but the entities in a lower mechanistic level are themselves decomposed into lower-level mechanisms.

Craver approaches synaptic efficacy through this multi-level perspective, focusing specifically on LTP. On this approach, LTP is a prominent example of a mechanism nested in a multi-level mechanism. Craver motivates the multi-level approach over reductive views by describing how causal knowledge is generated in neuroscience. Craver claims that different scientific fields contribute to the study of LTP by placing evidential constraints on possible entities in higher and lower levels of a single mechanism. Knowledge about mechanisms is generated in multi-level research programs. Like a mosaic, a multi-level mechanism is assembled in a piecemeal way.

How does Craver’s mosaic approach achieve explanatory unification? Unification is gained by different fields with different tools, concepts, and problems converging on a single explanatory mechanism -

“The central idea is that neuroscience is unified not by the reduction of all phenomena to a fundamental level, but rather by using results from different fields to constrain a multilevel mechanistic explanation...The search for mechanisms guides researchers to look for specific kinds of evidence (in the form of different constraints) and provides the scaffold by which these constraints are integrated piecemeal as research progresses” (2007, p.232).

.Research is unified when there is a common investigative cause on which different fields converge to evidentially support different components or levels of the same mechanism. Scientific fields such as electrophysiology and molecular biology are unified by having the same epistemic goal of elaborating different levels of the same multi-level mechanism.

The model of spatial memory is offered as an example of such multilevel mechanistic description. According to Craver, the mechanistic model of spatial memory includes four distinct mechanistic levels: the highest level is the behavioral-organismic level, which deals with the conditions under which different memories may be stored or retrieved; the anatomical-hippocampal level, which concerns the role of the hippocampus in spatial memory and involves descriptions of its anatomical and computational properties; the electrical-synaptic level, which includes entities such as synapses and processes they undergo, such as LTP; and the molecular level, where entities like NMDA receptors are responsible for inducing potentiation processes.

It is important to note that Craver’s account does not imply that the nature and number of levels are fixed; rather, they evolve with scientific investigation. Fields like computational and systems neuroscience may revise this traditional picture and

can alter our understanding of the relationships between levels, such as between the hippocampal and behavioral levels (Levy and Bechtel 2013).

Despite the evolving nature of levels, examining the spatial memory mechanism example can help us understand how the mosaic view accounts for the explanatory target and structure of synaptic mechanisms. Each level specifies a mechanism with entities and activities that are particular to that level. The explanatory target of a mechanism at a given level in the hierarchy is the mechanism at the level 'above'. Consequently, the explanatory target of the LTP mechanism is the formation of spatial maps in the hippocampus.

The determination of the explanatory target has consequences for the explanatory structure of the LTP model, as the mechanism's entities and activities constrain the type of entities and activities possible in the underlying mechanisms of lower levels. For example, for potentiation to occur at the synaptic level, NMDA receptors are required to allow calcium insertion into the neuron, which triggers protein synthesis to stabilize the synaptic connections. Thus, changes in synaptic transmission constrain what researchers expect to find at the molecular level.

Craver argues that LTP research does not provide evidence solely for a downward reductive trend, as it involves findings from different fields that contribute to understanding either the mechanism of LTP or the memory mechanisms that contain it. Although reductive downward-looking trends exist in LTP research, there are also upward-looking trends that are not accounted for in a reductive approach. Investigating memory requires more than just uncovering its molecular basis; it also involves situating it in higher levels, such as relevant brain structures and various behavioral tasks.

It is important to note that mechanisms are individuated by the phenomena they explain. The different mechanistic levels involved in fear memory or spatial memory imply different evidential constraints, and therefore, the multilevel mechanisms underlying these specific forms of memory are different. Craver argues that this allows for relevant differences in the causal details between instances of LTP mechanisms to be captured. For instance, LTP in the hippocampus may differ causally from LTP in the striatum since they are embedded in distinct multilevel structures. If there are different instances of LTP, or synaptic efficacy more generally, then describing the same mechanism does not explain how memories are formed and stored.

To understand Craver's objection, consider how a mechanism is determined to fit a phenomenon (Craver and Darden 2013). A key part of this process concerns the initiation and stopping conditions of the phenomenon, which can be manipulated through interventions on a mechanism's components. These conditions are known as *boundary conditions* and are crucial in determining the scope of a mechanistic explanation. Lack of knowledge about boundary conditions may indicate that a mechanistic explanation is incomplete (Craver and Kaplan 2020). In the case of synaptic efficacy, as more causally relevant details accumulated, the hypothesis, Craver claims, was found to mistakenly lump distinct mechanisms.

Craver argues that boundary conditions are a main reason why researchers should not expect synaptic mechanisms to be part of the same phenomenon -

“Different forms of memory often have different temporal features; that is, the memories of one kind take longer to acquire than those of another kind, and the two decay at different rates....If synaptic changes underlie these different forms of memory, then one would expect differences in the temporal properties of the varieties of synaptic plasticity observed in these brain regions to correspond to differences in the temporal properties of the different kinds of memory” (2007, p.263-4).

Craver denies that synaptic efficacy is actually a unitary memory mechanism because distinct forms of memory have different temporal properties. Roughly speaking, while memory of particular events is rapidly acquired, other forms of memory, such as habits, take much longer to form and decay. Questions regarding the duration it takes for memories to initiate and decay address the same phenomenon that synaptic efficacy aims to explain - memory formation and storage. Mistakenly grouping together rapid and slow memory mechanisms commits a lumping error. The temporal profile of various memory types is so distinct that their underlying mechanisms cannot be the same; thus, grouping them under the same description is a mistake.

Although Craver was one of the first philosophers to express concerns that LTP is not a unitary mechanism, others make similar claims. Sullivan (2016) argues that pragmatic factors are responsible for treating different instances of LTP as the same phenomenon. Even though researchers take different instances of LTP to be produced across various experimental protocols, they abstract from specific mechanistic details to provide a unified model of LTP. Colaço (2020) argues that discrepant mechanistic details led researchers to recharacterize LTP as involving diverse phenomena. Colaço, like Craver, emphasizes that it is the precipitation conditions of LTP that result in the recharacterization of LTP. Therefore, the boundary conditions of LTP pose a significant challenge for a unitary mechanism, regardless of whether one adopts Craver’s mosaic view.

## 4 Causal Unification and Synaptic Efficacy

In this section, I present an alternative approach to understanding the unifying character of synaptic efficacy, focusing on its explanatory role. I elucidate unification through the lens of invariant causal relationships, exploring how such relationships contribute to the hypothesis’ explanatory power. Building on this approach, I investigate the explanatory status of the synaptic efficacy hypothesis by examining its generality and the similarity of synaptic mechanisms.

According to Kitcher (1989), who offered the standard account of unification, explanations are argument patterns that allow for deriving descriptions of many phenomena.<sup>1</sup> When fewer argument patterns are used, the description becomes more

<sup>1</sup> Argument patterns consist of schematic sentences (where nonlogical terms are replaced by variables), filling instructions (which are rules for substituting these variables with specific values), and classifications (which define which sentences serve as premises and which ones act as conclusions). For example, consider the sentence ‘Organisms homozygous for the sickling allele develop sickle cell anemia.’ This can be restructured as ‘Organisms homozygous for A develop P,’ where ‘A’ and ‘P’ are variables. The



unified, thus providing an explanation of the phenomena of interest. Kitcher suggests that there is an explanatory store which is “the set of arguments that achieves the best tradeoff between minimizing the number of premises used and maximizing the number of conclusions obtained” (p. 431). Kitcher’s basic idea, then, is that explanation consists in minimizing the ratio of explanantia to explananda.

Kitcher’s view of unification is controversial, largely due to its limitations in accounting for various forms of explanations. For example, Kitcher’s account has faced criticism for its difficulty in assessing individual causal explanations. Philosophers attacking the relevance of unification to explanation often argue that unificationist views, like Kitcher’s, imply that explanations of single or low-probable events cannot be explanatory (Craver 2007). The objections are due to Kitcher’s identification of explanation with unification, casting doubt on the value of unification to explanation.

However, Kitcher’s idea can be reconceived as a *norm* rather than an *account* of explanation. Indeed, several philosophers have pointed out that unification can inform and enrich causal explanations. According to Levy (2016), it is plausible to ask how unified our stock of causal explanations is, a question that can be answered by applying Kitcher’s measure of unification. When a small set of causal-explanatory models can account for a broad range of phenomena within a given domain, we achieve unification insofar as these same models can be used to explain a wide range of phenomena.

Weber and Lefevere (2017) add another perspective. They argue that unification supplements causal conceptions of explanation by answering *resemblance questions*. Resemblance questions focus on similarities between individual causal explanations. For example, we can ask why different species learn associations through Pavlovian conditioning. So, unification can be seen as a way to evaluate what is shared by causal-explanatory models which enable them to explain their target phenomena.

Understood as a norm of causal explanation, unification does not imply that descriptions of single individual events cannot be explanatory. Adopting unification as a norm allows for explanations of individual events to carry explanatory power, especially if other norms are also considered. Unlike Kitcher’s view, seeing unification as a norm enables its application to different targets, such as local explanations that are pursued individually, rather than being the primary goal of scientific explanation. The emphasis here is not on seeking a global synthesis of various fields, but on achieving a specific epistemic goal through the integration of unification with causal explanations (cf. Brigandt 2010).

That said, unification does mean that, in some cases, we should *prefer* the more unifying explanations. Explanatory unification is tied to explanatory power, namely the idea that some explanations are better than others. Specifically, some explanatory models (or families thereof) score higher in terms of their explanatory virtues. One way in which a model exhibits explanatory virtue involves the extent to which it minimizes the ratio of explanantia to explananda.

---

filling instructions, in this case, would direct us to replace ‘A’ with a specific allele’s name and ‘P’ with a phenotypic trait.

But what does unification, as a norm of causal explanations, consist of? To evaluate unifying causal-explanatory models we can appeal to notions of *invariance, stability or insensitivity of causal relationships* (Woodward 2010, 2021; Ylikoski and Kuorikoski 2010).<sup>2</sup> Mechanists interpret causal relationships within the interventionist framework, according to which X causes Y iff an idealized experimental manipulation of variable X leads to a change in the value of variable Y under certain background circumstances. Invariance concerns whether changes to the value of Y, resulting from interventions on X, hold under a wide range of background circumstances.<sup>3</sup> These circumstances, implicitly represented in the X-Y relationship, may or may not hold causal relevance to Y. Thus, the crux of causal unification lies in the examination of invariant causal relationships.

With this understanding of unification as a norm of causal explanations, we can now explore the significance of invariance in the context of causal unification. Invariance is crucial for causal unification since it captures how mechanistic models employ similar causal variables in explaining a wide range of phenomena. That is, a relatively invariant causal relationship between a mechanism and its explanandum phenomenon, under a wide range of background circumstances, implies that the mechanism can produce the explanandum phenomenon across various behavioral tasks, species, and causal contexts (e.g., different regions of the nervous system), among others.

On this view, mechanistic models are deemed more explanatory when the causal relationships they describe exhibit relatively more invariance under a given range of background circumstances. These models provide a better explanation over a less-invariant counterpart, as they can answer more questions about the explananda. These inquiries may encompass not only the familiar what-if-things-had-been-different questions, namely circumstances under which the explanandum phenomenon would have been different, but also resemblance questions that arise when different explananda exhibit shared characteristics.

By identifying an invariant causal relationship, the ratio of explanantia to explananda can be minimized.<sup>4</sup> As said, this occurs when the same causal relationships remain stable under a wide range of background circumstances, such as producing consistent behavioral outcomes in different species. Such minimization enables a smaller set of mechanistic models to explain a variety of phenomena. When I say that synaptic efficacy constitutes a unifying hypothesis, I am proposing that the mechanistic model(s) involved in this hypothesis display invariance under a wide range of background circumstances.

<sup>2</sup> I will use the terms interchangeably.

<sup>3</sup> Woodward (2021, p. 264, ft. 41) acknowledges the tight connection between invariance and unification but dismisses it due to the standard characterization of unification being considered non-causal. In contrast, the perspective on unification presented here, which treats it as a norm of causal explanation, effectively addresses Woodward's concern.

<sup>4</sup> Invariance offers a means of interpreting what minimizing the ratio of explanantia to explananda implies within the context of causal explanations, specifically those within the interventionist framework. Notably, unification need not be conceived in the same terms for non-causal explanations or under a non-interventionist understanding of causal relations.

In the upcoming subsections, I examine the explanatory status of the synaptic efficacy hypothesis by exploring the role of invariance in determining generality and similarity, thereby contributing to its unifying character. First, I discuss how generality considerations indicate that synaptic mechanisms exhibit more explanatory power compared to other memory mechanisms. Next, I show how focusing on shared features of diverse synaptic mechanisms contributes to inferences about mechanistic types. While the main purpose of discussing these issues separately is to compare and contrast my position with Craver's, they also highlight a core issue at stake: how should we identify the relevant features that differentiate between explanatory representations? I elaborate further on this issue in the last section.

#### 4.1 Generality

In this subsection, I aim to explore the generality of synaptic efficacy in terms of its scope and breadth, shedding light on how generality plays a role in establishing the hypothesis' unifying character. Neuroscientists describe the synaptic efficacy hypothesis as a unifying hypothesis since it explains a wide range of mnemonic phenomena. The range of phenomena in which synaptic efficacy plays an explanatory role is a central way to evaluate synaptic efficacy's unifying character. Expressing this idea, Eric Kandel and colleagues say that -

“One of the key *unifying* findings emerging from the molecular study of implicit and explicit memory processes is the unexpected realization that these distinct forms of memory that differ not only in the neural systems involved but also in the nature of the information stored and in the role of attention in that storage, nevertheless share a *common set* of molecular mechanisms... Thus, whereas animals and humans are capable of a *wide variety* of learning processes... they may recruit the *same restricted set* of molecular logic for the storage of long-term memory” (Hawkins et al. 2017, p. 22, my emphasis).

How are we to understand the relationship between the wide variety of learning processes and the common set of mechanisms that underlie them? The relationship discussed by Kandel and colleagues immediately raises questions about the generality of mechanistic models. When considering the generality of mechanistic models, it is important to distinguish between scope and breadth. Breadth pertains to an *explanatory generalization* that holds under a wider range of background circumstances (Blanchard et al. 2018). A mechanistic model that exhibits breadth can be understood through the causal relations that remain invariant under a larger set of changes in the explanandum phenomenon. Another way to evaluate the generality of a mechanistic model involves its scope, which pertains to the set of mechanisms falling under the model's range of application. A mechanistic model with broader scope refers to a model that applies to a wider range of phenomena. Thus, to evaluate the generality of mechanistic models, one should differentiate between models describing stable explanatory generalizations and those describing a relatively wider range of phenomena.

When considering generality in terms of breadth, synaptic models of memory surpass other mechanistic models in terms of explanatory power. Researchers evaluate the explanatory power of synaptic mechanistic models by examining causal variables under varying sets of background circumstances, including different species, neural circuits, and behavioral tasks. To appreciate the significance of Kandel and colleagues' assertion, it is useful to consider the pervasive observation that intervening with synaptic mechanisms results in memory storage and retrieval deficits across an extensive range of behavioral tasks. The fact that these interventions occur in a wide range of background circumstances enables the synaptic efficacy hypothesis to address more what-if-things-had-been-different questions about the explanandum phenomenon, a main way for evaluating explanatory power.

A notable example is synaptic consolidation, which is the process by which stabilization of a memory involves gene expression and protein synthesis necessary to strengthen synaptic connections. The causal relation between synaptic consolidation and memory performance has been observed across various tasks, including the Morris water maze, auditory fear conditioning, and habituation. Suppose we want to understand why the causal relationship underlying synaptic consolidation is invariant. If, across various tasks, each involving a different set of background circumstances, inhibiting protein synthesis interferes with learned behavior, we can infer the presence of a stable causal relationship. In this sense, one can understand the generality of consolidation, a central activity of synaptic efficacy, as reflected in the invariance of its causal relationship, where the manifestation of learned behavior is causally dependent on protein synthesis. Memory formation may be operationalized through different tasks but interventions on synaptic mechanisms result in a similar outcome.<sup>5</sup>

It should be noted that not all causal variables in synaptic models exhibit the same degree of invariance. It is entirely possible (even to be expected) that a causal relationship is invariant under some changes and not under others. Stable causal relations must be understood as invariant only under a *specified* set of changes. The components and activities of synaptic efficacy, as causal variables in synaptic models, may exhibit varying levels of stability under different ranges of background circumstances.

However, not all background circumstances are equal, and their relative importance plays a crucial role in assessing the invariance of causal variables. Domain-specific knowledge helps us identify which changes in background circumstances are most significant for evaluating invariance. It informs us about which conditions are typically 'usual' or 'normal' within a given domain, thereby establishing the relevant range of background circumstances. (Woodward 2010). For example, although inducing synaptic changes is sensitive to temperature changes, temperature may not be a primary factor when evaluating the overall explanatory power of synaptic efficacy. Memory-dependent behavior resulting from changes in synaptic strength will not be systematically altered, provided a certain temperature range is maintained. In contrast, the range of behavioral tasks involved holds significance as background cir-

<sup>5</sup> What the subjects of these experiments are supposed to remember, and subsequently forget is not entirely independent from the mechanisms in which synaptic efficacy is nested. Synaptic efficacy determines whether memories are formed, it does not specify which memories are formed.

cumstances, since they shape the explanatory target that can be demonstrated. Thus, by relying on domain-specific knowledge, we can discern more or less relevant background conditions when analyzing the degree of invariance in synaptic models of memory formation.

Considering breadth is also important when evaluating the explanatory power of mechanistic explanations at *different levels*. Synaptic mechanisms exhibit a greater degree of invariance compared to other mechanisms in multi-level mechanistic explanations of memory. For instance, while synaptic mechanisms affect memory formation across a wide range of tasks, hippocampal mechanisms influence memory formation more selectively. Although hippocampal mechanisms contribute to various memory tasks beyond the traditionally-conceived spatial and episodic tasks (See, e.g., De Brigard 2019; Roy and Park 2010), their causal relationships are not as stable as those of synaptic mechanisms. To illustrate, in both auditory fear conditioning and contextual fear conditioning, inhibiting synaptic consolidation impairs performance, while hippocampal lesions affect contextual fear conditioning (Rudy et al. 2004). Notably, the stability of synaptic mechanisms remains uncompromised even when considering only hippocampal-based tasks. Such findings underscore synaptic efficacy as a *primary* mechanism for memory formation.

The mosaic view downplays the *explanatory asymmetry* between mechanistic levels, specifically the causal role that synaptic efficacy is expected to play across more mechanistic hierarchies compared to other memory mechanisms. This explanatory asymmetry posits that various mechanisms, although causally relevant to the explanandum, may not be related to it in a symmetric way. The mosaic view, it seems, holds to a causal parity claim, urging “all causally relevant factors be treated the same, in a spirit of “causal democracy” (Woodward 2010, p.316).

According to Woodward (2010), defenders of causal parity theses argue that in many cases, there is no principled basis for distinguishing between the causal roles played by different factors with respect to an explanandum, and that both should be regarded as playing a “symmetric” causal role. An analogy can be drawn from the debate about genetic causation. Developmental systems theorists maintain that since both genetic and extra-genetic mechanisms are relevant to explaining evolution, namely, the causal role that genes play in evolution are also played by other factors. However, some argue that there is a difference in the degree of stability exhibited by these mechanisms (Weber, Forthcoming). Similarly, for synaptic mechanisms, the relation of one mechanistic level to the explanandum phenomenon can be said to be more stable, and thus underscore synaptic efficacy’s primary role in explaining memory formation.

An appeal to generality might seem misplaced given that Craver rejects its role in evaluating mechanistic explanations (Craver and Kaiser 2013; Craver and Kaplan 2020). Specifically, Craver denies that models that apply to more phenomena are improved by expanding their scope. A model that explains memory formation in rodents would not be improved by the fact that the same model works for primates. Not only will expanding the model’s scope fail to add to its explanatory power, but, as Craver and Kaiser note, such a model “has difficulty accommodating the variability characteristic of biological mechanisms” (2013, p.134), thereby overlooking relevant explanatory details.

This is where distinguishing between breadth and scope matters. It is not simply that synaptic models can apply to a wider range of phenomena, but that the components and activities involved in synaptic efficacy are more stable under a wider range of background circumstances. For example, since AMPA receptors are the primary mediators of fast excitatory synaptic transmission and responsible for most of the postsynaptic current at glutamatergic synapses, they are likely to exhibit a stable value in downstream causal variables in most synaptic models of memory formation.

Synaptic efficacy derives its unifying explanatory character from the fact that some of its components and activities remain relatively insensitive to changes. While a model need not lack explanatory power for having a narrower scope, possessing invariant components and activities is not merely “a further interesting and important fact” (Craver & Kaplan 2019, p.307) for a mechanistic model. This fact reflects that the model has honed in on explanatorily relevant information, allowing it to capture explanatory generalizations that enable answering more what-if questions, as what would happen to memory formation if we knock out AMPA receptors in the cerebellum and striatum rather than in the hippocampus. The generality of the synaptic efficacy hypothesis is highlighted by the fact that the variables in synaptic models maintain their character and causal role, even when the range of background circumstances in which the model provides explanatory information about memory formation is expanded.

## 4.2 Similarity

In this subsection, I explore ways to resist the idea that diverse causally relevant details among synaptic mechanisms imply that they are not part of a unitary mechanism, offering, in its place, a view that treats synaptic efficacy as a mechanistic type. First, I will present the worry associated with considering synaptic efficacy as a mechanistic type and highlight some reasons for preferring a unifying approach. Then, I will underscore the significance of abstraction in mechanistic explanations, arguing for the explanatory role of mechanism schemas. Finally, to solidify my argument, I will focus on the case of spike-timing-dependent plasticity as an illustration of this view.

Diverse causally relevant details are at the basis of claims made by neuroscientists that different instances of synaptic efficacy “may share some, but certainly not all, of the properties and mechanisms of NMDAR-dependent LTP” (Malenka and Bear 2004, p.5). Such claims have been interpreted by Craver, Colaço and Sullivan as showing that synaptic efficacy involves different mechanisms. The variability characteristic of synaptic mechanisms gives the impression that a unitary mechanism is untenable.<sup>6</sup> When such different mechanisms are sometimes described as belonging to the same phenomenon, philosophers conceive this categorization as serving

<sup>6</sup> This impression might be conceived as overriding considerations of generality. The following discussion questions whether the diversity of synaptic mechanisms indeed hinders the feasibility of a generalized mechanism. However, my focus in this section is on identifying similarities through abstraction. It is crucial not to conflate abstraction with scope. A single mechanism can have a narrow scope yet remain highly abstract, or the other way around.

a pragmatic purpose, such as minimizing explanatory inconsistencies to facilitate communication.

Consider synaptic efficacy in the striatum, a key component in the larger network involved in procedural memory, which exhibits different properties compared to the meticulously studied hippocampal LTP (Kreitzer and Malenka 2008). Which mechanistic details differ in striatal synaptic mechanisms? One highly discussed discrepancy is in the different induction protocols applied. To induce potentiation in hippocampal synapses researchers apply high-frequency stimulation. To induce potentiation in striatal synapses, researchers apply experimental protocols that aim to capture the relative spike timing (Perrin and Venance 2019). Another potential discrepancy is that striatal synapses have different receptors than glutamate receptors. Striatal projection neurons have dopaminergic receptor expressions: D1-class and D2-class.

What motivates splitters' claims is an emphasis on differences over similarities. I want to suggest an alternative way to look at changes in synaptic strength. On this view, the mechanisms underlying synaptic efficacy form a mechanistic type. The shared properties of this mechanistic type inform inferences made from one mechanistic instance to other instances of synaptic efficacy. Moreover, the fact that researchers learn and infer from one instance to another, similar, mechanism reveals that the grouping of these *prima facie* different mechanisms goes beyond pragmatic considerations.

Unifying explanations, I argued, address resemblance questions. These explanations seek to answer why similarities are found among two or more regularities, such as why both mice and slugs depend on temporal contiguity to form memories. To tackle these questions, neuroscientists often resort to more abstract representations of the explanans. This approach is required because these questions revolve around shared features of mechanisms. By emphasizing stable regularities and omitting certain features, a model can effectively capture these shared aspects. In turn, these features impose evidential constraints on inferences about *instances within that mechanistic level*.

Nonetheless, the relationship between mechanistic models and abstraction is not straightforward. Descriptions of synaptic efficacy are understood to be the result of statistical summations describing potentiation effects in different organisms dispersed over space and time. Since the explanandum is a regularity rather than a particular occurrence, it is unsurprising that abstraction is involved. Consequently, what is at stake between the mosaic and unificationist views lies in whether abstraction plays an *explanatory* role. The central question is whether there is a shared explanatory-causal structure underlying synaptic efficacy and, if so, whether its abstract character adds to our explanation of the phenomenon. With this in mind, I now turn to examining the significance of abstraction in mechanistic explanations, specifically focusing on the explanatory role of mechanism schemas.

To make sense of similar synaptic mechanisms, it is important to consider the relationship between token and type level mechanistic explanations. Craver argues that type level models should be understood extensionally, as descriptions of the set of mechanism tokens explaining a phenomenon (Craver and Kaiser 2013). For example, the token mechanisms of different strains of a particular species, such as Long-Evans

and Sprague Dawley rats, will be jointly described in a type-level model of spatial memory. While researchers choose a type-level description based on the level of abstraction of the phenomenon they aim to explain, ultimately, in Craver's view, it is the mechanism tokens, rather than an abstract description of the mechanism type, that carry the explanatory weight.

However, this approach struggles to account for what makes type-level mechanistic models explanatory. After all, grouping tokens into types is based on their shared mechanistic features, abstracted from individual token mechanisms. In order to accommodate token mechanisms in type-level descriptions, one can employ a mechanism schema. This is an abstract description of a type of mechanism that can be supplied with descriptions of specific components when applied to a token (Craver and Darden 2013). Yet, this does not solve our problem, as mechanism schemas are generally not interpreted as playing an explanatory role, but rather as guiding the discovery of mechanisms by serving as a template on the way to more detailed explanations of a phenomenon.

An alternative approach to understanding type-level mechanistic explanations, and the role mechanism schemas can play in such explanations, involves examining how abstraction is used in generating type-mechanism models. Darden (2006) offers a view that highlights the explanatory potential of mechanism schemas (See, e.g., Stinson 2016 for a similar view). In Darden's perspective, the abstractness of schemas not only plays a role in discovering but also in unifying mechanistic research - "If one knows what kind of activity is needed to do something, then one seeks kinds of entities that can do it, and vice versa. Scientists in the field often recognize whether there are *known types of entities and activities* that can possibly accomplish the hypothesized changes" (2006, p.28, my emphasis). This suggests that mechanism schemas serve a crucial function in not only guiding discovery but also contributing to explanatory practices in mechanistic research.

The converse claim is important here - If one *seeks* the kind of entities that can perform a certain activity, then one *knows* the kinds of entities that can do it. Mechanism schemas can be understood as descriptions of mechanism tokens grouped together in a manner that enables explanatory generalizations over those groupings to be, and likely to be, evidentially supported.<sup>7</sup> A mechanism schema represents a generic causal structure that guides the construction of models explaining potentiation effects in synapses. These models can then be used to discover novel mechanisms because mechanism schemas describe a causal structure shared among the mechanism tokens. Consequently, mechanism schemas can feature in type-level explanatory models or, crucially, they can provide the justification for treating grouped mechanism tokens as instances of a generic causal structure at the type-level.

This generic structure incorporates type-variables, typically instantiated by *representative* components and activities. One way to determine type membership is through invariance. A token component or activity would be deemed representative

<sup>7</sup> In this regard, my approach is similar to approaches that emphasize the role of kinding in hypothesis formation (Colaço 2022). However, one difference is that my approach focuses more on how existing evidence plays a role in mechanism characterization rather than concentrating on how kinding guides and justifies investigative and classificatory practices. In this sense, compared to other approaches, my approach is more backward-looking than forward-looking.



of a type if it remains stable across changes in background circumstances, including changes in the composition of particular token mechanisms. Several philosophers (See, e.g., Bechtel 2009; Bickle 2003) have noted that generalizations of mechanisms responsible for similar phenomena are often justified by appealing to how components, activities, or the organization of a mechanism in an ancestral species that play the same functions are largely conserved in descendant species. By using invariance as a guideline, we can identify type-level components, activities, or organization of mechanisms despite the variability of token mechanisms, without requiring representative type-variables to be definitional.

Alongside recognizing the known types of entities required for mechanism schemas, we also need to understand why extrapolating from known tokens is intelligible. This has been an important aspect of grasping how more generalized explanatory models account for closely related phenomena that exhibit different causal details (Batterman and Rice 2014; Weisberg 2013). As such, these models can answer resemblance questions about their targets. What makes it possible for these models to do so is that they describe features that are required for the explanandum phenomena to occur, and the model explains how those features are instantiated despite expected variability between mechanistic tokens.

By examining how mechanism schemas account for different causal details and understanding the intelligibility of extrapolating from known tokens, we can develop a better understanding of how they function as explanatory models. Specifically, a mechanism schema plays an explanatory role if it satisfies two conditions:

- a. (a) Type-variables remain relatively invariant.
- b. (b) The model explains why type-variables are required for the explanandum phenomena to occur despite token-variability.

The extent to which type-level synaptic mechanistic models satisfy these conditions will tell us whether the mechanism schema of synaptic efficacy provides explanatory purchase. In what follows, I explore the case of spike-timing-dependent plasticity. To illustrate the unified character of the schema, I describe how distinct instances of synaptic efficacy can be demonstrated experimentally through interventions on this form of plasticity.

Faced by discrepancies of synaptic efficacy observed in the striatum researchers ask if these mechanisms can be characterized as *Hebbian*. Hebb (1949) formulated a basic organizing principle underlying the connection between memory formation and synaptic efficacy:

“When an axon of cell A excites cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells so that A’s efficiency as one of the cells firing B is increased” (1949, p.4)

According to this organizing principle, *co-occurrence* is fundamental for changes in synaptic strength. What co-occurrence means, experimentally and conceptually, has been extremely important in understanding different instances of synaptic efficacy. Hebbian plasticity relies on the ‘coincident’ activity on pre-and-postsynaptic sides.

When researchers explored the conceptual space of Hebb's rule, they revealed other ways in which synapses could coincide. This expansion involved taking co-occurrence to involve a precise temporal asymmetry between pre-and-postsynaptic spikes, a phenomenon called *spike-timing-dependent plasticity* (hereafter STDP).

In STDP, the timing of the presynaptic spike relative to the postsynaptic potential determines the occurrence of LTP or LTD. If the presynaptic spike precedes the postsynaptic spike, LTP ensues, while LTD manifests if the reverse happens. STDP paradigms have been used to demonstrate bidirectional plasticity in striatal synapses, where the type of synaptic mechanism depends on the order of pairings (Perrin and Venance 2019). By manipulating STDP pairings, researchers have observed various instances of synaptic efficacy, ranging from the well-known NMDA receptor-dependent LTP, initially discovered in the hippocampus, to other forms such as LTD and endocannabinoid-LTP (Caporale and Dan 2008; Feldman 2012).<sup>8</sup>

I claimed that mechanism schemas are explanatorily relevant when the type-variables they describe remain invariant. Using STDP induction protocols to demonstrate LTP in the striatum indicates that the same activity, i.e., co-occurrence of pre-and postsynaptic spikes, leads to the same outcome under different background circumstances. If the fact that diverse instances of synaptic mechanisms are precipitated differently actually indicates that synaptic efficacy involves distinct phenomena mistakenly lumped together, then applying Hebbian experimental paradigms to study striatal LTP would not make much sense. That STDP induction protocols are successful in demonstrating striatal LTP tells us that this activity is not limited to the token level and is stable across different circumstances.

However, not only does this type-variable remain relatively invariant, but the mechanism schema also explains why this type-variable is required for the explanandum to occur despite variability in tokens. The ability to induce LTP in the striatum was a result of considering how precise timing of pre-and postsynaptic spikes matters in determining the occurrence of changes in synaptic efficacy. The idea of co-occurrence at the basis of Hebb's rule, leads to the inference that the changes in the same type-variable, i.e., the order of pre-and postsynaptic spikes pairings, explains why some tokens exhibit LTP while others LTD. The different properties involved in hippocampal and striatal synaptic mechanisms, properties that appear to distinguish between types of synaptic efficacy, actually reinforce the belief that memories are formed through the joint activity of pre-and-postsynaptic neurons.

I conclude by addressing several objections. Recall that what is at stake is whether similarities are explanatory. One way to resist the STDP story is to argue that similarities are useful for discovery but not relevant for explanation. This aligns with the common idea that the function of schemas is discovery. However, this objection overlooks the fact that the mechanism schema of synaptic efficacy provides explanatory information; it answers why-questions by citing type-variables whose changes

<sup>8</sup> A potential worry is that the mechanistic model as described here is possible in principle but may not be an adequate depiction of scientific practice. According to Craver and Kaplan (2020), it is often "unwieldy to create a single 'uber-model'" (p.309). However, attempts to create a general model that conjoins multiple models of synaptic efficacy by appealing to STDP are part of a live research program. On this approach, STDP is a multifactorial rule in which the magnitude of LTP and LTD depends jointly on spike timing, presynaptic firing rate, postsynaptic voltage, and synaptic cooperativity (Feldman 2020).

in value explain the variability of token mechanisms. So, in contrast to a model that omits details to capture expected variability among token mechanisms, we have a model that accounts for how changes in its values (e.g., order of spike pairings) explain known variability, thus moving beyond discovery towards explanation.

While one may acknowledge the explanatory function of schemas, it is still possible to argue that the grouping of synaptic efficacy is driven by pragmatic considerations. Is it just a useful way to avoid an “unwieldy taxonomy containing as many different kinds of LTP as there were experiments for producing it” (Sullivan 2016, p.137)?<sup>9</sup> The STDP case exemplifies Darden’s remark. Researchers pursue Hebbian properties because they understand the *kind of activities needed for potentiation*. Full-blown pragmatists would need to explain why inferences that different tokens of synaptic efficacy share similar properties successfully inform inferences about closely related mechanisms if the grouping is purely for communicative purposes. The success of these inferences suggests that the grouping of synaptic mechanisms goes beyond practical considerations, uncovering systematic patterns within the nervous system.<sup>10</sup>

Pragmatism may present itself in a different manner. Craver contends that “[t]he appropriate grain of the explanans depends on how the explanandum is specified” (Craver and Kaplan 2020, p.308). However, he denies that abstraction plays an explanatory role, viewing it as mainly serving practical functions (e.g., psychological, communicative, etc.). Since “...which causal structures count as relevant parts of those mechanisms depends on...a conceptual decision made at the beginning rather than discovered within the causal order of things” (Francken et al. 2022, p.12), selecting a level of abstraction to individuate type mechanisms could be regarded as necessarily resulting from pragmatic considerations.

There are two ways to respond to this objection. First, choosing (for the same explanandum) a more abstract grouping allows for explaining a mechanism in a way that is, at least partially, independent of the level of detail required to specify the explanandum. For instance, the success in applying STDP paradigms across neural circuits suggests that synaptic learning rules are relevant across diverse local multi-level mechanisms (Magee and Grienberger 2020). This is an explanatory achievement, given that researchers reveal causal information that allows them to use the same explanation to answer diverse contrastive questions.

Additionally, abstraction guided by invariance considerations helps determine the most suitable grouping to strike a balance between relevant and irrelevant details.

<sup>9</sup> Sullivan’s position is too complex to be fully captured here. Sullivan (2009, 2016) argues that the diverse LTP-inducing stimulation protocols lead to differences in the cellular and molecular mechanisms responsible for LTP, making it unclear whether different labs are examining the same phenomenon or distinct ones. This diversity, according to Sullivan, indicates that the sciences of memory are not geared toward coordination. Consequently, it is unlikely that a single type of LTP will be discovered, given the nature of their investigative projects. Nonetheless, Sullivan acknowledges that unification is possible, at least in principle.

<sup>10</sup> Whether these deeper patterns say something about the causal structure of the world rather than the epistemic characterization of these mechanisms is an open question. Craver does not deny that there are abstract patterns in the causal structure of the world. Nevertheless, this view faces various issues in accommodating the relation between token and type mechanisms (See, e.g., Sheredos 2016). I return to this point in Sect. 5.

Grouping decisions may dictate the appropriate choice of the level of explanation, revealing how the explanans is *proportional* to the explanandum (Woodward 2021). For instance, if the effects associated with LTP and LTD are fully accounted for by the timing of spike order, then STDP would be more proportional to the explanandum of memory formation. In this case, how to represent the explanans would not be entailed by prior conceptualization but the outcomes of explanatory practices.

## 5 Beyond the Mosaic View

In the previous section, I presented the unificationist perspective and outlined several objections to the mosaic view. Now, I will further investigate these limitations and evaluate the compatibility of the mosaic view with a unificationist approach, taking into account both explanatory virtues and representations.

Consider why the mosaic view fails to fully account for the explanatory target of synaptic efficacy. In the mosaic view, synaptic efficacy is used to explain an explanandum phenomenon situated at a higher mechanistic level within a multi-level mechanistic hierarchy. However, the mosaic view does not accommodate the fact that the ‘same restricted set of molecular logic’ is employed to explain various instances of memory formation. The phenomenon of synaptic efficacy is invoked as an explanatory mechanism for memory formation across a broad range of background circumstances.

A guiding assumption of the mosaic view is that mechanistic levels place evidential constraints on other levels, and, consequently, these other levels are expected to exhibit properties that correspond to these constraints.<sup>11</sup> However, this assumption leads to an incomplete understanding of the explanatory target of synaptic efficacy. The mosaic view encounters difficulties when attempting to account for the fact that various high mechanistic levels appeal to synaptic efficacy as a mechanistic part. Overlooking this fact results from assuming that the explanatory purchase of synaptic efficacy is restricted to a specific local multi-level structure.

These issues raise the question of whether the mosaic view is compatible with a unificationist approach. Let us examine compatibility on two fronts: explanatory virtues and explanatory representations. Explanatory virtues refer to the desirable qualities that make an explanation powerful, qualities that can include, among other things, simplicity, accuracy, and unification. We can assess different mechanistic models by evaluating how well such models exhibit explanatory virtues. In examining explanatory representations, I will focus on the way in which mechanistic explanations are structured, specifically, the degree of abstraction involved in describing the explanans and explanandum. These representations can range from highly abstract models

<sup>11</sup> Craver’s mutual manipulability account is partially responsible for this expectation, but mechanists are not necessarily committed to this account, and several alternatives have been proposed (e.g., Serban & Holm, 2019). Alternative views propose that constitutive relevance relations are, at most, abductive since they can be sensitive to interventions that may falsify the relation, like the connection between hippocampal (or striatal) level and synaptic level. Consequently, higher-level constraints do not necessarily require splitting synaptic mechanisms. This is consistent with the view of neuroscientists who consider synaptic efficacy as the primary, but not the sole, *causal* mechanism explaining memory formation.

to more detailed descriptions. Considering both explanatory virtues and explanatory representations will help determine the extent to which the mosaic view aligns with a unificationist approach.

Treating synaptic efficacy as a unitary mechanism is related to whether there is one correct way to group distinct mnemonic phenomena together. Craver contends that the search for a single underlying mechanism for various memory types—such as episodic or procedural memory—revealed a lumping error in characterizing mnemonic phenomena. Craver views the division of memory into distinct types as a significant advancement in the sciences of memory, promoting the exploration of multiple, distinct memory mechanisms (Craver and Darden 2013).

So far, I aimed to show that decisions to group closely-similar phenomena are connected to the adoption of certain explanatory virtues. As such, these decisions should be sensitive to trade-offs in explanatory power. The extent to which there is an inherent incompatibility between the mosaic and unificationist views depends, in part, on how these trade-offs are addressed. This is why the tension between these views is not resolved simply by attributing a more ambitious ideal of unity in neuroscience to the unificationist view (cf. Levy 2016). Instead, they reflect genuine disagreements about which norms should guide explanatory practices.

Model-building often involves balancing competing explanatory virtues, such as accuracy and generality. These trade-offs occur when increasing one desideratum of model-building limits the possibility of increasing another desideratum (Weisberg 2013). A general model may be applicable to a broader range of phenomena, but it might not accurately represent the specific details of any single case. The existence of trade-offs implies that decisions about which explanatory virtue to sacrifice may involve compromising aspects of the model's explanatory power.

One way to capture the difference between mosaic and unificationist views concerns the distinction between ontic and epistemic views of explanation (Salmon 1984). Each view holds that explanations are guided by distinct norms that push in different directions. The ontic view holds that explanations try to capture the causal structure of the world. Ontic views demand explanations to be deeper or more complete. According to Strevens (2008), deeper explanations provide better understanding of the phenomena and are achieved when we have a mechanistic model of all relevant causal influences on the explanandum phenomenon. Craver and Kaplan (2020) raise similar considerations when invoking completeness. Explanatory grasp is evaluated according to how much causally relevant detail is known about the factors constitutively relevant in providing a mechanistic explanation.

The epistemic view takes explanation to be in the business of making phenomena intelligible. In other words, explanation is first and foremost a cognitive achievement (Sheredos 2016). On this view, unification is one of the norms guiding explanations since it requires synthesizing and systematizing the many patterns needed to account for the explanandum. As Friedman puts it, “this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given” (Friedman 1974, p.15).

Keeping in mind the trade-offs in model-building and the differing norms guiding epistemic and ontic views of explanation, the incompatibility of the mosaic and unifi-

cationist views boils down to two possibilities: strong and weak incompatibility. On a strong reading, the mosaic view is, *in principle*, unable to account for the explanatory power of synaptic efficacy. High-level evidential constraints indicate that grouping synaptic efficacy constitutes a lumping error, as they reveal a certain fact about the causal structure of the world - the world is devoid of a unitary synaptic mechanism. Due to these central assumptions, the mosaic view construes the explanatory role of synaptic efficacy as being confined to a specific, local multi-level structure. Consequently, it struggles to provide a satisfying account of the explanatory power of synaptic efficacy.

On a weak reading, mosaic and unificationist views are incompatible *in practice*. When confronted with multiple mechanistic models at a certain level, researchers might prefer models that minimize (in terms of invariance) the ratio of explanantia to explananda. This preference may arise even when a more complete model of the phenomenon exists, reflecting explanatory practices that aim to prioritize unification over completeness (or breadth over depth). Reconciling mosaic and unificationist views requires recognizing that “epistemic norms sometimes have priority (e.g., in cases of general and systematizing explanation) and that ontic norms may sometimes have priority (e.g., in cases of singular explanation)” (Sheredos 2016, p.947). The need to satisfy different norms in explanatory practices necessitates the use of different models for the same phenomenon.

On this reading, the incompatibility between mosaic and unificationist views comes at a price - each view comes with its own set of epistemic costs. From a unificationist perspective, explanatory power should take center stage when deciding to lump or split a scientific phenomenon. If such considerations are taken into account, it might uncover that classifying memory into distinct types limits certain explanatory practices (Najenson, [Forthcoming](#)). Given that, as shown, the search for a unitary mechanism for memory formation is a live research program, when evaluating advances in the memory sciences we must consider that splitting the phenomenon of memory formation and storage into distinct types may not provide the same explanatory power that lumping the phenomenon provides.

Note that this incompatibility is weak. Multiple models of similar phenomena can coexist, each guided by a different explanatory norm.<sup>12</sup> In some cases, one form of explanation can be achieved while the other is still being pursued. In other instances, a model may score high on both unification and completeness compared to other models. Finally, distinct explanatory practices need not be incompatible in the long run. This depends on the extent to which the characterization of the phenomenon is stable enough to allow the appropriate norms to be adopted.

Another source of potential incompatibility involves explanatory representations. Descriptions of mechanisms that differ in their level of abstraction do not always

<sup>12</sup> This normative conception is compatible with the mechanistic approach to explanation. Recognizing that some mechanistic models may be more explanatorily powerful does not diminish the importance of mechanistic explanatory strategies, such as local decomposition. The synaptic efficacy hypothesis does not represent a global model for all explanatory practices concerning memory formation. Moreover, the primacy of synaptic efficacy does not imply a return to reductionism. As I have emphasized throughout, memory phenomena cannot be explained exclusively by synaptic efficacy without considering other neural systems.

hold the same status in explanatory practices. One reason why, in some cases, more abstract mechanistic models may be preferred over more causally detailed mechanistic models is that abstract models allow for the identification of targets that apply to a wider range of tokens without sacrificing the robustness and proportionality of the constructs under study. It is the way these representations are structured that enables them to play this role. On the contrary, a more detailed representation may be limited by its ability to account for changes in background circumstances.

In a similar vein, these considerations also apply to how multiple mechanistic representations are combined. The mosaic view emphasizes hierarchical piecemeal integration, which might overlook the possibility of horizontal combination. In such cases, as Levy notes, “unity is attained via the identification of recurrent explanatory patterns” (2016, p.19). These representations describe mechanisms across modalities and anatomical structures. Knowledge about these mechanisms imposes evidential constraints, which justify inferences on the structure of similar systems and drive investigative trends that are not strictly downward or upward, but *horizontally*, involving tokens of the same type. This form of combining mechanistic representations focuses on what is shared by specific mechanisms across different local multi-level mechanistic hierarchies.

Recognizing that abstraction plays a key explanatory role could help reconcile these views. While Craver acknowledges the *methodological* role of abstraction, it remains an open question whether the mosaic view can accommodate the *explanatory* role of abstraction in specifying mechanistic representations. Somewhat ironically, this issue brings to light the idea that a philosophical theory of scientific conceptualization is required to adjudicate between opposing philosophical theories of explanatory norms.

## 6 Conclusion

Synaptic efficacy is widely regarded in neuroscience as a general mechanism for memory formation and storage, with changes in synaptic strength seen as a paradigm of successful explanation. Craver’s mosaic view, which presents LTP as a multi-level mechanism nested within a multi-level structure, has been central in understanding synaptic efficacy. However, I have argued that the mosaic view does not fully capture the explanatory power of the synaptic efficacy hypothesis due to assumptions about the multi-level structure of mechanistic explanation. Ultimately, these assumptions cause the mosaic view to miss the pattern for the pieces. By considering the explanatory function of unification, we can account for the widespread consensus in neuroscience regarding the synaptic efficacy hypothesis, capturing both its explanatory role across different mechanistic hierarchies and how explaining synaptic mechanisms is informed by causal principles such as those embodied by Hebb’s rule.

## Declarations

**Conflict of Interest** None.

## References

- Batterman, R. W., and C. C. Rice. 2014. Minimal model explanations. *Philosophy of Science* 81(3): 349–376.
- Bechtel, W. 2009. Generalization and Discovery by assuming conserved Mechanisms: Cross-Species Research on circadian oscillators. *Philosophy of Science* 76(5): 762–773.
- Bechtel, W., and R. C. Richardson. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT press.
- Bickle, J. 2003. *Philosophy and neuroscience: a ruthlessly reductive account*. Kluwer Academic Publishers.
- Blanchard, T., N. Vasilyeva, and T. Lombrozo. 2018. Stability, Breadth and Guidance. *Philosophical Studies* 175(9): 2263–2283.
- Boyd, R. 1999. Kinds, complexity and multiple realization. *Philosophical Studies* 95(1–2): 67–98.
- Brigandt, I. 2010. Beyond reduction and pluralism: toward an epistemology of Explanatory Integration in Biology. *Erkenntnis* 73(3): 295–311.
- Caporale, N., and Y. Dan. 2008. Spike timing-dependent plasticity: a hebbian learning rule. *Annual Review of Neuroscience* 31: 25–46.
- Colaço, D. 2022. What counts as a memory? Definitions, hypotheses, and “Kinding in Progress. *Philosophy of Science* 89(1): 89–106.
- Colaço, D., and J. Najenson. (In Preparation). *Where memory resides: Is there a rivalry between molecular and synaptic models of memory?*
- Craver, C. F. 2002. Interlevel experiments and Multilevel Mechanisms in the neuroscience of memory. *Philosophy of Science* 69(S3): 83–97.
- Craver, C. F. 2007. *Explaining the brain: mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.
- Craver, C. F. 2009. Mechanisms and natural kinds. *Philosophical Psychology* 22(5): 575–594.
- Craver, C. F., and L. Darden. 2013. *In search of mechanisms: discoveries across the Life Sciences*. University of Chicago Press.
- Craver, C. F., and M. I. Kaiser. 2013. Mechanisms and laws: clarifying the debate. In *Mechanism and causality in Biology and Economics*, eds. H.-K. Chao, S.-T. Chen, and R. L. Millstein. vol. 3 125–145. Springer Netherlands.
- Craver, C. F., and D. M. Kaplan. 2020. Are more details better? On the norms of completeness for mechanistic explanations. *British Journal for the Philosophy of Science* 71(1): 287–319.
- Darden, L. 2006. *Reasoning in Biological Discoveries: essays on mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge University Press.
- De Brigard, F. 2019. Know-how, intellectualism, and memory systems. *Philosophical Psychology* 32(5): 719–758.
- Feldman, D. E. 2012. The spike-timing dependence of plasticity. *Neuron* 75(4): 556–571.
- Feldman, D. E. 2020. Spike timing-dependent plasticity. In *Neural circuit and Cognitive Development (Second Edition)*, eds. J. Rubenstein, P. Rakic, B. Chen, and K. Y. Kwan. 127–141. Academic Press.
- Francken, J. C., M. Slors, and C. F. Craver. 2022. Cognitive ontology and the search for neural mechanisms: three foundational problems. *Synthese* 200(5): 1–22.
- Friedman, M. 1974. Explanation and scientific understanding. *The Journal of Philosophy* 71(1): 5.
- Gallistel, R., and P. King. 2009. *Memory and the computational brain: why cognitive science will transform neuroscience*. Wiley Press.
- Gershman, S. J. 2023. The molecular memory code and synaptic plasticity: A synthesis. *Biosystems*, 104825.
- Gold, A. R., and D. L. Glanzman. 2021. The central importance of nuclear mechanisms in the storage of memory. *Biochemical and Biophysical Research Communications*, 564: 103–113.
- Hawkins, K. E., C. F. Gavin, and D. Sweatt. 2017a. Long-term potentiation: a candidate Cellular mechanism for Information Storage in the CNS” in Byrne. In J. H. (ed). *Learning and memory: a comprehensive reference*, Academic Press.
- Hawkins, R. D., M. Mayford, and E. R. Kandel. 2017b. A comparative analysis of the Molecular Mechanisms contributing to Implicit and Explicit Memory Storage in Aplysia and in the Hippocampus” in Byrne. In J. H. (ed). *Learning and memory: a comprehensive reference*, Academic Press.
- Hebb, D. O. 1949. *The organization of behavior; a neuropsychological theory* (pp. xix, 335). Wiley.



- Kitcher, P. 1989. Explanatory Unification and the Causal structure of the World. In *Scientific explanation*, eds. P. Kitcher, and W. Salmon. 410–505. Minneapolis: University of Minnesota Press.
- Kreitzer, A. C., and R. C. Malenka. 2008. Striatal plasticity and basal ganglia circuit function. *Neuron* 60(4): 543–554.
- Levy, A. 2016. The Unity of Neuroscience: a flat view. *Synthese* 193(12): 3843–3863.
- Levy, A., and W. Bechtel. 2013. Abstraction and the Organization of Mechanisms. *Philosophy of Science* 80(2): 241–261.
- Magee, J. C., and C. Grienberger. 2020. Synaptic plasticity forms and functions. *Annual Review of Neuroscience* 43(1): 95–117.
- Malenka, R. C., and M. F. Bear. 2004. LTP and LTD: an embarrassment of riches. *Neuron* 44(1): 5–21.
- Morris, R. G. M. 2003. Long-term potentiation and memory. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences* 358(1432): 643–647.
- Najenson, J. (Forthcoming). Memory Systems and the Mnemic Character of Procedural Memory. *British Journal for the Philosophy of Science*
- Perrin, E., and L. Venance. 2019. Bridging the gap between striatal plasticity and learning. *Current Opinion in Neurobiology* 54: 104–112.
- Poo M., M. Pignatelli, T. J. Ryan, S. Tonegawa, T. Bonhoeffer, K. C. Martin, A. Rudenko, L.-H. Tsai, R. W. Tsien, G. Fishell, C. Mullins, J. T. Gonçalves, M. Shtrahman, S. T. Johnston, F. H. Gage, Y. Dan, J. Long, G. Buzsáki, C. Stevens. 2016. What is memory? The present state of the engram *BMC Biology* 14(1). <https://doi.org/10.1186/s12915-016-0261-6>
- Roy, S., and N. W. Park. 2010. Dissociating the memory systems mediating complex tool knowledge and skills. *Neuropsychologia* 48: 3026–3036.
- Rudy, J. W., N. C. Huff, and P. Matus-Amat. 2004. Understanding contextual fear conditioning: insights from a two-process model. *Neuroscience & Biobehavioral Reviews* 28(7): 675–685.
- Salmon, W. C. 1984. *Scientific explanation and the Causal structure of the World*. Princeton University Press.
- Sheredos, B. 2016. Re-reconciling the epistemic and ontic views of explanation. *Erkenntnis* 81(5): 919–949.
- Stinson, C. 2016. Back to the cradle: mechanism Schemata from Piaget to DNA. In *Eppur si muove: doing history and philosophy of Science with Peter Machamer*, eds. M. P. Adams, Z. Biener, U. Feest, and J. A. Sullivan. 183–194. Springer Verlag.
- Strevens, M. 2008. *Depth: an account of scientific explanation*. Harvard University Press.
- Sullivan, J. 2009. The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese* 167: 511–539.
- Sullivan, J. 2016. Long-term potentiation: one kind or many? In *Eppur Si Muove: doing history and philosophy of Science with Peter Machamer*, eds. M. P. Adams, Z. Biener, U. Feest, and J. A. Sullivan. 127–140. Springer Verlag.
- Theurer, K., and J. Bickle. 2013. What's Old is New Again: Kemeny–Oppenheim reduction at work in current molecular neuroscience. *Philosophia Scientiae* 17: 89–113.
- Weber, E., and M. Lefevre. 2017. Unification, the answer to resemblance questions. *Synthese* 194(9): 3501–3521.
- Weber, E., M. Lefevre (Forthcoming), J. Brian, and Hanley, eds. *Philosophical perspectives on causal reasoning in Biology: Minnesota Studies in Philosophy of Science. Vol. XXI*. Minneapolis: University of Minnesota Press.
- Weisberg, M., and M. Weisberg. 2013. *Simulation and Similarity: using models to understand the World*. Oxford University Press.
- Woodward, J. 2010. Causation in Biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy* 25(3): 287–318.
- Woodward, J. 2021. *Causation with a human face: normative theory and descriptive psychology*. Oxford University Press.
- Ylikoski, P., and J. Kuorikoski. 2010. Dissecting Explanatory Power. *Philosophical Studies* 148(2): 201–219.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted

manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.