**UNIVERSIDADE FEDERAL DE SANTA MARIA**
**CENTRO DE CIÊNCIAS SOCIAIS E HUMANAS**
**PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA**

Matheus Diesel Werberich

# THE MECHANISMS OF TIME:
## DELINEATING THE SYSTEMS FOR EPISODIC MEMORY AND IMAGINATION

Santa Maria, RS
2023

Matheus Diesel Werberich

# The mechanisms of time: delineating the systems for episodic memory and imagination

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial de obtenção do título de **Mestre em Filosofia**.

Orientador: Prof. Dr. César Schirmer dos Santos

Santa Maria, RS

2023

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001

Werberich, Matheus Diesel
The mechanisms of time: delineating the systems for episodic memory and imagination / Matheus Diesel Werberich.- 2023.
81 p.; 30 cm

Orientador: Cesar Schirmer dos Santos
Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Ciências Sociais e Humanas, Programa de Pós-Graduação em Filosofia, RS, 2023

1. Episodic memory 2. Imagination 3. Philosophy of memory 4. Mechanisms 5. Philosophy of the cognitive sciences I. Santos, Cesar Schirmer dos II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

Declaro, MATHEUS DIESEL WERBERICH, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Matheus Diesel Werberich

# The mechanisms of time: delineating the systems for episodic memory and imagination

Dissertação apresentada ao Programa de Pós-Graduação em Filosofia, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial de obtenção do título de **Mestre em Filosofia**.

Aprovado em 06 de Julho de 2023

---

**César Schirmer dos Santos, Doutor em Filosofia, UFSM**
Orientador

---

**Raquel Krempel, Doutora em Filosofia, Universidade Federal de São Paulo**
(por videoconferência)

---

**André Sant'Anna, Doutor em Filosofia, Université de Genève**
(por videoconferência)

Santa Maria, RS
2023

*À Ligia e Vera Lúcia (in memoriam)*

## AGRADECIMENTOS

Agradeço, primeiramente, à CAPES pelo apoio financeiro durante a realização do mestrado. Também sou grato ao Departamento de Filosofia da UFSM e a todos os professores que me apoiaram durante o período de Licenciatura e Mestrado em Filosofia.

Dentre estes, merecem menção especial o professor César, orientador desta dissertação e mentor de toda minha trajetória na pesquisa até então, a professora Mitieli e o professor Flávio, cujos incentivos ao longo da Licenciatura foram muito especiais para mim.

Além disso, sou grato à minha família, sem a qual eu não poderia ter me dedicado tanto aos estudos. Meus avós, minha mãe, tia Lila, Max P. e Charlotte. deram-me muito suporte durante esses anos, cada um à sua maneira. Ademais, esta dissertação não teria sido possível sem a ajuda do meu psicólogo, Matheus Rizzatti, e psiquiatra, Iron Pedro Giacomelli. Sem eles, não poderia ter chegado até onde estou agora.

Agradeço também às minha amigas, Ana, Emiliano, Sarah e Wellington, por terem suportado por tanto tempo meus desvaneios sobre a memória e o cérebro — além de terem me dado alguns retrucos que, percebo agora, faziam muito sentido.

Ao meu namorado Max F., dedico todo meu carinho. Sem você, minha vida não teria tantas flores, nem tanto afeto.

Saibam que as seguintes páginas, e todas as outras que ainda virão, são frutos da companhia de cada um de vocês. Muito obrigado.

*This analysis of memory is probably extremely faulty, but I do not know how to improve it.*
Bertrand Russell, The Analysis of Mind, chapter IX

*In the end, we are all in a very intricate game of inference to the best explanation*
Michael Anderson, Brain Inspired, n. 152

## RESUMO

**OS MECANISMOS DO TEMPO**: DELIMITANDO OS SISTEMAS DA MEMÓRIA E IMAGINAÇÃO EPISÓDICAS

**AUTOR**: Matheus Diesel Werberich

**ORIENTADOR**: César Schirmer dos Santos

A memória episódica é um estado mental em que o sujeito possui uma representação imagética de algum evento do seu passado pessoal. Tal representação é geralmente rica em detalhes perceptuais, emocionais e fenomenológicos, além de ser crucial para nossa noção de identidade pessoal ao longo do tempo. Desde Aristóteles, filósofas e filósofos têm se questionado acerca da natureza da memória, em especial sobre a sua relação com a imaginação. No último século, a pergunta sobre se a memória episódica é um tipo de imaginação ganhou considerável destaque, principalmente devido a achados da neurociência cognitiva de que lembrar o passado e imaginar o futuro empregam as mesmas regiões cerebrais. Essa questão, conhecida hoje como *problema (des)continuísta*, dividiu pesquisadores entre *continuistas*, os quais defendem que não há uma diferença fundamental entre memória e imaginação, e *descontinuistas*, os quais defendem que memória e imaginação são estados e processos mentais fundamentalmente distintos. Contudo, na literatura contemporânea se dedicou pouca atenção ao sentido da expressão "fundamentalmente distinto", tampouco a quais critérios são relevantes para delimitar os mecanismos da memória e imaginação episódicas. A presente dissertação preenche essa lacuna ao traçar um diálogo entre a filosofia da memória e a filosofia das ciências cognitivas. Através de três artigos independentes, defendo que o conceito de "mecanismo" é uma ferramenta frutífera para se compreender e responder o problema (des)continuísta. Partindo desta análise mecanicista, argumento que não há critérios livres de interesses pragmáticos para a delimitação de mecanismos neurocognitivos. Por conseguinte, qualquer solução ao problema (des)continuísta é contingente a um determinado *framework* de pesquisa, e devemos ser pluralistas sobre a delimitação entre memória e imaginação episódicas.



**Palavras-chave**: Memória episódica. Imaginação. Filosofia da memória. Mecanismos. Filosofia das ciências cognitivas.

## ABSTRACT

**THE MECHANISMS OF TIME**: DELINEATING THE SYSTEMS FOR EPISODIC MEMORY AND IMAGINATION

**AUTHOR**: Matheus Diesel Werberich

**SUPERVISOR**: César Schirmer dos Santos

Episodic memory is a mental state in which the subject has an imagistic representation of some event from his or her personal past. Such representation is usually rich in perceptual, emotional, and phenomenological details, and is crucial to our notion of personal identity over time. Since Aristotle, philosophers have wondered about the nature of memory, in particular about its relationship with imagination. In the last century, the question of whether episodic memory is a type of imagination has gained considerable prominence, mainly due to findings from cognitive neuroscience that remembering the past and imagining the future employ the same brain regions. This issue, known today as the *(dis)continuist problem*, has divided researchers between *continuists*, who argue that there is no fundamental difference between memory and imagination, and *discontinuists*, who argue that memory and imagination are fundamentally distinct mental states and processes. However, in contemporary literature little attention has been devoted to the meaning of the term "fundamentally distinct", nor to what criteria are relevant for delimiting the mechanisms of episodic memory and imagination. The present dissertation fills this gap by drawing a dialogue between the philosophy of memory and the philosophy of cognitive science. Through three independent papers, I argue that the concept of "mechanism" is a fruitful tool for understanding and answering the (dis)continuist problem. Starting from this mechanistic analysis, I argue that there are no criteria free of pragmatic interests for the delineation of neurocognitive mechanisms. Therefore, any solution to the (dis)continuist problem is contingent on a particular framework of research, and we should be pluralists about the delimitation between episodic memory and imagination.



**Keywords**: Episodic Memory; Imagination; Philosophy of memory; Mechanisms; Philosophy of the cognitive sciences

## CONTENTS

# 1 INTRODUCTION

Noam Chomsky (1976) famously distinguished between two kinds of questions of any intellectual enterprise: *problems* and *mysteries*. Problems refer to questions that, no matter how complex, are understandable and tractable with a certain research framework. They usually arise from the interaction between theory and experience in such a way as to give us (at least) a hint on how to solve them. Mysteries, on the other hand, are largely intractable. They reflect a question that might in principle be unsolvable, or that is currently not solvable with our current methods and concepts.

In philosophy, we deal with both problems and mysteries, often times explicating mysteries into more tractable problems. Philosophical questions often come in an unpacked, disorganized, and messy formulation, generally with undefined terms along the lines which make them all the more challenging to approach. This state of affairs obviates the need to first understand and reformulate the question, and only then to begin to tackle it.

This dissertation is focused on formulating and understanding the (dis)continuism question in philosophy of memory. This question concerns the the relation between episodic memory and imagination. How similar are they? Are they only two instances of the same cognitive capacity? Should we lump memory and imagination into a single category, or keep them split? In the contemporary literature, *continuism* is the view that memory and imagination are fundamentally the same, while *discontinuism* takes that there are fundamentally different capacities.

Constructed as such, the (dis)continuism question could only be considered as a mystery. What is "episodic memory"? What do we mean by "imagination"? Should we think of them as personal cognitive states, or subpersonal systems? If the former, what should we consider when attempting to differentiate them? If the latter, then which properties are relevant when answering whether two capacities are underlined by the same mechanism? These questions are not to be left unanswered. Without them, a great deal of uncertainty will plague (dis)continuism and render the task of solving it all the more difficult.

The present dissertation is an attempt at getting a more precise formulation of this issue via an analysis of the mechanistic literature on the philosophy of neuroscience and the cognitive sciences. My contention is that once we consider how mechanisms are delineated in the cognitive sciences[1], there are inherent pragmatic and perspective-laden constraints that should be accounted for when tackling the (dis)continuism problem. Such constraints are inevitable

[1] I should note that here my scope is what could be called "traditional cognitive science" - that is, the framework that takes the center of cognition to be the brain, wherein representations are manipulated and processed to guide bodily behavior. One of the reasons why this scope is more appropriate for the (dis)continuism problem is that enactive and embodied approaches to cognition have a significant difficulty in explaining memory and imagination without referring to mental representations. Another, more pragmatic reason is that most researchers working on that problem are under the traditional framework.

once we consider how the task of characterizing phenomena is a matter of recognizing patterns, which, by their very nature, are dependent on some perspective to be picked up. Based on these considerations, I conclude that there is no single answer to (dis)continuism - instead, there are multiple and empirically supported ways of fleshing out the differences between episodic memory and imagination.

This introduction is divided as follows. Section 1 contains an introduction for episodic memory and its differences to other forms of memory. Section 2 introduces the notion of imagination as it pertains to (dis)continuism, and discusses the mental time travel framework, which has been greatly influential in contemporary studies of memory and imagination. Section 3 is about (dis)continuism proper, its multiple formulations, and common arguments for continuism and discontinuism. Lastly, section 4 discusses how the dissertation is organized as to answer how can we delineate the mechanisms for episodic memory and imagination?

## 1.1 MEMORY

Philosophy of memory has a very interesting history. While it only recently gained the status of an independent discipline from general philosophy of mind, memory has been a topic of concern for philosophers ever since Ancient Greece (and possibly even before that). Aristotle's *De memoria et reminiscentia*, for example, is a significant work for the development of philosophy of memory, since it is one of the first discussions on how the act of remembering may not be as simple as retriving information from the mind's storehouse (see CHAPPELL, 2017). Picking up on some of Aristotle's considerations, Descartes further developed an account of how past information can be stored in brain tissue. His book on the functioning of the human body, named *L'Homme*, includes a very intricate description of how fluids carried over from our sense organs reach the brain in such a way to create carvings in its tissue. Once these carvings are filled with other fluids, they activate that representation which consists in the subject's memory content (SUTTON, 1998). Further philosophical developments on memory appeared after Descartes' *L'Homme*, but they were mostly smaller parts of more comprehensive theories of mind. Up until recently, memory was not the central subject of philosophical inquiry.

Aside from works of Wrinch (1920) and Russell (1922), our capacity for remembering was not yet in the central stage of philosophical investigation in the first half of the 20th century. It was partially the work of psychologists that changed this memory barren landscape. In the 1950s, Canadian researcher Brenda Milner started her groundbreaking studies with patient HM, who lost his capacity to form new memories (a condition called *anterograde amnesia*) and to retrieve some memories he already had (*retrograde amnesia*). Milner and colleagues studied HM extensively and noted that, despite his severely impaired memory capacities, he was able to learn new motor tasks and retain a list of words for a short-period of time (SQUIRE, 2009; PENFIELD; MILNER, 1958). Such findings prompted researchers to classify memory not as a single activity, but as an umbrella term that includes multiple types of memory. The fact that

patient HM was capable of storing information in short periods of time, but still incapable of keeping them for longer than a few hours, suggested that memory should be split into two general types: *short-term* and *long-term* memories. Moreover, HM's intact ability to learn new motor skills indicated that there are two further kinds of memory: *procedural* and *declarative* memories. These classifications are illustrated as follows:

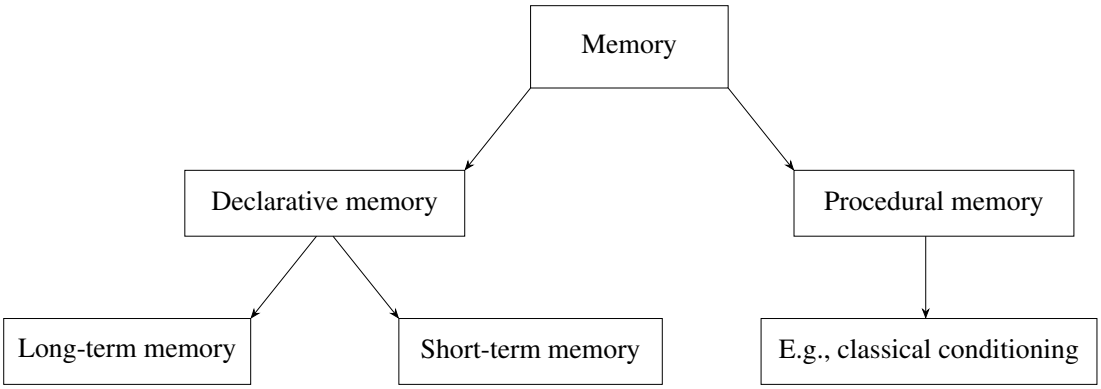


Figure 1 – Memory classification after HM

Contemporarily, there are further divisions within the class of long-term memories. Primarily due to the work of Endel Tulving (1972), long-term memory was divided between *episodic memory* and *semantic memory*[2]. Episodic memories are typically about events located in one's personal past and often consist in imagistic representations about that event. For instance, whenever I remember my 6th birthday party, I can picture with "my mind's eye" the Sponge Bob decoration my mother had made, the smell of cake, and the sound of other children running around. In contrast, semantic memories are void of any significant imagistic content and may not be about one's personal past. For example, I can semantically remember that "the atmosphere of Venus is partially made of sulfur dioxide", even though that information, by itself, does not bring any associated imagery.

The distinction between episodic and semantic memory gained significant more support after studies with patient KC. After a motorcylce accident and severe head injury in the early 1980s, KC lost his ability to form and retrive episodic memories, while his semantic memory capacity was mostly intact (TULVING, 1985). Neuroimaging studies later showed that he had lost significant parts of his medial temporal lobes, including both hippocampi. Not only did such injury cause KC to lose his episodic memory, but also it significantly impaired his ability

[2] Importantly, there are a few authors today that question whether there really is a clear cut distinction between episodic and semantic memory. Aronowitz (2019), in particular, argues that the semantization of episodic memories is a strong indication that they consist in the same general capacity. Despite being a very important issue on its own, the episodic-semantic distinction is not the main topic of the present dissertation and, for the time being, will be assumed to be the case.

to imagine future scenarios. This finding prompted Tulving to hypothesize whether the same neurocognitive mechanism supports both episodic memory and imagination (e.g. (TULVING, 2005); see also (HASSABIS et al., 2007)).

Tulving's hypothesis was further developed with the advent of more precise neuroimaging techniques. Functional magnetic resonance imaging (fMRI) studies showed that there is a strong overlap of active brain areas when someone is remembering the past *vis-à-vis* imagining the future (ATANCE; O'NEILL, 2001; OKUDA et al., 2003; SCHACTER; ADDIS, 2007). These findings also prompted philosophers to ask if there is any significant difference between remembering the past and imagining the future (MICHAELIAN, 2016b; MICHAELIAN; PERRIN, 2017). This debate also hinges on a proper characterization of what we mean by "imagining", which is the topic of the next section.

## 1.2 IMAGINATION

As is the case with the memory literature, debates on nature and processes of imagination take it to be an umbrella-term that encompasses more specific cases of imagining. For instance, Van Leeuwen (2013) distinguishes between three senses of the word "imagine":

- *Constructive imagining* refers to the cognitive process of creating in one's mind a representation of a certain object or event;
- *Attitudinal imagining* consists in taking a certain content as fictional, as not necessarily representing the world as it is;
- *Imagistic imagining* is the mental state of entertaining a representation with some imagistic content.

These types of imagining are not mutually exclusive: we can, and often do, hold mental states that qualify as constructive, attitudinal, and imagistic imaginings. Such is the case when I read Érico Veríssimo's *O Tempo e o Vento*: I mentally contruct the scene of Ana Terra arriving at Santa Fé, which makes me entertain as a fiction the visual representation of a woman amidst a small village. However, as Van Leeuwen (2013) makes clear, not every case of constructive imagination will also be attitudinal imagining, nor will every imagistic imagining necessarily be constructive or attitudinal imagination.

Whenever memory researchers are talking about imagination, they typically are not referring to the multitude of possible imaginings. Instead, they use the term "imagination" as a shorthand for "picturing a future or counterfactual scenario". This is particularly the case in the mental time travel literature, in which to mentally picture a non-present event is taken as a specialized form of imagination that is necessarily self-related and may be about a possible past scenario, or a future event (see SUDDENDORF; CORBALLIS, 2007; SANT'ANNA, 2018).

As noted by Langland-Hassan (2020), this use of the term "imagination" is "not always clear" (p. 67). Does mental time travel involve constructive, attitudinal, or imagistic imagining? Moreover, when we ask if episodic memory is a type of imagination, what do we mean by "imagination"? Recently, Langland-Hassan (2021) argued that the main sense of imagining in the (dis)continuism debate is constructive imagining. His contention is that, when asking if episodic memory is fundamentally the same as imagination, the question is only relevant if we take "imagination" in the constructive sense. If, instead, we take it as attitudinal imagining, then the question has an obviously false answer, for remembering episodically is not the same as considering the content to be fictional. Conversely, the answer becomes naturally true if "imagination" is taken as imagistic imagining: both episodic memory and episodic future, or counterfactual, thought involve mental imagery as described in the concept of imagistic imagining. In the pages that follow, I adopt the constructive sense of "imagination".

## 1.3 (DIS)CONTINUISM

We have thus two of the necessary ingredients for our (dis)continuist recipe: we understand better what the terms "episodic memory" and "imagination" mean and to which mental states they make reference. What is missing to turn the (dis)continuism mystery into a tractable problem is to understand how exactly should we formulate the relation in question. When we ask if episodic memory is fundamentally the same as imagination, what do we mean by "fundamentally the same"? What sort of explication can we attempt to make for making (dis)continuism more understandable? Here, I present four possible solutions:

1. *Is memory imagination?* This formulation is not very precise: what do "memory" and "imagination" mean exactly? Do we mean episodic or semantic memory? Attitudinal or constructive imagining?

2. *Is episodic memory identical to imagination?* Following Leibniz's Law, if $x$ is identical to $y$, then, for any properties of $x$, they are also present in $y$ and vice-versa. Memory is clearly not identical ot imagination in this sense, given that only the latter can represent scenarios that belong to the future or to some possible world. As such, taking (dis)continuism as being about identity trivializes and undermines any potentially fruitful discussion on the relation between memory and imagination. Moreover, the identity relation is inadequate for analyzing neurocognitive systems since, even during the "lifespan" of a particular memory token, multiple processes will change the memory's form and content, so much so that any future remembrance will not be identical to a past memory of the same event.

3. *Is episodic memory relevantly similar to imagination?* Instead of thinking about identity relations, which are unlikely to happen in biological and cognitive systems, the notion of being *relevantly* similar is more inline with how dynamic biological systems tend to be. However, this formulation is still not precise enough, for the notion of "relevant

similarity" is dependent on some epistemic goal. For example, tomatoes are relevantly similar to grapes for the sake of a biologist's classification (i.e., they are both botanical berries (BRITANNICA, 2023)), but not relevantly similar for a chef to make a fruit salad. As such, we need to be clearer under which framework we are talking about memory being sufficiently (dis)similar to imagination. In this dissertation, I approach this issue as it pertains to natural kinds in chapter 3 and to cognitive systems in chapter 2.

4. *Is the attitude involved in episodic memory the same as the attitude of imagining?* Considering other characterizations of the attitude of imagining other than as "taking a representation as fictional" (see LANGLAND-HASSAN, 2021), this formulation of the problem naturally depends on how we characterize propositional attitudes, which is further discussed in chapter 4.

Going forward, I further develop the view that the (dis)continuism problem is one about the natural kinds and mechanisms of episodic memory and imagination. The dissertation is divided into three independent papers. The first paper, "Top-down and bottom-up constraints in mechanistic inquiry", contains a general introduction to the recent philosophy of mechanisms, and how it contributes to the discussion of delineating neurocognitive systems in the philosophy of neuroscience and the cognitive sciences. This article is currently under the first round of reviews at the journal *Filosofia Unisinos*. The second paper, "Do episodic memory and imagination belong to the same natural kind?", is about the formulation of (dis)continuism as a problem about natural kinds. The text was submitted to the journal *Philosophy and the Mind Sciences*, special issue "Successful and Unsuccessful Remembering and Imagining". The third and final article, "Remembering as attitude: an interpretivist view", considers whether a naturalist philosopher of memory should take propositional attitudes into account and makes further conclusions to the view of (dis)continuism as a problem about delineating mechanisms. It currently under second stage of reviews at *Synthese*, special issue "Scientific Realism in Cognitive Neuroscience".

# 2 TOP-DOWN AND BOTTOM-UP CONSTRAINTS IN MECHANISTIC INQUIRY

## 2.1 INTRODUCTION*

Mechanisms are ubiquitous in the empirical sciences, specially in the biological and cognitive sciences. Scientists often describe themselves as uncovering mechanisms for a variety of phenomena - from protein synthesis to global economic recessions. It should then come as no surprise that philosophers of science are particularly interested in the epistemic and ontological role of mechanisms in scientific theories (MACHAMER; DARDEN; CRAVER, 2000; GLENNAN, 2017). Mechanistic philosophy is primarily concerned with questions such as how do mechanisms explain, how do scientists separate mechanisms from their surrounding environment, whether mechanisms are real or just explanatory tools for empirical research, among others.

While mechanistic philosophers have many disagreements on these issues, they widely agree that individuating mechanisms is highly dependent on what we take its phenomenon to be. In short, mechanisms are always mechanisms *for* some phenomenon (MACHAMER; DARDEN; CRAVER, 2000; DARDEN, 2008; GLENNAN, 2017). In this context, "phenomenon" indicates the event that researchers want to explain: from an initial description of the phenomenon, researchers will try to figure out which elements and interactions within that system are relevant for the phenomenon. The complete description of these elements and the interactions between them is a mechanistic explanation for that phenomenon. So constructed, the phenomenon is nothing more than the behavior of the mechanism.

The dependence of mechanistic delineation upon the characterization of phenomena can entail some problems for the objectivity of mechanisms. Considering that characterizing phenomena is dependent on a research project, and that such are dependent on some pragmatic considerations, then mechanistic inquiry is also dependent on pragmatic interests of researchers. As such, mechanistic inquiry will vary significantly between research projects and, hence, may not constitute a completely objective framework for the empirical sciences, as well as inevitably leading to an anti-realist view of mechanisms. In this article, I present some defenses of mechanistic inquiry from such critiques.

The paper is structured as follows. In section 1, I give an overview of the new mechanistic literature, specially as it pertains to the relation between phenomenon delineation and mechanistic inquiry. In section 2, I show how researchers' interests and pragmatic concerns are intrinsically at play when delineating phenomena. How the phenomenon is constructed grounds what Bechtel and Richardson (2010) call *top-down constraints*. The doubts about the realism of mechanistic inquiry stem primarily from such constraints and their intrinsic pragmatic considerations. As

* This article was submitted to the journal *Filosofia Unisinos*.

a response to these doubts, I argue in section 3 that constraints at the physical level also limit the range of possible descriptions of phenomena, thus amounting to an empirical and objective counter-weight to the perspectivalist nature of top-down constraints. Such physical considerations are called *bottom-up constraints* (BECHTEL; RICHARDSON, 2010). In section 4, I analyze two examples of how top-down and bottom-up constraints are employed in actual scientific practice. As an example of top-down constraints, I overview how abstract computational accounts of cognitive processes guide researchers' inquiry into their underlying mechanisms (SHAGRIR; BECHTEL, 2017; DEWHURST, 2018). As an example of bottom-up constraints, I discuss how Biderman and Shohamy (2021)'s work on memory and decision making influenced how the phenomenon of decision making is delineated - instead of having the final behavioral decision as a terminating condition, Biderman and Shohamy (2021) suggest that the mechanism for deliberation is still active well after the behavioral output. Finally, in section 5, I summarize the previous discussions and explore some implications of the present framework to the task of delineating cognitive mechanisms — memory and imagination included. I argue that this issue is primarily, though not only, a matter of perspective.

## 2.2 MECHANISMS

In the dawn of philosophy of science, there was a significant emphasis on how to unify scientific knowledge in a large, overarching framework. Exemplified by Carnap (2017)'s *Der logische Aufbau der Welt*, a main goal of philosophers at the time was to find a way to translate all true scientific statements into a purely empirical language, composed of sense-data and logical laws. Still influenced by the lofty goals of Carnap and others, Oppenheim and Putnam (1958) conceived a hierarchy of scientific enterprises, where "higher" disciplines (such as psychology) could be reduced to "lower" disciplines (such as chemistry and physics). In this framework, the reduction relation carries most of science's explanatory weight: to explain a phenomenon means to reduce it to more fundamental entities and describe their interactions with the laws of physics.

Throughout the twentieth century, however, the reductionist framework in the philosophy of science has steadily declined in popularity. One reason for such decline is how scientific practices in biology, psychology, and other "higher" sciences are notoriously difficult to understand as reductions of a target *explanandum*. Moreover, such sciences often lack the rigid structure necessary for a reductionist explanation: their laws often have exceptions, their classifications are cross-cutting in different ways, and their explanatory tools are not primarily mathematical. Instead, the life sciences seemed more interested in investigating the *causal* relations that produce their *explananda*. Thus, Wesley C. Salmon (1984b), Woodward (1984), and Cummins (1985), just to name a few, shifted their focus to how causal and constitutive relations are prevalent in the life sciences. Not only had reductionism fallen out of flavor, the idea of physics as the ultimate model for the scientific enterprise was also replaced by an emphasis on how fuzzy boundaries are prevalent in "higher" sciences, as well as how individualistic explanations can be.

The transition from physics and reductionism to biology and causal relations paved the way for a mechanistic renaissance in the late 1990s and early 2000s. Instead of focusing on how scientific explanations present law-like generalizations, mechanistic philosophers began to pay attention to how explanations often are about how one particular system works in virtue of its underlying components. In this framework, to explain a system means to decompose it into smaller parts, whose interactions and overall organization make up the behavior of the system as a whole (MACHAMER; DARDEN; CRAVER, 2000). Such decomposition amounts to describing the underlying *mechanism* for that system. In the words of Bechtel and Abrahamsen:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena (BECHTEL; ABRAHAMSEN, 2005, p. 423).

This description of mechanisms makes clear that, for any mechanistic explanation, there are three elements that are individually necessary and jointly sufficient: 1. A description of the phenomenon to be explained; 2. A description of which components are responsible for the phenomenon; 3. An account of how these components are coordinated in such a way to be responsible for the phenomenon.

There are several challenges researchers must face when tackling these aspects of mechanistic explanation. Firstly, phenomena are often not clearly delineated from their surrounding environment. Consider, for example, the phenomenon of a beating heart: not only it is connected with several different parts of the body, it is affected by our breathing reate, stress levels, muscular movements, and so on. This obviates the need for scientists to separate the *explanandum* from other contextual factors that, nevertheless, are causally interactive with the phenomenon. Scientists thus need to separate what they want to explain from the "busy and buzzing confusion that constitutes the causal structure of the world" (CRAVER, 2013, p. 140).

Secondly, and relatedly, among all the causal factors observed interacting with the phenomenon, not all of them will be relevant for its mechanism. For instance, the heart beating is affected by whether we are moving, but bodily movement is not a necessary condition for the heart to beat. Separating what is *actually* relevant for a phenomenon from the myriad of background conditions and spurious effects is a necessary step towards an accurate mechanistic explanation. For this task, we need clear criteria on which components and operations we should include in a mechanistic description, and which we can just abstract away.

Thirdly, after component individuation, we need an account of how they work together in such a way that they are responsible for the phenomenon in question. How the mechanism is related to its phenomenon is the basis upon which we classify general types of mechanisms. *Constitutive* mechanisms *underlie* their phenomena; *etiological* mechanisms *cause* their phenomena; and *maintaining* mechanisms *keep* their phenomena in a homeostatic state (KÄSTNER, 2021; CRAVER; DARDEN, 2013). Importantly, while there may be constitutive and etiological vari-

eties of maintaining mechanisms (KÄSTNER, 2021), constitutive and etiological mechanisms are polar opposites of each other (BAUMGARTNER; CASINI, 2017; KAISER; KRICKEL, 2017). Etiological mechanisms necessarily have a diachronic relation with their phenomena, in such a way that they are temporally prior to their effects; while constitutive mechanisms are necessarily synchronic. Moreover, constitutive mechanisms are not mere aggregates that form their phenomena irrespective of the organization of their parts. Instead, the phenomenon only is produced when its underlying components are coordinated in a very specific way. The *explanandum* thus is related to its mechanism by way of "organizational emergence" (CRAVER, 2015).

These three elements of mechanistic explanations highlight a major feature of this framework: mechanisms are always *for* a given phenomenon (GLENNAN, 2017; DARDEN, 2008; CRAVER, 2015). This platitude indicates how characterizing mechanisms necessarily involves describing a set of activities that, together, make up the phenomenon. In this regard, a heart that doesn't beat, a clock that doesn't move, or a neuron that doesn't fire are not mechanisms precisely because they are not active in any meaningful way - i.e., they don't produce any phenomena. Mechanisms require phenomena, and phenomena are activities that the mechanism does. The initial characterization of the phenomenon, together with underlying assumptions on how it works, are initial and crucial elements that guide researchers' inquiry on how possibly the phenomenon is underlined. Such guidance are what is sometimes called *top-down* constraints, which are the topic of the following section.

## 2.3 TOP-DOWN CONSTRAINTS

As emphasized in the previous section, a necessary part of mechanistic explanations is properly characterizing the phenomenon to be accounted for. Such step is necessary, albeit not solely sufficient, for the fact that there are no mechanisms which are not active in some way. And it is this activity that characterizes the *explanandum* phenomenon - i.e., what the mechanism *does* and what it is *for*.

Given that the end goal of mechanistic inquiry is to achieve a description of how a mechanism is responsible for some phenomenon, an investigation cannot start with mechanisms themselves. They are out there to be empirically discovered and not gratuitously stipulated from the armchair. In this regard, we must start from some preliminary characterizations, assumptions, and observations of the target phenomenon. As we'll see in what follows, these elements constrain the range of possible mechanisms we are able to come up with. They are usually called *top-down constraints* (BECHTEL; RICHARDSON, 2010).

There are multiple ways of delineating a phenomenon and, hence, of establishing top-down constraints. We can characterize phenomena either as a particular causal role in a larger system (CUMMINS, 1985; STICH, 1985); as an etiological function that promotes the survival and fit of a given system (MILLIKAN, 1984; CRAVER, 2013); or as a computational function

that manipulates inputs based on given set of rules and equations (SHAGRIR; BECHTEL, 2017; DEWHURST, 2018; KAPLAN; CRAVER, 2011). While this is not an exhaustive list of all possible ways of delineating a phenomena (see, e.g., Glennan (2017), chapter 5, for a tentative taxonomy of phenomena descriptions), they represent some of the most common ways to fix *explananda* in both the life and mind sciences. These ways of characterizing phenomena should not be understood as mutually exclusive, but rather as possibly compatible ways of delineating a system (KÄSTNER; HAUEIS, 2021). In regard, such plurality of phenomena descriptions is to be expected when we consider that if the causal structure of the world is not readily demarcated for us to discern, there are bound to be more than one possible way of carving a phenomenon apart from the rest of its environment.

Regarding causal role and etiological accounts of phenomena, there is an crucial and somewhat obvious way in which the general environment of the system is a crucial element in delineating *explananda*. On causal role descriptions, the phenomenon is characterized in function of its interactions with the environment: for example, the behavior of the heart to pump blood can only be determined if we analyze how it interacts with other neighboring parts of the system, such as veins, arteries, and blood. Meanwhile, on etiological descriptions, the phenomenon's function in sustaining a given system is dependent on contextual constraints: for example, the etiological function of the heart is to aid in the gas exchange between cells and respiratory system, thus maintaining the organism alive. This function can only be fulfilled if the organism in within an environment that allows for such gas exchange.

Moreover, some philosophers convincingly argue that computational descriptions are also crucially dependent on feature of the surrounding environment. Harbecke and Shagrir (2019), for instance, claim that computational accounts are strongly dependent on the context of explanation in the sense that, if the system were to be placed in a different environment, receiving different types of inputs, it would have to perform a different computational function to get the same output. Such computational contextualism is exemplified in Shagrir and Bechtel (2017)'s analysis on Marr (1981–2010)'s account of edge detection in the human retina: the mathematical function proposed by Marr was only possible by "the observation that in our perceived environment sharp changes in light reflectance occur along physical edges such as boundaries of objects" (SHAGRIR; BECHTEL, 2017, p. 200).

In short, causal role, etiological, and computational descriptions of phenomena require consideration of the surrounding environment to fix mechanistic *explananda*. From this, we can extract some very important features of phenomena delineation and top-down constraints in general. Firstly, phenomena delineation is an empirical matter. We need empirical evidence to make these characterizations, since contextual features are necessary to take into account. Secondly, and relatedly, it is likely that our delineation of the phenomenon will change as research progresses. Thirdly, for re-occurent phenomena, multiple instances of the same phenomena will interact with different environmental elements and, hence, are bound to have some different

properties. Researchers then need a way to say that the same phenomenon is happening in these different circunstances. One way of guaranteeing that is to appeal to how a certain pattern is repeated along these instances.

The concept of *pattern* is relatively recent in the history of philosophy. Philosophers have increasingly taken interest in it after the publication of Dennett's seminal paper *Real Patterns* (1991). In his article, Dennett proposes a response to the problem of whether beliefs exist as real mental entities, or they are only conceptual constructs of an outdated understanding of human psychology, soon to be replaced by neurocognitive descriptions. His answer claims that beliefs, along with other propositional attitudes, are specific patterns of thought and behavior which, despite having multiple different realizations in the brain, are useful abstractions for explaining and predicting behavior.

Beyond the application of patterns to the problem about propositional attitudes, Dennett provides a formal account of what a pattern is. He claims that patterns are a way of compressing information in such a way that unnecessary details are left out and only the most useful pieces of information survive. For example, consider that we want a computer to produce a digital copy of Van Gogh's *Starry Night*. A very unefficient way of instructing the computer would be to make a list of all the pixels it has to fill in, each with a particular color. The computer would get the job done, and its reproduction of the *Starry Night* would be extremely accurate, but that would be tremendously time and energy consuming: a program that instructs how to fill in precisely each pixel would be needlessly long. Another, vastly more efficient way would be to describe a mathematical function on how the stars in the painting are arranged, another on how the village is layed out, etc. This description would be significantly shorter than the previous one, even though it may not be as complete as the former. Once the computer finishes processing these functions, there will probably be spots in the picture that are not identical to the original painting, but it is nevertheless recognizable as similar to Van Gogh's famous work. In this case, the most efficient description is the pattern of the *Starry Night*: it conveys only the relevant information to be close enough to the painting, while some details will be abstracted away.

Interestingly, if we were to ask an human adult for a copy of *Starry Night*, we could just say "Starry Night" and, if they were painting from memory, they would produce something even less accurate than the computer's copy. But, in this case, our description was far more efficient than either the program or the pixel-list, which indicates that there is a trade-off between fidelity of information and efficiency of transmission. These examples illustrate how the same object or system can be described by more than one pattern. Which pattern we ultimately decide to be the most adequate depends on our pragmatic concerns. If we want a perfect copy of the *Starry Night*, the pixel-list would be better; if we don't need such precision, a set of functions roughly describing what the arrangement of elements is would be the most appropriate.

Moreover, the example from the *Starry Night* illustrates two important aspects of patterns: (1) they are real entities in the world, in so far as they can be empirically investigated and, often,

mathematically described; and (2) they are perspective-dependent, in so far as they are relative to the particular methods and frameworks of researchers. Haugeland (1998, p. 274) distinguishes these two aspects of pattern in terms of (1) "orderly and non-random arrangement" and (2) "candidate for recognition". These characteristics are individually necessary, and jointly sufficient, for a given arrangement to constitute a pattern. In particular, the necessity of (2) indicates that an "unobservable pattern" is a misnomer: if there is a pattern in a given system, it should be observable given the adequate methodological tools.

Applying the idea of patterns to mechanistic inquiry, several philosophers have argued that phenomena description is just a way to figure out which patterns there are in a given system (LEE; DEWHURST, 2021). In this regard, considering how overlapping and chaotic causal relations tend to be in the life and mind sciences, delineating a phenomenon involves figuring out which relations and elements tend to persist despite the surrounding noise. Once this pattern is made explicit, researchers are able to make some predictions about the behavior of the system, since patterns are intrinsically non-random regularities that endure as the surrounding context changes (cf. DENNETT, 1991).

It is important to note that recognizing a certain pattern in a causal system is not the same as uncovering the mechanism that produces that pattern. Following Dennett (1991), to understand a pattern we don't need to describe details on how it was produced. Going back to the *Starry Night* example, the pattern of stars arrangement can be described and studied without reference to how the pattern is manifested empirically (either through paint on a canvas, or pixels on a computer screen). This feature is mirrored in phenomena delineation, since, at this stage in mechanistic inquiry, we also still don't know how the phenomenon is implemented; we are just able to describe it in a higher level of abstraction. As such, when we are delienating phenomena, we are describing higher-level characteristics via a pattern meanwhile suspending judgement on the finer details of implementation.

Taking seriously Haugeland (1998)'s idea of pattern as a "candidate for recognition", delineation of phenomena is also dependent on perspective and pragmatic interests. Kästner and Haueis (2021), as well as Potochnik and Oliveira (2020), highlight that how such account of phenomenon delineation has significant implications for mechanistic inquiry. Given that recognizing a pattern is dependent on some perspective, delineating phenomena from their surrounding environment is also dependent on researchers' perspectives and methodologies. For example, Kästner and Haueis (2021, p. 1649) discuss how a "pattern recognition practice" is composed of multiple epistemic activities, such as characterizing a system via operational definitions, or creating schematic models of interactions within that system. Among a community of researchers, these activities need to be coordinated in such a way that they do not become mutually exclusive, but capable of integration into a more definite and repeatable pattern.

As such, not only does pattern recognition depends on the researchers' perspective, but also it is a highly empirical endeavour. It requires both theoretical modelling of known data, as

well as experimental approaches that test these models and gathers more relevant data. Given the interaction between empirical methods and how phenomena are delineated, it is likely that the initial characterization of a phenomenon will be improved as research develops. Bechtel and Richardson (2010) highlight this feature of mechanistic inquiry. For instance, they note how the classical Mendelian view of genes as autonomous determinants of phenotypes had to be significantly altered once deviations from this characterization were discovered (BECHTEL; RICHARDSON, 2010, chapter 8). Such findings not only lead to a greater understanding of the underlying mechanism of genetic determination, but also significantly changed the general research question of genetics: now the phenomenon is couched in probabilistic terms, with much greater appreciation on how genes are not the solely determinants of phenotypes, but instead produce biochemical changes that may lead up to a change of characteristic.

Let's take stock. As we saw in section 1, mechanisms are always characterized in reference to a certain phenomenon. More succinctly, mechanisms are always *for* some phenomenon (MACHAMER; DARDEN; CRAVER, 2000; GLENNAN, 2017). This entails that uncovering mechanisms necessitates an adequate description of the phenomenon under investigation. Such descriptions involve the delineation of the phenomenon from its environment and, as such, are empirical tasks that aim to figure out which pattern exists in a given system. Moreover, if phenomenon delineation is about figuring out patterns, and if patterns are intrinsically perspective dependent, then characterizing phenomena must also be so dependent. This brings forth an important problem for the epistemic adequacy of mechanistic explanations. If phenomena delineation is necessary for mechanistic inquiry and is perspective-laden, then how can we maintain that mechanistic explanations posit real entities? Aren't we thus forced to accept an anti-realist position on mechanisms? In the next section, I argue that the concept of *organizational emergence* provides an useful explication on how empirical evidence constraints mechanistic inquiry and phenomena delineation. Such criteria are called *bottom-up constraints*, and they provide sufficient realist grounds for mechanistic inquiry to block such objection.

## 2.4 BOTTOM-UP CONSTRAINTS

Reference to mechanisms are ubiquitous in the natural sciences. From forming galaxies to synthesizing proteins, mechanisms are often held to be responsible for these and other phenomena. In this vein, mechanistic explanations always start with a characterization of some phenomenon, which, as described in the previous section, is a feature that might bring serious doubts on the objectivity of mechanistic inquiry. If phenomena delineation, the critique goes, is a type of pattern recognition, and if patterns are always dependent on some perspective, then mechanistic explanations are grounded in a perspective-laden activity, thus forcing us to adopt an anti-realist position about mechanisms. Such is the *anti-realist objection*, which is formulated as follows:

*P1.* Mechanistic inquiry explains a phenomenon by describing how a mechanism is responsible for it;
*P2.* If (P1), then delineating mechanisms is dependent on how its phenomenon is characterized;
*P3.* Characterizing phenomena is a matter of pattern recognition;
*P4.* Detecting patterns inherently depends on the perspective of the subject;
$C_1$. *Therefore*, delineating phenomena inherently depends on the perspective of researchers;
*P5.* If (P2) and ($C_1$), then delineating mechanisms inherently depends on the perspective of researchers;
*P6.* If (P5), then mechanisms don't exist independently on some perspective - they are best seen as useful explanatory tools for some research project;
$C_2$. *Therefore*, anti-realism about mechanisms is true.

In this section, I defend that this argument does not hold for mechanistic inquiry. Based on the conception of *organizational emergence*, I argue that P7 does not follow. That is, even if mechanistic inquiry depends on the perspective of researchers, it does not necessarily follow that mechanisms are only explanatory tools. Instead, I defend that it is precisely because mechanisms have firm basis on empirical evidence that they have their explanatory weight. Such evidence, in my view, constrain in a *bottom-up* way the functioning of the mechanism and the production of the phenomenon.

My argument starts with the observation that investigators often move between levels of abstraction to make sure that their description of upper-levels is consistent with what is happening at lower-levels (BECHTEL; RICHARDSON, 2010). The way in which physical descriptions limit the range of higher-level phenomena is called *bottom-up constraint.* As such, bottom-up constraints limit the range of relevant and useful perspectives for the phenomenon. Investigating a system entails figuring out which patterns are applicable to it or not. In this sense, patterns may be perspective dependent, but only partially so: they also depend on which empirical evidence there is available.

To make sense of researchers moving between levels, we need a more detailed account of how mechanistic levels are organized and how they relate to each other. Craver (2015) and Povich and Craver (2018) argue that levels in a mechanism are related to one another via *organizational emergence*. They present their account by contrasting it with a reductionist view of levels, according to which higher-level phenomena are nothing more than the sum of lower-level parts. For example, a reductionist would claim that the temperature of a room simply consists in the average kinetic energy of all molecules in that room. In this case, there is nothing over and above a certain temperature than the movement of molecules; we can therefore say that temperature is only an aggregate function of molecules' energy. Following Craver and Povich, an aggregation of parts does not necessitate any particular organization of such parts. For instance, if $\psi$ is an aggregate of elements in set $X = [x_1, x_2, ..., x_n]$, $\psi$ is reducible to $X$ in such a way that the organization of $x_i$ is irrelevant for forming $\psi$. In sum, if the higher-level is only an effect

of the culmination of lower-level elements, then only the latter carry any explanatory weight. Reductionism is thus incompatible with the idea that higher-level descriptions are explanatory and, therefore, that there are significant top-down constraints in mechanistic inquiry[1].

Let us assume, for the sake of argument, that reductionism is true about levels of mechanisms. As such, any phenomenon, $\psi$, is reducible to the components $[x_1, x_2, ..., x_n]$ in the sense that $\psi$, as well as its properties, are nothing over an above the collection of $x_i$, and any organization of $x_i$ will be sufficient for $\psi$. If such is the case, then a very effective way of studying $\psi$ would be to analyze the properties of each $x_i$ individually, given that the interactions between each $x_i$ do not matter. The end goal of this reductive approach is to ultimately derive $\psi$ from the collection of $x_i$ (POVICH; CRAVER, 2018, p. 191).

However, to understand the mechanistic relation between $\psi$ and $x_i$ as reductive is significantly at odds with the standard characterization of mechanisms: that is, the behavior of the phenomenon, $\psi$, is determined by the properties of $x_i$ *and* their coordinated interactions. In this characterization, $\psi$ is literally more than the simple aggregate of $x_i$, for the interactions between each element $x_i$ is necessary for there to be $\psi$. Consequently, while it might be useful study each $x_i$ in isolation to understand some features of the mechanism for $\psi$, it is not necessary nor sufficient to do so in a proper mechanistic inquiry.

In this scenario, if the reductionist model for mechanisms fail, what should we replace it with? A standard and promising approach is organizational emergence. It consists in the thesis that the phenomenon of a mechanism is dependent not only on the individual parts of the system, but also on how these parts interact with each other. The phenomenon $\psi$ thus emerges from the collection of $x_i$ only if its members are organized and interactive in very particular ways. For example, imagine the audience of a soccer match doing the "stadium wave": successive groups of people stand up, raise their arms, and sit down. From a distance, the coordinated movement of these groups of people forms a wave that can "travel" across the stadium. The wave cannot be reduced to the movement of a particular person, for the very fact that it requires an orchestrated behavior means that it can only be observed once we abstract away from the particularities of each person to the movement of the whole. In this case, the height, speed, and position of a particular member is not relevant for explaining how the wave is formed. We can only explain it from a relatively higher level of abstraction.

Such characterization of organizational emergence fits well with the notion of pattern, discussed in the previous section. The pattern cannot be reduced to the behavior of individual parts because it requires a more distant standpoint to be picked out from (DENNETT, 1991;

[1] Importantly, this defense of organizational emergence and against reductionism concerns only the *epistemic* variants of these theories. In this context, organizational emergence and reductionism are about how we should go about explaining mechanisms: should we focus and reduce them to the behavior of a lower-level component, or should we try to get a bigger picture of the interaction between components? This debate, being about explanatory practices, is distinct from the discussion on *ontological* reductionism and emergentism (FAZEKAS, 2014). In what follows, I will only be referring to the *epistemic* notion of these theories.

HAUGELAND, 1998). If we focus too much on individual parts, we are bound to miss out how the collection of parts is orchestrated as to produce the phenomenon. For example, neuroscientists are now realizing that single-neuron analysis and focusing on one particular brain region at a time is not enough to capture how the enormously complex collection of these elements make up cognitive processes (cf. ANDERSON, 2014; PESSOA, 2022). In this regard, understanding cognitive processes requires a level of granularity that is able to encompass how complex the interaction between brain networks really is. If we focus on single-neuron or even single-area analysis, we don't see the entire network in action and, hence, miss the neurocognitive mechanism we want to capture.

Going back to the anti-realist objection against the pattern account of mechanistic inquiry, organizational emergence provides us with an interesting way of answering that critique. If organizational emergence is true about phenomena and their mechanisms, then to explain phenomenon $\psi$ we need to know how its component parts, $x_i$, are organized. That we cannot do without methodical empirical work. If $\psi$ is strongly constrained by the organization and physical properties of $x_i$, then, were we to discover that cannot support $\psi$ in the way we currently characterize it, then we must change our conception of $\psi$ to accommodate what we know about $x_i$.

Hence, mechanistic inquiry is not completely perspectival - it necessitates bottom-up constraints. Still, if both bottom-up and top-down constraints are necessary for a proper mechanistic explanation, how do they relate with each other? Does one take precedence over the other? How do they interact in actual scientific practice? These are the topics of the next section.

## 2.5 MULTI-LEVEL CONSTRAINTS IN MECHANISTIC INQUIRY

From the previous sections, we have seen how mechanistic inquiry must be *multi-level constrained*: a mechanistic account of a phenomenon must be consistent with our best description of the *explanandum* and with the physical structure of the component parts. In this framework, if an account does not meet either of both constraints, then it does not carry sufficient explanatory weight.

There is an interesting parallel between multi-level constraints and the discussion on whether mechanistic explanations are *ontic* or *epistemic*. Wesley C Salmon (1984a) introduced these distinctions as a way to more clearly differentiate his account with Hempel (1965)'s deductive-nomological view of scientific explanation. Following Salmon, Hempel's account is *epistemic* in so far as it characterizes explanations as arguments, wherein the conclusion amounts to the *explanandum*, and the premises, the *explanantia*. Meanwhile, *ontic* accounts of explanations take them as "exhibitions of the ways in which what is to be explained fits into natural patterns or regularities" (SALMON, Wesley C, 1984a, p. 293). In other words, ontic explanations highlight how the *explanandum* phenomenon is inserted in the causal structure of the world.

In the context of the mechanistic literature, Salmon's terminology has been slightly modified as to be more inclusive. In this framing, epistemic accounts of mechanistic explanations are more concerned with viewing them as an human activity: explanations thus describe mechanisms as a way to improve knowledge about empirical phenomena (BECHTEL, 2008). Alternatively, ontic accounts understand the mechanisms themselves as explanatory, for they are the basis upon which phenomena are inserted into the causal structure of the world (CRAVER, 2007). While most mechanist philosophers tend to agree with Craver, there are still substantial debates between proponents of the epistemic and ontic accounts (see, e.g., HALINA, 2018).

Illari (2013) defends that such characterization of the debate is changing as mechanistic philosophy develops. According to her, new mechanists are increasingly moving away from the traditional understanding of the debate (i.e., as one about what explanations are), towards a framework that asks which explanatory aspects should be given priority. In this vein, epistemic accounts argue that the important constraints on mechanistic inquiry are methodological and, to some degree, relative to the psychological underpinnings of scientific practice; meanwhile ontic views defend that empirical and causal criteria should take precedence over epistemic ones (ILLARI, 2013).

Given such normative twist on the ontic-epistemic debate, it becomes clear how, in reality, both positions are not mutually exclusive. In fact, Illari (2013) herself shows how both Craver and Bechtel can accept each others views, which diminishes the importance of on their disagreement about what criteria are more fundamental. According to Illari:

> It seems that the most sensible conclusion to draw is that neither aim of mechanistic explanation is prior to the other. Ontic and epistemic constraints are both ineliminable, as both aims must be met, to generate a successful mechanistic explanation:
>
> - Describe the (causal) structure of the world: to be distinctively mechanistic, describe the entities and activities and the organization by which they produce the phenomenon or phenomena;
>
> - Build a model of the activities, entities and their organization that scientists can understand, model, manipulate and communicate, so that it is suitable for the ongoing process of knowledge-gathering in the sciences (ILLARI, 2013, p. 250).

Kästner and Haueis (2021) similarly argue that such debate constitutes a false dilemma: mechanistic explanations must attend to *both* ontic and epistemic criteria. Their premises, however, are more metaphysically flavored. According to them, ontic and epistemic criteria are equally important because they constitute fundamental, yet completely distinct, ways of picking out patterns in phenomena and mechanisms. In their words:

> It is not enough for a pattern recognition practice to simply characterize a phenomenon, i.e. what is salient from above. Mechanistic explanation also requires researchers to specify the elements of a pattern, i.e. what makes the pattern persist from below. To achieve this goal, researchers must introduce

> various *epistemic operations* that track the entities and activities constituting the pattern. The selection of such epistemic operations is ontically constrained: scientists must tailor them to the particular spatiotemporal characteristics of the entities and activities they are supposed to track (KÄSTNER; HAUEIS, 2021, p. 1649, emphasis from the original).

Ontic, or bottom-up, constraints are ubiquitous in the empirical sciences. As an example of such criteria, consider the recent research on the relation between memory and decision-making. Starting from a common sense understanding, decision-making consists in a process of evaluating the pros and cons of a set of alternatives in order to eventually make a choice. In this scenario, the process of decision-making has very distinct starting and terminating conditions: we begin with a range of alternatives and, ideally, stop with only one. However, some psychological studies suggest that decision-making does not quite end after an alternative is picked out. In an experiment conducted by Biderman and Shohamy (2021), participants were presented a pair of paintings, A and B, and had to choose which one would generate a higher profit in an auction. After the participant made their decision (say, painting A) and learned the consequent outcome (say, made a profit), they were asked what they thought the value for painting B was. The majority of participants answered that painting B was less profitable than painting A, lending support for Biderman and Shohamy (2021)'s hypothesis that there is an underlying memory component for decision making, which updates the values of unchosen alternatives by association with chosen ones. This indicates that, even after the decision has been made, the system that produced that decision is still active in considering other alternatives. The experiment thus entails something about a component, in this case memory, of the decision making mechanism. From data about this component, we have to change how we think about the phenomenon: in this case, how it is still active well after what we first thought it was the terminating point.

Importantly, changing the description of a phenomenon motivated by evidence of lower-level functioning only necessarily applies to a scientific understanding of these phenomena. With regards to changing any common sense description of a cognitive state, the multi-level framework presented here is neutral. It may be the case that, for some phenomena, there is a real need to change how common sense makes reference to them, but that is not necessarily the case. What is relevant for the present purposes is how the way in which we scientifically understand cognitive phenomena is constrained by lower levels of explanation.

Changing how we conceive of starting and terminating conditions, however, is not the only task of mechanistic inquiry. Top-down and epistemic constraints are also at play when delineating the mechanisms for a given phenomenon. These constraints work by narrowing down the range of possible mechanisms for a phenomenon: from a particular description of a behavior, only some configurations of components and operations are able to produce a phenomenon that matches that description. Once these configurations are properly characterized, scientists can conduct experiments that aim to identify which configuration is actually implemented in the system they are investigating.

One example of top-down constraints is Marr (1981–2010)'s computational analysis of information processing in the human retina. Famously, Marr divided explanations of cognitive phenomena into three levels:

1. *Computational*: characterizes the problem a system has to solve, as well as the mathematical function that resolves such problem;
2. *Algorithmic*: describes how the mathematical function from the computational level can be executed in a series of steps;
3. *Implementational*: details how the algorithmic steps above are implemented in physical hardware.

Note how the Marrian levels form a hierarchy, wherein the lower-levels stand for explanations of how the upper-level works. As such, the algorithmic and implementational levels are finer-grained descriptions of how the system behaves, and a mathematical description of that behavior is given by the computational level. Hence, given that the *explanandum* phenomenon is the overall behavior of the mechanism (MACHAMER; DARDEN; CRAVER, 2000; GLENNAN, 2017), the computational level is nothing over and above than a mathematical description of the phenomenon (see also SHAGRIR; BECHTEL, 2017).

Such quantitative description of the phenomenon is greatly relevant for mechanistic inquiry. From such description, scientists have a more clearly defined target to guide their mechanistic inquiry. The mathematical description of the phenomenon then becomes a parameter against which any mechanistic description must be upheld. As such, if a proposed mechanism does not properly implement the function associated with the phenomenon, it must be discarded as a likely explanation.

Still, there are important caveats with this view. Firstly, scientists may too revise their phenomenon description to accommodate evidence on the physical characteristics of the mechanism, instead of searching for mechanisms elsewhere and saving their phenomenon description. This is what happened with the classical computational theory of cognition, due to McCulloch and Pitts (1943): once we learned that the brain is not suited to serial processing of information (or, at least, not to the same extent as a digital computer), the description of cognition as computation had to be altered to fit the brain's aptitude for parallel and distributed processing (cf. CHURCHLAND; SEJNOWSKI, 1992; PICCININI, 2020; COLOMBO; PICCININI, Forthcoming).

Secondly, a computational description of a phenomenon falls short of a complete mechanistic explanation. Considering how full mechanistic explanations have to account for how the coordinated operations of components yield the phenomenon, any account that does not mention the physical implementation of a process is not properly mechanistic (PICCININI; CRAVER, 2011) In this sense, describing the phenomenon quantitatively and remaining neutral about its physical implementation is only an incomplete mechanistic account - even though it remains as a very important stage in mechanistic inquiry.

In sum, both top-down and bottom-up constraints are individually necessary, and jointly sufficient, for a proper mechanistic explanation. As such, these explanations are *multi-level constrained*, a property which highlights not only how difficult it is to achieve an adequate mechanistic account, but also how valuable these explanations are. They reflect our epistemic criteria and necessities, as well as the causal and constitutive structure of the world.

### 2.5.1 A slight detour into forest metaphysics

The previous considerations on multi-level constraints on mechanistic inquiry have interesting parallels with a wider movement in the philosophy of science, as well as in metaphysics, to admit some amount of perspectivalism into a realist framework (MASSIMI, 2022; LADYMAN et al., 2007; DUPRE, 1996; GLENNAN, 2017). This context allows us to have a deeper understanding of what organizational emergence is, and how realism about mechanisms still holds despite its inherent perspectival character.

Imagine you are a tourist-guider, walking a group of people through the Amazon rainforest. You point to individual trees, animals, comment on their species and importance to the local environment, as a group a curious individuals walks behind you. At some point, one particular tourist asks you: “this is all well and good. You’ve shown us a lot of trees. But where is *the* rainforest?”. This is a category mistake. The tourist, quite literally, missed the forest for the trees. They were seeing the Amazon all along, but only piece by piece. To see the entirety of the Amazon, one cannot be among the trees that make it up.

The forest is an ecological system. This means that it has characteristics that, while not being entirely independent on particular trees, one can only understand these patterns by looking at the larger picture. One needs to change the stance towards the forest. Does the rainforest exist? Or is it a heuristic tool to abstract away the details of individual trees, animals, soil, and so on? An entity/system that is composed of other, smaller, entities/systems still exists. Yet, it is more practical to abstract away the details of particular parts when talking about the whole. The fact that forest-talk is heuristically useful is not incompatible with the view that it exists - despite only existing as a composed or complex entity.

In fact, the very fact that forest-talk is more heuristically useful allows us to discover properties that belong only to the forest, not to any individual tree. For example, the forest is an energetically closed system - i.e., the amount of "waste" as a by-product of energy production is re-used to produce more useful stuff. Think how the forest, by itself, uses the same amount of $CO_2$ as it produced. Nothing escapes. In contrast, individual organisms are biologically incapable of using energy in that way. They produce truly useless waste and need the environment to stock up their energy supply.

We can call the relation between the forest and the organisms that constitute it *organizational emergence*. While the behavior of any individual tree is not determinant for the functioning of the entire forest, the way in which a (large enough) collection of trees behaves is capable of

changing the entire forest. So there seems to be a hierarchy of relevance between an individual tree, a collection of trees, and the entire forest. The entire forest is maximally relevant for the functioning of the forest. The individual tree is minimally, barely, relevant for the forest. A collection of trees may be sufficiently relevant to change the forest, depending on its size and organization.

A such, the entire forest ($A$) is dependent on (some) collection(s) of trees ($B$), and a particular collection is dependent on an individual tree ($C$). Such dependence relation is not transitive: because $A$ depends on $B$, and $B$ depends on $C$, it does not follow that $A$ depends on $C$. And this relation is true of any composed system that exhibits behaviors that are not present in its individual parts.

## 2.6 SUMMARY AND CONCLUSIONS

In the past few decades, the near-omnipresence of mechanisms in the sciences has been thoroughly analyzed by philosophers of science. Such blooming research topic yielded impactful insights for our understanding of the art of explanations. In this chapter, we looked at how mechanistic explanations must be developed in such a way that both meet our epistemic demands and reflect the causal structure of the world. These constraints are called, respectively, *top-down* and *bottom-up*, and together they make sure that mechanistic accounts are adequate for multiple levels of explanation.

Regarding top-down constraints, they consist in researchers' considerations on what the *explanandum* phenomenon is and how it causally interacts with the environment. Given that mechanisms are always mechanisms *for* a given phenomenon, how we understand the latter greatly influences how we go about discovering the former. In this sense, any mechanistic inquiry must start with attempts at delineating the phenomenon in question. This separation of the phenomenon from its environment consists in an abstract understanding of the system under scrutiny, thus limiting the range of possible mechanisms that might be responsible for the system's behavior.

The fact that delineating the phenomenon from its surroundings is a major step during mechanistic inquiry indicates that it is not a task that can be accomplished a priori. As Bechtel and Richardson (2010) note, characterizing the phenomenon is an empirical task precisely because of the fact that we have pick it out from the causal confusion that is the material world. Moreover, considering that these mechanistic phenomena are repeatable across a range of environments (MACHAMER; DARDEN; CRAVER, 2000), phenomenon delineation has to be able to detect what remains the same across these tokens. Kästner and Haueis (2021) argue that such task fundamentally consists in recognizing patterns, where they are understood both as a non-random organization of elements and a candidate for recognition (HAUGELAND, 1998; DENNETT, 1991).

One possible criticism of the patterns account is that it might leave us no choice but to accept anti-realism about mechanisms. If, the critique goes, delineating phenomena is a necessary task for mechanistic inquiry and amounts to a perspective-laden process of pattern recognition, then the discovery of mechanisms is also perspective-laden and may not reflect any real structure in nature. As such, it seems that the inherently perspectival character of patterns is incompatible with a realist position on mechanisms.

Such counter-argument can be blocked by attending to how empirical constraints also have an effect in mechanistic inquiry. These criteria are called *bottom-up constraints*, and they serve as the objective and empirical counterweight to the top-down theory-laden constraints. In this regard, mechanistic inquiry must also attend to whether the physical and organizational structure of the (proposed) components can support the characterization of phenomena that come from the top down. In the same way that cognition cannot be thought of as a serial process of computation because the brain's architecture cannot support this type of processing, an abstract description of a phenomenon has to be abandoned if the underlying hardware is not capable of producing the same behavior.

These considerations have an important impact on how we delineate cognitive systems. Top-down considerations for cognitive science include folk psychological considerations of mental phenomena, computational modelling of cognitive capacities, the overall theoretical framework in which these considerations are inserted, among others. Bottom-up constraints is also quite diverse, for they include consideration of brain structure, the impact of particular neurotransmitters in neural processing, the organization and functioning of perceptual and motor structures, and so on.

# 3 DO EPISODIC MEMORY AND IMAGINATION BELONG TO THE SAME NATURAL KIND?

## 3.1 INTRODUCTION*

In the introduction of *The Order of Things*, Michel Foucault (2005–1966) cites a rather peculiar taxonomy. Taken from a short story by Jorge Luis Borges, the fictional taxonomy includes categories such as "animals that belong to the Emperor", "the ones that from far away look like flies", "mermaids", among many others (FOUCAULT, 2005–1966, p. xvi). The French philosopher uses Borges' example to show how there are multiple ways of organizing entities into distinct categories, not all of which necessarily track real differences in reality.

In contemporary philosophy of science, empirically informed classifications of natural entities are called *natural kinds* (BOYD, 1991). Natural kinds are philosophically interesting for they aim to reflect actual differences and similarities in the world. Such parallel with empirical reality grants natural kinds with an important explanatory and predictive role in the empirical sciences (BIRD; TOBIN, 2022). Paradigmatic examples of natural kinds include fundamental particles, chemical elements and, though more controversely, biological species (ELDER, 2008).

Discussion on classifications can also be found in the philosophy of mind and the cognitive sciences. Questions regarding how to classify mental states and cognitive processes are widespread in the philosophical literature on perception, action and, in particular, memory. Philosophers of memory question whether episodic memory (i.e, the perception-like representations of events in our personal past) and episodic imagination (i.e., the construction of episodic representations of future and counter-factual scenarios) belong to the same kind. *Continuism* is the claim that memory and imagination belong to the same natural kind, while *discontinuism* argues that memory is a different kind of mental state.

Both continuists and discontinuists defend their views on empirical grounds. A common defense of continuism appeals to how episodic memory employs a very similar (if not the same) constructive mechanism to the one of imagination — thereby urging us to lump them into a single kind (see, e.g., De Brigard (2014) and Michaelian (2016b)). Similarly, a frequent defense of discontinuism stems from the claim that only episodic memory, and not imagination, is dependent on a causal mechanism that links past experience with current retrieval (see, e.g., Michaelian and Perrin (2017) and Werning (2020)).

A recurrent feature of the (dis)continuism debate is the homeostatic property cluster theory (HPC) of natural kinds (BOYD, 1991). The HPC account is often used as a theoretical framework upon which we can defend whether episodic memory forms a distinct natural kind

* This article was submitted to the journal *Philosophy and the Mind Sciences*, special issue "Successful and Unsuccessful Remembering and Imagining".

from imagination (CHENG; WERNING, 2016; ANDONOVSKI, 2018). In this paper, I propose an analysis of HPC and how it relates to mechanistic discovery in the life sciences, specially cognitive psychology and neuroscience. I argue that, specially given the dynamic and ever-changing nature of neurocognitive systems (ANDERSON, 2015; DE BRIGARD, 2017; PESSOA, 2022), there is no theory-free demarcation criteria for these mechanisms and, consequently, for natural kinds in the sciences of the mind.

Such account of HPC and mechanisms allows for a more substantive defense of a pluralist solution for (dis)continuism. In this context, pluralism is understood as the claim that there is no uniquely correct way of carving natural kinds. It can be interpreted in two different ways: epistemically or metaphysically (NATHAN, 2018). Epistemic pluralism is the weaker interpretation, according to which the correct use and demarcation of natural kinds is due to the specific characteristics and limitations of a research project. An epistemic pluralist, thus, is not necessarily committed to the view that multiple natural kinds reflect real differences in nature, but only to the the weaker claim that natural kind ascription is correct only if it meets the pragmatic concerns of a specific research project. In comparison, metaphysical pluralism goes a step further and makes the stronger claim that distinct kinds actually do carve nature at its joints. It consists in the thesis that, for a given population of objects under scrutiny, there is more than one way of carving this population into kinds, and that each proposed kind may be consistent with empirical evidence available.

In this paper, I argue that metaphysical pluralism is correct for (dis)continuism about episodic memory and imagination. The paper is structured as follows. Section 2 includes a review of the philosophical literature on (dis)continuism and its different formulations. In this paper, I focus on (dis)continuism as a discussion about the natural kinds of memory and imagination. To understand more clearly what (dis)continuism about natural kinds entails, section 3 contains a discussion on the most widely used notion of natural kinds by philosophers of memory: the homeostatic property cluster theory (HPC). Following Boyd (1991), HPC characterizes natural kinds in terms of a regularly occurring cluster of properties via a homeostatic mechanism. In section 4, I argue for *mechanistic pragmatism*, according to which mechanistic discovery intrinsically involves a pragmatic component. In section 5, I defend that mechanistic pragmatism entails metaphysical pluralism about natural kinds and defend this view from some possible objections. Section 6 is devoted to how metaphysical pluralism influences (dis)continuism about episodic memory and imagination. In particular, I defend that metaphysical pluralism provides promising solutions to the arbitration challenge brought up by Andonovski (2018). Section 7, finally, is reserved for summary and conclusions.

## 3.2 VARIETIES OF (DIS)CONTINUISM

The relation between memory and imagination is not necessarily a new topic in philosophy. Thomas Hobbes, for instance, claimed in the *Leviathan* that “imagination and memory

are but one thing, which for diverse considerations hath diverse names". Recently, the issue of whether memory is a kind of imagining was rekindled partly due to empirical evidence that episodic memory shares a lot of similarities with future and counterfactual imagination. In particular, studies from cognitive psychology and neuroscience suggest that these mental states are dependent on the same brain regions and cognitive processes (MCLELLAND; SCHACTER; ADDIS, 2015; ADDIS, 2020; SCHACTER; ADDIS, 2007). These findings provided further motivation for the *mental time travel* framework in the empirical sciences, according to which memory is just one way among others of mentally transporting oneself to a non-present scenario (TULVING, 1993, 2005).

Drawing from this framework, continuists argued that a philosophical account of episodic memory and imagination should treat them as instances of the same general capacity (MICHAELIAN, 2016b; DE BRIGARD, 2014). According to Michaelian (2016; 2021), the only differences between memory and imagination are that only the former is oriented to the past, while the latter may be about the past or the future; and generally more reliable (or virtuous) in producing accurate representations. Still, continuism claims that these are not sufficient differences to say that memory is different in kind from imagination. Instead, they are merely differences in the degree of accuracy and of temporal directions.

While the previous discussion seems to suggest that the available evidence favors continuism, empirically oriented defenses of discontinuism are also available. Perrin (2016), for instance, argues that the subjective character of memory and imagination are subsumed by different cognitive processes. He supports this claim by noting evidence that suggest important phenomenological differences between episodic memory and imagination (the latter tends to invoke more demanding processing and to fulfill different functions with regards to behavior and emotional regulation), and how the processes for the phenomenology of remembering may be epistemically and causally distinct (PERRIN, 2016, pp. 44, 48).

In a similar vein, Werning (2020) claims that episodic memory and imagination are different in kind because only the former is constructed with guidance of a minimal trace that was formed during the original experience. His account is largely based on the Sequence Analysis for Episodic Memory (CHENG; WERNING, 2016), according to which episodic memory is a temporally extended process that consolidates an event representation from perception for future retrieval. The authors' claim is that episodic memory forms a distinct natural kind on the basis that only it employs a mechanism that instantiates the Sequence Analysis, and that any malfunction of this mechanism would result in a memory deficit.

An important element of the previous discussion is how precisely to individuate a cognitive kind. What are the relevant mechanistic components and operations that we should consider when arguing for the (dis)continuity between memory and imagination? Robins (2020) approaches this issue by defending that (dis)continuism should pay closer attention to the mental attitudes involved in remembering and imagining. Such framework led to a distinction between

*Processual-(dis)continuism*, which asks whether the processes and kinds of episodic memory are (dis)continuous to imagination, and *Attitudinal-(dis)continuism*, which asks whether the attitudes of remembering and imagining are (dis)continuous (MICHAELIAN; PERRIN, et al., 2023). While it is certainly a fruitful venue of research (see, e.g., LANGLAND-HASSAN, 2022; SANT'ANNA, 2021), Attitudinal-(dis)continuism shifts the focus from the mechanisms that generate representations of specific episodes to their associated *attitudes* towards a given representation. While some might take a realist position with regards to propositional attitudes and claim that they are directly instantiated by discrete cognitive processes (e.g. FODOR, 1992), this position is far from forming a consensus in the philosophy of mind[1]. As such, (dis)continuism about attitudes is a different question from (dis)continuism about mechanisms, and it is not clear how exactly they relate to each other.

Therefore, despite being important on its own, I leave Attitudinal-(dis)continuism aside because it is not clear whether propositional attitudes are related to neurocognitive mechanisms or natural kinds. In what follows, I focus exclusively on Processual-(dis)continuism (henceforth, "(dis)continuism" for short) as it pertains to natural kinds. For starters, I analyze an influential account of these kinds to philosophy of memory — the homeostatic property cluster theory.

## 3.3 THE HPC THEORY OF NATURAL KINDS

The question of whether episodic memory forms a natural kind is a relevant issue for (dis)continuism, but it has not been frequently and explicitly analyzed in the literature. A notable exception is the characterization proposed by Werning (2020) and Cheng and Werning (2016) of what counts as a natural kind. According to them, any set, $x$, is a natural kind if, and only if:

1. The members of $x$ are likely to have the same cluster of properties that is explanatory and inductively relevant;
2. Such cluster of properties is subsumed by an uniform causal mechanism;
3. $x$ is the maximal set of entities with such a cluster - i.e., there is no further set which includes $x$ and other members that aren't included in $x$.

Besides Werning's analysis, one of the earliest analysis of memory as a natural kind was made by Michaelian (2011). He defended that the category of "memory" (which includes episodic, semantic and procedural memory) is not a natural kind, even though some declarative memory systems may in fact count as natural kinds. In order to characterize what a "memory system" might be, Michaelian uses Marr's tri-level view of cognitive systems, which identifies a mental process as composed of a computational, algorithmic and implementational levels (MARR, 1981–2010).

[1] See De Brigard (2015) for a comprehensive analysis of this issue.

The previous characterizations of natural kinds are distinct formulations of Boyd (1991)'s homeostatic property cluster theory (HPC). Following HPC, a natural kind is a category of entities that tend to have the same cluster of properties due to an underlying mechanism. Boyd's theory stands in contrast with two other views on natural kinds: nominalism and essentialism.

Nominalism is the claim that there are no "natural" kinds - i.e., our classification practices do not reflect any meaningful distinctions in nature, but are completely based on human interests and purposes (ELDER, 2006). In short, for a nominalist all classifications are nominal kinds. The nominalist position, sometimes also called "anti-realism", is commonly criticized for not being able to account for the predictive and explanatory usefulness of some kinds. If kinds are not based in any empirical consideration, the critique goes, then the reliability of these kinds for scientific purposes may be severely undermined. The HPC theory avoids such threats of nominalism by claiming that natural kinds do refer to genuine differences in nature, namely the mechanism responsible for the clustering of properties, thereby also being useful for making explanations and predictions.

In contrast to nominalism, essentialism affirms that natural kinds reflect deep essences in reality, so much so that they should be described using necessary and sufficient conditions. Boyd's notion of the property cluster being homeostatic is supposed to allow for a bit more variability among members of a natural kind, thus denying that any set of properties is necessary and sufficient to define a natural kind.

Boyd (1991) explicitly argues for HPC as a middle way between the strictness of essentialism, which he condemns as unsuited for sciences such as biology and psychology, and the radical conventionalism applied to natural kinds by nominalists. Boyd thus intends that HPC kinds actually refer to objective characteristics of reality (contrary to nominalism), but in such a way that permits some degree of variability between members of that kind (contrary to essentialism). Primary examples of HPC kinds are biological species. For instance, while there are innumerable differences between each individual human being, we all tend to have roughly the same cluster of properties in part due to genetics, our environment, and our evolutionary history. The lack of any necessary and sufficient conditions for being a *Homo sapiens* permits some degree of overlap between our species and, e.g., *Homo neanderthalensis*, making the boundaries between species vague and fuzzy. HPC acknowledges and welcomes this implication, for it is designed specifically to deal with these issues (BOYD, 1999).

One might question if such permissibility of HPC undermines the predictive usefulness of natural kinds. If there are no necessary and sufficient conditions for a natural kind, then how could we be certain that our explanations and predictions are reliable? According to this line of reasoning, the boundaries of kinds being vague is at odds with the explanatory and predictive functions of HPC kinds. Boyd (1999, 2021) deals with this worry through the notion of *accommodation*. The concept of accommodation is simply that the cluster of properties associated with an HPC kind is grouped by an uniform mechanism. The fact that this mechanism

is, at least partially, discovered by empirical evidence indicates that natural kinds are not "free floating" - i.e., their predictive and explanatory success depends on there being a mechanism underlying it.

However, to meet the permissibility worry with mechanisms could be seen as exchanging one problem for another, since mechanisms themselves may also have vague and fuzzy boundaries. Craver (2009), for example, admits pragmatic concerns into mechanistic discovery and, hence, opens the space for more permissibility about mechanisms. However, it is important to note that his conclusions do not undermine the objectiveness and realist position about mechanisms. The fact that pragmatic concerns play a *partial* role in mechanistic discovery does not entail that they are *all* that matter and that there is absolutely no role for empirical evidence.

To illustrate the concept of accommodation, suppose a group of marine biologists are studying the behavior of whales. After some time collecting data, they are able to make reliable predictions of the behavior of only a certain group of whales (let's call this group *W*). The biologists know that the category *W* is particularly useful for making predictions and that members of *W* tend to habitat the same area and share some physiological properties. However, at this point they cannot be sure if *W* forms a natural kind. This is because the biologists' predictions haven't been *accommodated* in any causal factor - i.e., no uniform mechanism has been discovered that accounts for the detected uniformity among members of *W*. In other words, they don't know if *W* marks a real difference among whales, or if the predictions based on *W* are just statistically lucky correlations with no significant causal import.

Let's now suppose that, after careful genetic analysis, the biologists discover that there is a mutation in gene *y* that is correlated with a significant change in a part of whales' cerebral cortex which is thought to be responsible for their behavior. They also discover that gene *y* is connected with the physiological properties regularly encountered among members of *W*. It is only after these discoveries that our biologists are in a better position to claim that *W* is a natural kind, given that:

1. There is a regularly occurring cluster of properties among members of *W*, such that:
2. If the biologists observe that a given whale, *a*, has some of the properties of *W*, they are able to reliably infer that *a* also has other properties of *W*, including the mutation in *y*;
3. The cluster of properties characteristic of *W* is explained by the mutation in *y*.

In this (admittedly rough) scenario, *W* forms a natural kind because it accommodates the biologists' predictions with a regularly occurring mechanism - in this case, a genetic mutation. This accommodation is precisely the link between a real regularity in nature and our theorizing of it. Accommodating predictions with causal structure is what allows for natural kinds, in the HPC sense, to account for why successful inferences are more than just cases of statistical luck.

Such predictive and explanatory usefulness, together with the flexibility of a homeostatic

cluster, is precisely what makes the HPC theory so appealing to philosophers interested in the special sciences, in particular the cognitive sciences. Fuzzy boundaries are abound in the sciences of the mind, thus making HPC the primary candidate for undertaking the task of carving (neuro)psychological kinds. It is not surprising, then, that authors such as Cheng and Werning (2016) and Werning (2020) defend that episodic memory is a single natural kind with reference to HPC.

The concept of accommodation, together with the characterization of mechanistic discovery from Craver (2009) and Glennan (2017), allow us to draw important conclusions about natural kinds more broadly, and about (dis)continuism more specifically. In the following sections, I argue that metaphysical pluralism is true about natural kinds and, consequently, there is no theory-free way of carving the kinds of episodic memory and imagination. In other words, there are intrinsic pragmatic concerns for deciding what Andonovski (2018) calls the *arbitration criteria* between continuism and discontinuism.

## 3.4 PRAGMATISM AND MECHANISMS

In this section, I defend what I call *mechanistic pragmatism*. It is the view that pragmatic concerns are partially, but intrinsically, involved in setting the boundaries of mechanisms and the phenonema they are responsible for. While this thesis is not novel in the mechanistic philosophy, it tends not to be emphasized enough when dicussing explanations and natural kinds. In what follows, the pragmatist turn to mechanisms takes a more center stage. Moreover, I present a novel argument for mechanistic pragmatism, which connects it with the subdetermination of theories by empirical evidence. The argument can be formalized as follows:

*P1.* Delineating mechanisms require three necessary conditions: (a) components, (b) operations, and (c) phenomenon (MACHAMER; DARDEN; CRAVER, 2000);
*P2.* The phenomenon is the behavior of the mechanism. It is usually characterized as a process with starting and ending conditions (MACHAMER; DARDEN; CRAVER, 2000) or what the mechanism produces when it's active (GLENNAN, 2017);
*P3.* Dynamic systems, such as the brain, are composed of multiple overlapping causal chains;
*P4.* If (P3), then there are multiple possible ways of delineating the starting and ending conditions of a phenomenon in dynamic systems;
*P5.* If (P4), then empirical evidence alone is not sufficient to delineate a single set of starting and ending conditions;
*P6.* If (P1) and (P5), then empirical evidence alone is not sufficient to delineate a phenomenon and, *a fortiori*, a mechanism;
*P7.* If (P6), then theoretical and pragmatic concerns are at play when delineating phenomena;
*C.* *Therefore,* pragmatic interests are partially, but intrinsically, involved in mechanistic discovery (*mechanistic pragmatism*).

There is a lot to unpack in this argument. In what follows, I offer a more precise characterization of the premisses above to better defend mechanistic pragmatism.

### 3.4.1 Minimal mechanisms and phenomena

The first premise above stems from a common conception of mechanisms in the literature, called "minimal mechanisms". According to this conception, mechanisms are an organized collection of component entities whose coordinated operations are responsible for a given phenomenon (MACHAMER; DARDEN; CRAVER, 2000; GLENNAN, 2017). Such characterization is called *minimal* because there are no further constraints on how components should be organized, how long the mechanistic process should take, nor whether the mechanism should be causally or constitutively related to the phenomenon. The lack of such constraints makes the concept of "mechanism" broadly applicable whenever there is a certain phenomenon to be explained by decomposing it into smaller parts (CRAVER, 2009; ANDONOVSKI, 2018).

Most relevant for the present purposes, the minimal characterization of mechanisms makes intrinsic the connection between mechanism and its phenomenon. In other words, for each mechanism, necessarily there is a phenomenon that is produced by it. To say that a mechanism is not causally responsible for any entity or event would be an oxymoron (GLENNAN, 2017).

It is important to note that the previous discussion does not entail that delineating a phenomenon is *sufficient* to delineate a mechanism, but only that it is *necessary*. A complete description of a mechanism must include not only a characterization of the relevant phenomenon, but also a description of the mechanism's working parts and coordinated operations.

### 3.4.2 Phenomena, empirical evidence, and pragmatism

From the previous discussion, it is clear that the first and most important task in mechanistic discovery is delineating what the phenomenon is. Setting the boundaries of what we want to explain restricts the range of possible entites and operations that are likely to be responsible for the phenomenon (GLENNAN, 2017). In this regard, one might ask whether empirical evidence alone is sufficient for delineating mechanistic phenomena. Following Darden (2008), such task is most commonly influenced by perspectival concerns and interests. In this context, even with all empirical data collected, we have to set initial and terminating conditions for what we are studying. This is because such neatly demarcated boundaries are often not there to be found in nature — empirical evidence alone is not sufficient for setting these boundaries (DARDEN, 2008).

Although there is a largely pragmatic flavor in delineating some phenomena and mechanisms, Darden (2008) argues that this is not all there is to mechanistic explanations. Some systems, she claims, have boundaries that are more naturally set than others. For instance, the mechanism for photosynthesis is naturally constrained by a plant's cell, since it is only inside it that we observe that process taking place. However, not all systems necessarily have such clearly

distinguished boundaries. Evolutionary processes, for example, are not easily delimited in either space or time, even though they can be objective and empirically studied.

With regards to neurocognitive phenomena, setting spatial and temporal boundaries is more difficult than in cellular or molecular systems. This is because the distributed nature of how the brain processes information, together with the dynamically changing networks that make up neural systems (DE BRIGARD, 2017), are features that are not quite compatible with clearly distinct boundaries. Assuming the characterization of phenomena due to Machamer, Darden, and Craver (2000), which takes the boundaries of phenomena to be starting and terminating conditions, neurocognitive processes by themselves are not sufficient for us to set their boundaries.

As such, given the dynamicity of the brain, the same cognitive capacity may be instantiated by multiple neural mechanisms and have different sets of starting and terminating conditions in different occasions (ANDERSON, 2015; DE BRIGARD, 2017; PESSOA, 2022). This indicates that mechanisms are not delimited by empirical data alone, but are also constrained by the range of operations we are most interested in explaining. Such is the thesis of *mechanistic pragmatism*.

## 3.5 PLURALISM ABOUT NATURAL KINDS

The previous section argued for *mechanistic pragmatism*, according to which pragmatic interests specific to a research project influence the way through which we carve phenomena and their mechanisms. While such interests alone are not sufficient to delineate phenomena, they are specially important when we're dealing with dynamically changing systems, such as neurocognitive processes. This has important consequences for how we go about classifying cognitive systems and, more precisely, how the (dis)continuism problem might be solved.

As was mentioned previously, an influential formulation of (dis)continuism is based on the HPC theory of natural kinds. Following Boyd (1991), natural kinds are characterized in terms of a homeostatic clustering of properties common across a given population of individuals. What makes this cluster homeostatic, according to HPC, is a common underlying mechanism - it is responsible for the common co-occurrence of properties in such a way that they are not just lucky statistical correlations.

The HPC theory thus relies on mechanisms as an ontological and causal foundation for natural kindhood. However, if the argument of the previous section is correct, such foundation is not entirely free of pragmatic and theoretical concerns. Specially with regards to dynamic systems that have no clear cut boundaries, marking initial and terminating conditions is partially but necessarily a pragmatic task. As such, for each research project there is a specific set of boundaries for a phenomenon that delimits the range of possible mechanisms and, hence, the natural kinds that are possibly marked.

The inherent perspectivalism of mechanistic demarcation entails that there is more than one correct way of carving boundaries of mechanisms and, consequently, of HPC kinds. Such pluralism about natural kinds can be understood in either of two ways: *epistemically* or *metaphysically* (NATHAN, 2018). According to epistemic pluralism, there are multiple correct natural kinds (about a specific population) because each best serves the explanatory and predictive goals of a specific research project (see Kitcher (1984) for an epistemic pluralist account of biological species). Conversely, metaphysical pluralism goes a step further and claims that those natural kinds reflect actual differences in nature. Dupre (1996) names this view "promiscuous realism", and it highlights how natural kinds may be interest-dependent but still have a counter-part in reality.

Metaphysical pluralism is more in line with the previous discussion on *mechanistic pragmatism*, since the perspectivalism of the latter cannot be dispelled simply as a methodological issue. The fact that more than one perspective is compatible with empirical data suggests that more than one demarcation can carve a real difference in nature and, hence, that there being multiple natural kinds is not a by-product of our incomplete knowledge. Instead, specially when it comes to dynamic systems, there are multiple possible sets of starting and terminating conditions that are compatible with data and yield productive classifications. Importantly, the claim that different research projects carve the boundaries of mechanisms differently does not necessarily entail that they are at different levels of abstraction. Rather, it is possible that two research projects focus on different properties of a system but at the same level of abstraction, possibly delineating distinct kinds if there are different mechanisms responsible for these properties.

One might object to metaphysical pluralism by claiming that it is only nominalism in disguise. Following a previous characterization of nominalism, it is the claim that there are no "natural" kinds - instead, all proposed classifications are only nominal and, as such, do not reflect any real difference in a given domain. The main difference between nominalism, so understood, and the metaphysical pluralist view defended here is that the latter takes empirical evidence as an essential part of carving natural kinds, while the former does not. Metaphysical pluralism, thus, is a hybrid view that takes pragmatic *and* empirical considerations as intrinsically relevant for natural kindhood and, hence, avoids the arbitrary characterization of nominalist kinds. Moreover, while nominalism entails that there is no strictly objective way of demarcating natural kinds, metaphysical pluralism is compatible with such objectivity, but only limits it to a particular research project - i.e., for one particular explanatory goal, there are objective reasons, such as empirical evidence and explanatory fruitfulness, to adopt one classification over another.

Another objection to metaphysical pluralism concerns whether the fact that there may be multiple correct natural kinds is just a consequence of imperfect knowledge. Following this counter-argument, if we were to have complete knowledge about a certain set of phenomena, there would be no question as to which natural kinds we should adopt. Hence, the possibility of there being more than one correct natural kind is not a necessary fact of nature, but only a

consequence of our incomplete knowledge. A metaphysical pluralist can respond to this argument by noting that, following *mechanistic pragmatism*, delineating phenomena and mechanisms is heavily influenced by pragmatic interests. It is partially in our hands to decide which are the relevant starting and ending conditions, as well as to decide at which level of abstraction we should be investigating. In this sense, there is no escaping imperfect knowledge - setting natural kinds is partially, but intrinsically, a matter of deciding which commonalities and differences we should take into account. These concerns are essential when creating taxonomies and models that are supposed to be generally applicable to a certain population - if we were to take every single property into account, every individual would be a kind into itself.

## 3.6 METAPHYSICAL PLURALISM AND (DIS)CONTINUISM

The previous sections argued for *metaphysical pluralism*, the view that there are multiple natural kinds which reflect real differences in nature. The argument started from the premise that the pragmatic concerns of carving natural kinds lie precisely on the ontological bedrock of these categories. Such concerns are not to be dismissed as simple methodological issues, but are intrinsic to the way in which we carve mechanisms. Therefore, metaphysical pluralism is the case for natural kinds.

Metaphysical pluralism has important consequences for (dis)continuism. Considering that (dis)continuism is about whether episodic memory and imagination belong to the same natural kind, the pluralist view defended here entails that there is no interest-free way of answering this question. In this section, I analyze the prospects of a pluralist account of (dis)continuism. In particular, I argue that pragmatic interests play an intrinsic role in Andonovski (2018)'s arbitration criteria.

As was stated in previous sections, continuists claim that episodic memory and imagination belong to the same natural kind, while discontinuists claim that they don't. One influential defense of discontinuism stems from the Sequence Analysis of Episodic Memory, defended by Cheng and Werning (2016), whose underlying mechanism involves a Minimal Trace, which carries information from the past experience to current remembering[2].

The Sequence Analysis is quite similar to the standard view in psychology that divides memory processing into perception, encoding, storage and retrieval (CHENG; WERNING, 2016, p. 10). The properties of the Sequence Analysis are thought to be accommodated by a mechanism of encoding and retrieving a minimal trace. A closer view on the Minimal Traces mechanism indicates that it can be divided into three distinct stages (WERNING, 2020, p. 326): (a) *Alignment*

[2] In the memory literature, traces are what is formed during a past experience and are causally active in current remembering. What exactly is the ontological status of memory traces, and how they carry information about the past event, is still a heavily debated topic among memory researchers (see Robins (2017) for a comprehensive review). Werning calls his account "minimal traces" because they only carry minimal and non-representational information about the past experience. Whether minimal traces are actually able to carry such information without representations is an important topic, but only tangential to present purposes.

(when hippocampal place cells form the minimal trace); (b) *Consolidation* (stabilization of the memory trace); and (c) *Reconstruction* (the activation of the hippocampal trace coordinates the reconstruction of perceptual representations).

Werning's analysis of the mechanism for episodic memory also explains why there are so many similarities with episodic imagination. Because both memory and imagination employ *Reconstruction*, one can expect that the brain regions employed in remembering and imagining greatly overlap (SCHACTER; ADDIS, 2007) and that there are significant phenomenological similarities between the two mental states (D'ARGEMBEAU; VAN DER LINDEN, 2006).

Andonovski (2018) critiques such defense of discontinuism. He notes that, while the Sequence Analysis is able to *track* a homeostatic mechanism for episodic memory, it ultimately fails to *arbitrate* between competing accounts. While it may provide reasons to accept the existence of such mechanism (the tracking condition), it is not sufficient to conclude that we should adopt this taxonomy over another (the arbitration condition). For Andonovski, the Sequence Analysis does not meet the arbitration condition because it stems from a conceptual analysis of memory - since there are several ways of conducting such analysis and Cheng and Werning (2016) do not specify why we should endorse their analysis over others, there is nothing in their account that prevents us from lumping episodic memory and imagination into a single natural kind. Andonovski illustrates this by noting how the CRISP theory of hippocampal function, formulated by Cheng (2013), supports continuism by providing a mechanism that accommodates the features highlighted by simulationism (MICHAELIAN, 2016b).

This discussion prompts Andonovski (2018, p. 11) to ask "which taxonomy should we choose? Are there reasons to privilege one (description of a) homeostatic mechanism?". In what follows, I explore how metaphysical pluralism can help us navigate this question. I argue that different ways of carving the natural kind of episodic memory stem from different considerations and interests of what the function of memory is and which level of explanation is required to explain these functions.

### 3.6.1 Pluralism, functions, and levels of explanations

While Andonovski (2018) seems to consider some possible satisfactory set of conditions for *arbitration*, metaphysical pluralism suggests that such conditions are highly specific to a research project and its pragmatic concerns. As argued earlier, the dynamic nature of neurocognitive processes suggests that there are multiple possible and empirically adequate ways of carving cognitive kinds. The choice of one over the other doesn't borne out from data alone, but also from what we want to explain and how fruitful a given classification is.

My contention with metaphysical pluralism is that meeting the arbitration *desiratum* is theory relative. If individuation of phenomena is project relative and involves setting up starting and ending conditions, then it involves looking at the effects of memory to the rest of the cognitive environment. These effects track different possible functions of memory (BOYLE,

2021; MAHR, 2022), some of which are the same with regards to imagination. So mechanistic and kind individuation of episodic memory depends on what we take the functions of memory to be. Some of these functions will be more higher level than others.

For instance, proponents of the modelling function claim that the role of memory is to learn models and regularities from past experience in a way that improves the likelihood of relevant retrieval (ARONOWITZ, 2019). This function is at a lower level than the communicative function (which claims that memory is a way to exert epistemic authority about the past (MAHR; CSIBRA, 2018)) - since the latter envolves whole individuals, while the former deals with representations. Alternatively for other functions, such as providing elements for future mental time travel (SUDDENDORF; CORBALLIS, 2007), it would be more fruitful to highlight memory and imagination in a way that is more coarse-grained and makes clear how they are connected.

When taking the proposed functions of memory into account, discontinuism seems finer-grained than continuism. The former takes the process of generating a memory representation as constructive, but still different from imagination on the grounds that it necessarily involves a memory trace. Continuism disregards this (possible) difference, on the grounds that it not relevant when accounting for memory errors or the observed overlap between episodic memory and imagination (SCHACTER; ADDIS, 2007). Thus, to solve Andonovski (2018)'s concern on arbitration, we need to demarcate which level of abstraction is more adequate for our classificatory purposes. But, as I argued previously, pragmatic interests are intrinsically at play in this regard.

From the previous discussion, one could counter argue that natural kinds, in the HPC sense, are simply unable to arbitrate between continuism and discontinuism[3]. It would be better, the counter-argument goes, to eliminate the formulation of (dis)continuism in terms of natural kinds instead of adopting a metaphysical pluralist account. While this is certainly an alternative to consider, I argue that it is not necessary. The pluralist account of natural kinds as partially interest-relative does not exclude their explanatory and predictive fruitfulness *tout court*, but only restricts them to a particular research project. As also noted by Taylor (2021) and Boyle (2022), the relativity of classifications does not entail that "anything goes", or that natural kinds should be eliminated for lack of objectivity. In short, the claim that more than one natural kind carve real differences does not entail that *all* do so equally well for *all* explanatory projects.

Another possible counter-argument relates to how there are multiple levels in any given taxonomy. Often taxonomies are comprised of higher order classifications (e.g., mammals) that include more specific *taxons* (e.g., *Homo sapiens*). If that is the case, and if the kind of *episodic simulation* includes processes that a discontinuist would say that belong to the kind *episodic memory*, then why not consider *episodic simulation* as a *genus* that includes the species *episodic memory*? Why should we adopt pluralism in this case?

Considering that higher-order classifications can still be natural kinds in the HPC sense

[3] A similar proposal for emotions and concepts is analyzed by Taylor (2020).

(BOYD, 2021), we have to delineate the mechanism that regularly clusters the properties associated with a given *genus*. As such, clearly characterizing the *genus* of *episodic simulation* entails setting the boundaries of the mechanism associated with this classification.

Whether or not the species of episodic memory fits well within that *genus* ultimately depends on (1) what are the mechanism and the property cluster of *episodic simulation*, and (2) if episodic memory actually has this property cluster and is underlined (at least partially) by that mechanism. Both of these issues involve setting the boundaries of mechanisms, which, as stated previously, is highly dependent on our research interests. This means that, on some research projects, episodic memory will fall under the scope of *episodic simulation*, but, on others, it might not. Therefore, even if we take higher-order taxa into account, we have to accept some form of pluralism.

## 3.7 SUMMARY AND CONCLUSIONS

The main topic of this paper was the (dis)continuism problem in philosophy of memory. While there are many different formulations of (dis)continuism currently available (MICHAELIAN; PERRIN, et al., 2023), I focused on an more standard, though still influential view of the problem as one about natural kinds (WERNING, 2020; ANDONOVSKI, 2018). Under this framework, philosophers ask whether episodic memory belongs to the same natural kind as imagination, with continuism being the affirmative answer, and discontinuism, the negative.

I argued that a closer analysis of how natural kinds relate to mechanisms leads us to the conclusion that continuism and discontinuism are not mutually exclusive. Dedicating more attention to the HPC theory of natural kinds (which is broadly used by philosophers of memory) reveals to us that a certain category, *c*, is only a natural kind if the inferential practices about members of *c* are accommodated by a regularly occurring mechanism. Given that inferential practices are heavily influenced by research interests (BOYD, 2021), and that there is no single objective way of setting the boundaries of mechanisms (CRAVER, 2009), delineating natural kinds becomes a task that is heavily project specific.

This opens the logical space for metaphysical pluralism about (dis)continuism. In this view, whether episodic memory is a kind of imagination strongly depends on which aspect of these mechanisms we are focusing on. For example, if we are interested at how the mechanism for episodic memory is influenced by previously stored trace, then it is more useful to consider memory as being different in kind from imagination. On the contrary, if we are researching how remembering is a highly manipulative and active process, then the similarities with imagination become all the more important as to unite them in a single natural kind.

# 4 REMEMBERING AND IMAGINING AS ATTITUDES: AN INTERPRETIVIST VIEW

## 4.1 INTRODUCTION*

One of the main questions of philosophy of memory today is the relation between episodic memory and imagination. This problem, named *(dis)continuism*, has resulted in numerous discussions on the particular similarities and differences between remembering the past and imagining the future (MICHAELIAN; PERRIN, 2017). Philosophers are particularly drawn to findings in the cognitive sciences, which tend to highlight the overlap between memory and imagination. In this context, developing a philosophical account that is in line with the empirical sciences is one of the main tenets of methodological naturalism, which has been influential in how we philosophically think about memory today (MICHAELIAN, 2016b).

In light of this methodology, Robins (2020) argues that we should consider the different attitudes involved in episodic memory and imagination. Given that, she claims, attitude terms are frequently used to coordinate participants' behavior in cognitive science experiments, a naturalistic philosopher of memory should take them into account. While I agree with her assessment, I worry that just pointing to the use of attitude terms in the sciences of the mind is not sufficient for taking them seriously in a naturalist background. In particular, what seems to be missing from Robins' argument is a more general account of how these attitudes are individuated and which explanatory role they play in our theorizing about cognition. The main purpose of this paper is to fill this gap. In the pages that follow, I argue for *interpretivism* about mental attitudes and, with it, provide a more robust defense of mental attitudes in naturalistic philosophy of memory.

This paper is structured as follows. In section 4.2, I offer an overview of the background for (dis)continuism and the current state of philosophy of memory. In section 4.3, I present an argument against the inclusion of attitudes to naturalistic philosophy of memory, as well as the relevant notions of methodological naturalism, mental attitudes, and folk psychology. In section 4.4, I present the interpretivist view on mental attitudes, and defend it against possible objections. In section 4.5, I explore how interpretivism leads us to adopt a pluralist view on (dis)continuism, as well as how it provides a novel argument for the relation between epistemic and empirical meanings of remembering. Finally, section 4.6 is reserved for summary and conclusions.

## 4.2 BACKGROUND

Among the many capacities of the human mind, the ability to re-experience events from the past is one that has puzzled philosophers and scientists for centuries. Ever since the early

* This article was submitted to the journal *Synthese*, special issue "Scientific Realism in Cognitive Neuroscience".

1970s, this type of remembering has been dubbed "episodic memory", which consists in the state of entertaining a mental representation about an event in one's personal past (TULVING, 1972, 1985). Contemporary philosophy of memory has predominantly focused on several issues regarding this form of memory, in particular about its relation with imagining the personal future and counterfactual scenarios (MICHAELIAN; PERRIN, 2017). Functional imaging data from cognitive neuroscience strongly suggest that there is a substantive overlap in the brain regions supportive of remembering the past, imagining the future, and picturing a counterfactual event (SCHACTER; ADDIS, 2007; MCLELLAND; SCHACTER; ADDIS, 2015; DE BRIGARD et al., 2013). Moreover, there is substantive evidence that episodic memory and imagination share a similar phenomenology (D'ARGEMBEAU; VAN DER LINDEN, 2006), and that they have similar development trajectories in children (ATANCE; SOMMERVILLE, 2014). Motivated in part by these studies, some philosophers defend *continuism*, the view that there are no fundamental differences between episodic memory and imagination (MICHAELIAN, 2016a; DE BRIGARD, 2013). In contrast, other researchers argue for *discontinuism*, which claims that memory and imagination are fundamentally distinct (PERRIN, 2016; BERNECKER, 2010). The debate between both theories is currently named *(dis)continuism*.

The (dis)continuism debate sparkled a series of related methodological discussions between philosophers of memory. Given its close proximity with the cognitive sciences, it has been standard to view philosophy of memory and (dis)continuism as instances of naturalized philosophy (MICHAELIAN, 2016b; ROBINS, 2020). In this scenario, the methodology of naturalism stands for a type of theoretical constraint: only accept into an ontology the entities present in the relevant scientific explanations — such as cognitive psychology and neuroscience (ROBINS, 2020; KEELEY, 2016; KORNBLITH, 2002).

From the previous discussion, it seems as if methodological naturalism can only support continuism about memory and imagination. If we are only allowed to make ontological commitments about entities that have empirical support, and if the relevant evidence from cognitive science suggests that there are no major differences between the systems for remembering and imagining, then a naturalistic philosopher would have to commit to the view that there is no major difference between memory and imagination and, hence, that they are fundamentally the same.

Against this view, Robins (2020) argues that a naturalist methodology does not necessarily preclude discontinuism. Her claim is based on the fact that many of the neuroimaging studies cited by continuists (e.g., SCHACTER; ADDIS, 2007; DE BRIGARD et al., 2013) use a vocabulary that implies a difference between the attitudes of remembering and imagining. Robins notes that, once inside a fMRI machine, participants are asked to remember a past event, or to imagine a future or counterfactual scenario. Given that participants are able to tell when they are remembering instead of imagining, and that they competently follow researchers' instructions, it is likely that there is a distinction at the personal level between memory and imagination.

According to Robins this distinction is due to their different mental attitudes. As such, given that neuroimaging experiments exploit this differentiation, a naturalist philosopher of memory should take it into account and, therefore, not jump into the conclusion that episodic memory and imagination are fundamentally the same.

While I agree with Robin's general claim (i.e., that neuroimaging studies alone do not necessarily entail that continuism is true), I believe that there are a few potential issues with her argument. Even assuming that mental attitude terms are present in the framework of cognitive science, it is not clear if they have a relevant explanatory role, or are just tools for coordinating the participants' behavior. The main idea here is that not every concept in a scientific framework serves a *relevant* explanatory purpose. Following a standard characterization of the naturalist methodology, we should only make ontological commitments to those entities and characterizations that have relevant explanatory roles in out best scientific theory available. In this regard, given that scientists often postulate entities and create characterizations that are, sometimes knowingly, filler elements to be replaced in the future[1], not every concept in an empirical framework has the same explanatory weight. Hence, there are some elements in such a framework that should not be accounted for by a naturalist philosopher. Robins' argument takes for granted that mental attitudes are relevant terms in cognitive science, but it is not obvious that it is so. This concern has been voiced before in the philosophy of mind, most famously by Churchland (1989) and Stich (1985), which indicates that a more explicit argument for the relevance of mental attitudes to cognitive science and naturalistic philosophy is in order. This critique forms the basis of the *exclusion argument*, which will be developed in detail in later sections.

Therefore, even if a naturalistic philosopher should take both personal and sub-personal phenomena into account, we still need a conceptual structure that more clearly articulates the explanatory function of mental attitudes to cognitive science, as well as its connection between mental attitudes at the personal level and the search for neurocognitive mechanisms at the sub-personal level. In this paper, I argue for *interpretivism* as a solution for this problem. Based on Dennett's (1991) and Haugeland's (1998) discussion on patterns, and their application to mechanistic philosophy of cognitive science (KÄSTNER; HAUEIS, 2021; FRANCKEN; SLORS, 2014), interpretivism consists in the thesis that mental attitudes are ways of talking about real patterns in subjects' general behavior (which includes bodily and mental behavior) and their underlying mechanistic structures. In this scenario, mental attitudes are understood as patterns that, once detected, constrain the range of possible mechanisms for which researchers should investigate. This criterion, called *top-down constraint* (BECHTEL; RICHARDSON, 2010), serves as a clear connection point between talk of attitudes at the personal level and the search for mechanisms at the sub-personal level, thus answering the exclusion argument. Such framework is also capable of making clearer how epistemic and empirical meanings of

[1] See Bechtel and Richardson (2010) for more discussion on such cases.

'remember" are connected: arguing against the incompatibility thesis of McCarroll, Michaelian, and Nanay (2022), interpretivism supports the claim (also made by Craver (2020)) that epistemic and empirical remembering are compatible descriptions under the same perspective.

## 4.3 THE EXCLUSION ARGUMENT

The argument for the exclusion of attitudes from naturalistic philosophy of memory can be so formalized:

1. Methodological naturalism consists in the normative thesis that, for any $x$, it is necessary that $x$ plays a relevant explanatory role in our best scientific theory for philosophers to make ontological commitments about $x$;
2. Mental attitudes, such as "remembering" and "imagining", are expressions from folk psychology used to coordinate intersubjective behavior;
3. If (P2), then scientists use the terms "remembering" and "imagining" as a way to coordinate participants' behavior and are not explanatory in any relevant sense;
4. If (P1) and (P3), then a naturalistic philosopher of memory should not make ontological commitments based on mental attitudes;

*C.* Therefore, naturalistic philosophy of memory should not be concerned about the attitudes of "remembering" and "imagining".

There is a lot to unpack in this argument. In the subsections that follow, I analyze each proposition to show how they support the exclusion of attitude-talk from naturalistic philosophy of memory. After that, I present the interpretivist account as a way of arguing against (C).

### 4.3.1 P1. Methodological naturalism

The overarching naturalistic claim with respect to methodology is that philosophy should be in continuity with the empirical sciences. Following Kornblith (2002, 2017), naturalism is vindicated by a particular metaphilosophical thesis: philosophy's object of study are the phenomena in themselves, not the concepts we use to refer to these phenomena. In this sense, epistemology is not concerned with the concept of "knowledge", but with knowledge itself - i.e., how individuals come to acquire it, and which are the common characteristics across tokens of knowledge. Similarly, the metaphysics of time tries to understand time itself, and not the concepts we use to refer to time.

While the naturalistic thesis about the object of philosophical study may not be entirely accepted by some philosophers (e.g. THOMASSON, 2017), it does lead to some interesting results and fruitful venues of research. In particular, taking philosophy as concerned with

phenomena in themselves leads us to endorse that the philosophical enterprise is, ultimately, an empirical matter and should be conducted in close proximity with the empirical sciences. As such, metaphysicians working on time should work with physicists to discover what the nature of time is; epistemologists should collaborate with psychologists to understand the cognitive processes behind knowledge, and so on. As Quine (2004) puts it, philosophy is a highly abstract branch of empirical science.

So far, what we have is a general account of how philosophy should be conducted in relation to the sciences. But pointing to the relation does not make clearer how it should be conducted. What kind of evidence should we take into consideration? What is the role of folk concepts and intuitions in naturalistic philosophy?

In the first chapter of *Knowledge and its place in nature*, Kornblith (2002) discusses these issues directly. In particular, he claims that whenever we are investigating a given phenomenon, our research progresses as we gather more empirical data and are able to come up with theories that accommodate and explain what we observed. While our end goal is to create a theory that is adequately informed by empirical data, to get there we need some way to refer and talk about the phenomenon we are studying, even if such way of speaking is generally imprecise and can be revised as more research is done. As such, folk concepts and intuitions play a very important role at the beginning of a research project, mostly as tools to fix *explananda* and to heuristically create hypothesis about what's being observed. As research progresses, and more experiments are made, our initial and ordinary understanding may be revised in light of what was found. This is what Bechtel and Richardson (2010) call "reconstituting the [*explanandum*] phenomenon".

As such, while philosophy may start with a folk conception of the phenomenon, its end goal is to create a theory that is informed by and continuous with the empirical sciences (following (P1) from *the exclusion argument*). Such theory will make ontological commitments based on our best scientific theory available and may be entirely different from our common sense understanding of the phenomenon. This is because, in a naturalistic methodology, accordance with folk concepts and intuitions is not a parameter to decide which theories we should endorse. If the preceding account is correct, then naturalistic philosophy of memory may not take common sense conceptions of memory and imagination as relevant parameters. Do such conceptions include mental attitudes? This is the topic of the next section.

### 4.3.2 P2. Folk psychology and mental attitudes

Folk psychology is ubiquitous in everyday life. If, say, Alina desires to have ice-cream and believes that there is some leftover in the fridge, we are thereby allowed to infer that her going to the kitchen had the intention of getting ice-cream. We make such inferences and predictions fairly often and, most of the time, we are successful in doing so. We connect overt behavior with certain mental states ("believing that there is ice-cream in the fridge") in such a way that the latter is predictive of the former. Folk psychology, then, is this practice of giving explanations of

other people's behavior by reference to their internal states (FODOR, 1987).

The mental states cited by folk psychology are composed of two parts: a content and an attitude. In the traditional representational theory of mind, a mental content is a proposition, usually preceded by a "that-clause", towards which a subject bears some relation (such as believing, desiring, remembering, and so on). These relations are called "mental attitudes"[2]. Discussion on mental attitudes spans various domains across philosophy, ranging from the philosophy of mind, epistemology, philosophy of language, among others (see SCHROEDER, 2006, for a review). Although there is little agreement on what is the nature of mental attitudes and which role should they fulfill in our scientific understanding of the mind, philosophers typically agree that mental attitudes, among with other postulates of our folk psychology, are reliable for understanding and predicting behavior in ordinary contexts.

For example, suppose two people, Alina and Alice, are watching the classical play *Medea*. Let's imagine that Alina possess regular folk psychological concepts, while Alice does not. From this, we can infer that Alina would fare much better than Alice in understanding the plot - given that she is more skilled in associating the character's behavior with a particular set of beliefs, desires, and so on. Only Alina is able to comprehend that Medea poisoning her husband is a consequence of her belief that he betrayed her, together with her anger and desire not to be betrayed. The causal link between the husband's betrayal and Medea's murder can only be traced by Alina, since it requires reference to Medea's internal states that cause her behavior. Alice, on the other hand, will see Medea poisoning her husband as an unmotivated behavior.

The example above illustrates a crucial feature of folk psychology and mental attitudes: they are highly effective for navigating ordinary contexts. With respect to predicting and understanding other's behavior, assuming that they possess internal states yields reliable predictions and, hence, facilitates interpersonal communication. This fact is further highlighted when considering how children diagnosed with Autism Spectrum Disorder fare considerably worse in socializing than their peers. Since they typically posses some deficits in theory of mind, they cannot understand why some people act the way they do, which makes communication all the more difficult (KIMHI, 2014; ROSELLO et al., 2020). These deficits in folk psychological skill, as it were, illustrate how much folk psychology is important for interaction with other people.

Here it is important to disambiguate different uses of the term "mental attitudes" in folk and technical contexts. While folk ascriptions of mental states may not explicitly conceive of these states as mental attitudes, they nevertheless carry important characteristics which are picked out by technical uses of the term. For example, folk psychology marks a clear distinction between *believing* and *desiring*: this much is evident from the intuitive difference there is between

[2] In the literature on the philosophy of mind, it is more traditional to use the term "propositional attitudes". I chose to substitute the term "propositional" for "mental" since the latter better encompasses the view that mental content can have forms of presentation other than propositional. This is important because, given the the fact that episodic memory often has imagistic content (see, e.g., TERONI, 2018; ROBINS, 2020), it is better to choose a term that reflects the multiple forms of content that memory can have.

“Alina believes that $p$" and “Alina desires that $p$". But the intuitive and folk distinction between these sentences is explicated by the technical conception of beliefs and desires having different directions of fit (ANSCOMBE, 1963). As such, folk psychology may not explicitly use the term “mental attitudes", but it employs the concept in accordance with the technical understanding of these attitudes.

In this context, “remembering" and “imagining" are to be understood as mental attitudes from the ordinary way we understand and predict the behavior of other people — as stated in P2 from the *exclusion argument* (section 4.3). For instance, to describe someone as “remembering that $p$" entails assigning to them the possible belief that $p$, as well as taking $p$ as having been learned in the past. While this form of description can be useful in ordinary contexts, it alone may not be sufficient for a proper scientific theory of memory and imagination. Whether this is the case is the topic of the next section.

### 4.3.3 P3 & P4. Folk terms in scientific psychology

While the previous subsection may not have been enough to cover the vast range of discussions regarding mental attitudes, it explored a very important point for the purposes of this paper: that attitudes and other folk mental posits are primarily useful for daily communication and coordination of behavior. If this is so, a pressing question then arises as to what is the relation between folk psychology and more scientific ways of understanding the mind. In the same way that the folk understanding of tomatoes as vegetables, not fruits, is not particularly useful for botany, the folk conception of remembering and imagining may fall short of being incorporated into a complete and robust account of cognition. If there isn’t a comprehensive framework that articulates the proper explanatory role of attitude terms to cognitive science, such terms will have to be eliminated from the former.

P3 in the *exclusion argument* highlights one straightforward way of mental attitudes being included in cognitive science experiments: researchers use them to direct participants’ behavior in remembering and imagining. This brings us back to Robins’ (2020) defense of mental attitudes in naturalistic philosophy of memory. She argues that the mental attitudes of “remembering” and “imagining” have to be taken seriously in a naturalistic philosophy of memory given that these terms are used in the cognitive sciences’ research program. Particularly in neuroimaging experiments, Robins notes that scientists instruct participants’ behavior using attitude terms, thus implying that they are incorporated into that particular research program. Further, this incorporation consists in the use of attitude terms in experimental design — i.e., the studies rely on the ability of participants to distinguish between remembering and imagining states (ROBINS, 2020, p. 15).

However, the use of a concept in experimental design does not guarantee that it will be present in the final theory. As the examples in Bechtel and Richardson (2010) illustrate, a particular concept or understanding of a phenomenon, characteristic of initial stages of research

when experiments are designed, are likely to be radically changed once a full account is developed. Thus, in the present framework, there is no guarantee that folk terms will be relevantly incorporated into an elaborate theory of cognition. This lack of any robust epistemic role for mental attitudes connects P3 and P4 in the exclusion argument. The latter premise stems from considering that the use of attitude terms in the experiments discussed is not epistemically relevant for the purposes of cognitive science and, hence, should not be incorporated in naturalistic philosophy.

As such, it is not clear whether the cases cited by Robins (2020) exemplify a genuine importance of attitudes to naturalistic philosophy of memory. Attitude terms, such as "remembering" and "imagining", would only be relevant for naturalistic (dis)continuism if their use is actually relevant for explanation in the cognitive sciences, and not just as tools to coordinate the behavior of research participants. Without a more detailed account of how attitude terms can be fruitfully incorporated into the program of cognitive science, they run the risk of being eliminated from the latter. Developing such account is the goal of the next section.

## 4.4 AN INTERPRETIVIST ACCOUNT AGAINST THE EXCLUSION ARGUMENT

In this section, I show that the exclusion argument can be blocked in two different but related ways. The first of which, here named *the compatibility argument* consists in arguing that the consequent of P3 does not follow from the antecedent - i.e., just because mental attitudes are terms used primarily in ordinary contexts, it does not necessarily follow that they don't have any significant epistemic role in the sciences of the mind. While denying P3 is sufficient to conclude that C is false, it is not sufficient to claim that the contrary of C is true, i.e., that naturalistic philosophy of memory should take mental attitudes into account. To argue for such inclusion, I present the *interpretivist argument*. Given that both arguments deal with mental attitudes at a high level of generality, the conclusions drawn from them can only be tentative. Nevertheless, they provide the necessary groundwork for defending important points in the philosophy of memory, as presented in section 4.5.

### 4.4.1 The compatibility argument

The compatibility argument is an attempt to block (C) by denying (P3) - i.e., it argues that there is no incompatibility between folk psychology and cognitive science in general. In the philosophical literature on mental attitudes, there two general ways for defending such compatibility. The first way is *intentional realism*, according to which the attitudes posited by folk psychology can be identified with real processes in the mind/brain. As such, there would be no incompatibility between folk psychology and the cognitive sciences. The second way is the *intentional stance theory*, according to which folk psychology is just a way of describing certain patterns in thought and behavior. As such, folk psychology is not a theory in any strict sense and it makes no claims about the mechanisms that underlie these patterns. Such account allows it

and mental attitudes to be, in principle, compatible with the cognitive sciences. I analyze each of these theories in turn.

#### 4.4.1.1 Intentional realism

It is tempting to think that the predictive success of folk psychology entails that the entities and processes postulated by it are vindicated by reference to actual cognitive processes in the brain. This position, here called *intentional realism*, is famously championed by Fodor (1987). He argues that the best explanation for why folk psychology and attitude-ascriptions are so successful is that they refer to discrete and real cognitive processes. While the inner workings of such processes still have to be uncovered by scientific psychology, Fodor maintains that our folk understanding already points us to the correct direction. In particular, Fodor (1981) argues how sentences that express attitude-ascriptions (e.g., "Alina believes that the actress who plays Medea won an Oscar") have a structure that indicates a relation (e.g., "believes that") between a subject (Alina) and a mental representation ("the actress who plays Medea won an Oscar"). Based on such linguistic analysis, Fodor highlights how we can infer "Alina believes something" and "there is something that Alina believes", thereby supporting his ontological commitment to mental attitudes and mental representations (see DE BRIGARD, 2015, for more discussion).

Supposing that Fodor is correct with regards to the existence of mental attitudes, it entails that *the exclusion argument* is not sound. The claim that the attitudes of "remembering" and "imagining" directly refer to states in our mental economy is sufficient for a naturalistic philosopher to taken them seriously. Moreover, if intentional realism were true, it would be senseless for cognitive scientists to dismiss attitude talk as mere common sense, thus also denying that folk psychology can be safely eliminated in a scientific context.

However, there are reasons to remain skeptical of Fodor's arguments for intentional realism. Firstly, his linguistic analysis of mental attitude expressions does not apply for all mental states that, supposedly, employ attitudes like beliefs and desires (BEN-YAMI, 1997). Secondly, given the wide range of attitudes we attribute to ourselves and other people, intentional realism would have to be committed to the existence of an extremely large collection of neurocognitive systems. It is unclear whether admitting the existence of such a large variety of attitudes would yield significant explanatory benefits. In fact, if there is a less permissive and equally, or even more, viable alternative to intentional realism, Occam's razor would certainly prefer the former.

Thirdly, intentional realism is in contrast with most recent developments in the cognitive sciences. As was first highlighted by Price and Friston (2005), most of our psychological terms do not have a strict relation with neural structure. This finding lead to the current discussion on creating a new taxonomy of psychological processes, commonly know as "cognitive ontology", that is better informed by what we currently know about the brain (POLDRACK, 2006; ANDERSON, 2015; DEWHURST, 2021). While the jury is still out on how we can accomplish this goal, it is sufficient for present purposes to note how, if even *scientific* psychological concepts do

not map very well with the brain, then *folk* psychological terms would not fare any better. In fact, it is perfectly plausible that the postulates of folk psychology are incompatible in light of cognitive neuroscience (CHURCHLAND, 1981, 1989). In short, if we grant that folk psychology and mental attitudes are supposed to refer to actual processes in the mind/brain, we assume the likely risk of having to replace these concepts in light of their incompatibility with findings in neuroscience.

#### 4.4.1.2 Intentional stance theory

Standing in between Fodor's (1987, 1992) intentional realism and Churchland's (1989, 1981) eliminative materialism, Dennett (1988, 2009) argues that mental attitudes do not directly refer to discrete processes in the mind-brain. His argument to that effect is that attitudes, along with other folk psychological postulations, are particular ways of speaking about a system that instantiates a certain pattern of behavior. No strong ontological commitments about underlying mechanisms are necessary to interpret people as if they believe, desire, or remember - instead, what is necessary is to view them from the intentional stance.

In Dennett's theory, the intentional stance is a way of describing and predicting the behavior of a given system based on the assumption that it is a rational agent. For example, suppose I installed a digital security system in my house. The system is composed of a camera next to my front door, who is connected to a computer which processes the data that come from the camera. If the computer detects someone in front of my house at night, it will trigger an alarm. One night, I wake up to the alarm and check the computer to see who is outside. Much to my surprise, only the neighbor's dog is outside, not so gently asking to come in.

From the intentional stance, my home's security system is described as *seeing* some suspicious movement and, thereby, forming the *belief* that there was a person knocking on the front door, which *motivates* the system to trigger an alarm *in order to* let me know that there is someone outside. The highlighted terms indicate conceptions from folk psychology, which enable us to make useful approximations of the security system as a rational agent. Such approximations are what make the intentional stance an effective framework for explaining behavior, since they abstract away the physical and design particularities of a given system. Following Dennett, the fact that a system works *as if* it were an intentional system is sufficient for it to *be* an intentional system.

Despite the apparent reasonableness of Dennett's theory, it still does not give us an answer for *why* the intentional stance is so effective. While the previous discussion indicates that the intentional stance might be a particularly useful form of speaking, it doesn't exclude the possibility that these explanations are just lucky, or that they only reflect a human tendency to draw inferences about internal states when, in fact, there aren't any. Indeed, Dennett addresses this difficulty in *Real Patterns* (1991), where he argues that intentional descriptions refer to general patterns of behavior. While different instances of such patterns may differ in detail, what

matters for the intentional stance is the fact that the pattern is repeatable and, therefore, we are capable of making reliable predictions based on it. As such, while mental attitudes do not directly refer to discrete processes in the mind/brain, they indicate a certain regularity of behavior that can only be detected when abstracting away from specific details.

This account is sufficient to argue for compatibility between, and the relevant epistemic role of, mental attitudes to the cognitive sciences, consequently denying P3 in the exclusion argument (section 4.3). Given that mental attitudes refer to patterns of behavior, which are more coarse-grained descriptions in comparison to cognitive explanations, there may not be any incompatibility between folk and scientific explanations of the mind. In this framework, they are just different ways of referring to the same phenomena at different levels of abstraction. However, to say that there are compatible *in principle* does not entail that they are *actually* so, nor that there are any significant explanatory advantages in describing cognitive states with mental attitudes. In particular, the claim that attitude terms such as "remembering" and "imagining" refer to general patterns of behavior does not, by itself, indicate any particular relevance of mental attitudes to the cognitive science of memory and imagination. In the next subsection, I argue that mental attitudes, understood as terms that refer to patterns of behavior, are important for our scientific understanding of the mind: they serve as *top-down constraints* on our search for neurocognitive mechanisms (cf. BECHTEL; RICHARDSON, 2010).

### 4.4.2 The interpretivist argument

Before exposing the interpretivist account, I should first analyze how cognitive science explains mental phenomena. In the contemporary philosophical literature on this matter, it is standard to view cognitive science as uncovering the *mechanisms* that underlie cognition. According to a traditional view, mechanisms are "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (MACHAMER; DARDEN; CRAVER, 2000, p. 3). The overall behavior of the mechanism, delimited by such starting and ending conditions, is also called "phenomenon" (GLENNAN, 2017; BECHTEL; ABRAHAMSEN, 2005).

In this characterization, mechanisms are always systems *for* a certain phenomenon (DARDEN, 2008). As such, whenever we are trying to uncover a mechanism for a certain phenomenon (say, $\psi$), we should begin with an adequate description of what $\psi$ is, which effects it has, and what are the typical environmental elements that allow $\psi$ to take place. Once such description of $\psi$ is made, the process of discovering the mechanisms that underlie it begins with the detection of certain elements whenever $\psi$ occurs. After thorough empirical analysis and manipulation of such elements, we figure out that there is a set of elements (say, $\phi$) which are consistently active whenever $\psi$ is also active[3]. Then we are closer to the claim that the mechanism for $\psi$ is nothing more than the coordinated functioning of the elements in $\phi$ (CRAVER, 2015).

[3] See Craver (2007), chapter 4, and Baumgartner and Casini (2017) for more discussion.

Naturally, the story presented above is vastly simplified. Still, it highlights how mechanism discovery is highly dependent on what we take the higher-order phenomenon to be. For example, if we want to discover what is the mechanism behind how a pocket calculator divides two numbers, we first have to understand how division works. Then we take such description of division and compose a list of steps a system must take in order to divide two numbers. Such list is, naturally, constrained by what we take division to be, but it also constrains which mechanisms are candidates for division. All possible systems that are not capable of implementing our list of steps are automatically ruled out from our investigation. This is what is commonly called the *top-down constraint* (BECHTEL; RICHARDSON, 2010):

> *Top-down constraint:* we start with a description of a system, $S$, and its behavior $\psi$-ing. These characterizations allow us to formulate "how-possibly" models of how $S$ $\psi$-ies. These models are only implementable on compatible hardware, thereby limiting the range of possible mechanisms that are able to fulfill $\psi$.

Importantly, top-down constraints are always dependent on a research project and its particularities. The conceptual framework of researchers, the empirical tools available to them, as well as their methods for mathematical modelling, all play a decisive role in constraining their research question and the range of hypotheses that can be formulated and tested (KÄSTNER; HAUEIS, 2021). In this context, constraining the research question consists in delimiting our characterization of $S$ and its behavior to $\psi$, which, in turn, bounds which mechanisms for $\psi$ can be discovered.

Kästner and Haueis (2021) explicate this relation between research methods, phenomenon characterization, and mechanistic discovery as three elements of a larger *pattern recognition practice*. Their use of concept "pattern" is directly influenced by the works of Dennett (1991) and Haugeland (1998). In this context, a given collection forms a pattern if, and only if, its individual elements are organized in a given arrangement, such that it can be recognized from a higher level of abstraction that captures the collection of elements as a whole, instead of focusing on individual elements. As such, a pattern is both (1) an orderly arrangement of elements, and (2) a candidate for recognition (HAUGELAND, 1998, pp. 273-274). Kästner and Haueis apply this conception to understand how ontic (i.e., bottom-up) and epistemic (i.e., top-down) norms constrain mechanistic inquiry:

> On the one hand, scientific practice with its methods and tools epistemically constrains what patterns in the causal structure of the world can be recognized as mechanisms. On the other hand, patterns in the causal structure ontically constrain which scientific tools will serve to recognize them as orderly arrangements persistent from below and salient from above, respectively (KÄSTNER; HAUEIS, 2021, p. 1645).

The way in which top-down constraints are related to practices of pattern recognition supports the thesis that mental attitudes cannot be excluded from cognitive science, nor from

naturalistic philosophy. Attitude-terms provide the initial and rough formulation of a certain mental state. After that, we start looking for the patterns of behavior associated with that mental state, usually via a set of experimental tasks that require the employment of the cognitive capacity in question (FRANCKEN; SLORS, 2014; KÄSTNER; HAUEIS, 2021). Then we use whichever empirical methods are more adequate to uncover the mechanism (or set of mechanisms) that correlate with the behavioral pattern. While it is likely that there may be multiple possible mechanisms that can underlie the pattern associated with a mental attitude, what matters is that they are all grouped together under the heading of a certain mental attitude term precisely because they produce the pattern of overt behavior associated with that term.

In sum, and in agreement with Robins (2020), there is there is no incompatibility between describing a mental state using folk psychology and using terms of the cognitive sciences. Quite the contrary, mental attitudes offer a way to refer to general patterns of thought and behavior in such a way that they constrain our mechanistic inquiry of particular cognitive capacities. Given that pattern recognition is inherently perspectival, ascribing mental attitudes and uncovering their mechanisms are also perspective-dependent tasks. Such is the thesis of *interpretivism*: people have beliefs and desires in virtue of the fact that their behavior conforms to a pattern associated with beliefs, desires, and so on. Interpretivism, so constructed, entails that ascribing mental attitudes consists in identifying patterns from a given perspective. This perspective serves as a top-down constraint on mechanistic inquiry, thus granting an important epistemic role for the mental attitudes in the empirical sciences of the mind. Therefore, the conditional of P3 in the exclusion argument (section 4.3) is false: the fact that mental attitudes come from folk psychological talk does not entail that they serve no purpose in cognitive science. As such, interpretivism agrees with Robins' (2020) general conclusion and expands her account by saying more on what are mental attitudes, how they can be individuated, and how they serve an important epistemic constraint in the investigations of cognitive science.

### 4.4.3 Possible objections

In light of interpretivism and the pattern view of mental attitudes, one could argue that such account is not sufficient for preventing the exclusion of mental attitudes from the empirical sciences of the mind. Even if interpretivism is true, the objection goes, it could still be the case that a complete theory of neurocognitive mechanisms does not mention mental attitudes in any relevant sense, thus eliminating them from a true theory of the mind.

Against this objection, I highlight how mental attitudes, taken as patterns of thought and behavior, serve an epistemic function in the sciences of the mind that cannot be eliminated by a complete description of neurocognitive mechanisms. As mentioned previously, patterns are only detectable once we take a more distant perspective from the *explanandum* phenomenon and, thereby, abstract away details from the functioning of its parts. By definition, we can only detect patterns in coarser-grained descriptions. Such vantage point brings several different epistemic

advantages that are not present in finer-grained descriptions of mechanisms. Computational tractability, for example, is only achieved once we consider only the overall behavior of a system and, thereby, disregard the particularities of each individual part (see, e.g., CHIRIMUUTA, 2014). The same is true with regards to mental attitudes: they highlight important patterns that may otherwise be missed in a finer-grained mechanistic description. Just like we cannot infer what a computer program is doing solely by looking at its string of binary numbers, being too fine-grained with our description of cognitive states leads us to miss the bigger picture.

Another objection concerns the status of mental attitudes inside the framework of the cognitive sciences. Given that interpretivism takes these attitudes as patterns of thought and behavior, one might question whether scientists should treat them as real entities, or only as useful terms that abstract away some of the nuances of particular mental states. Against this (supposed) ambiguity, I argue that scientists should, and often do, treat patterns as real entities. The metaphysics of patterns allows for them to have properties and causal powers that are not present in the constituent parts of the system. For example, biologists recognize how forests, understood as closed ecological systems, have properties that are not present in any individual organism that makes up the forest. In this instance, biologists treat the forest as a real entity independent of any individual organism. My contention with the interpretivist view is that, similar to the case in biology, cognitive scientists should, and often do, take mental attitudes into account as entities that emerge from neurocognitive processes, and not just as useful concepts for abstraction purposes.

## 4.5 INTERPRETIVISM IN PHILOSOPHY OF MEMORY

Let's take stock. The main issue that sparkled this debate was the apparent incompatibility between a naturalistic methodology for philosophy of memory and its recent interest in taking mental attitudes seriously. Such apparent incompatibility is highlighted in *the exclusion argument*. In short, the argument claims that, since mental attitudes are matters of folk psychology and are not intended for explanations of neurocognitive mechanisms, they are neither useful for explanations in cognitive science nor relevant to naturalistic philosophy of memory.

In the previous section, I argued that the premises of the exclusion argument are not sufficient to support its conclusion. I defended that the folk origins of mental attitudes do not exclude the possibility that they are nevertheless necessary for uncovering the mechanisms of cognition. Following interpretivism, mental attitudes are ways of referring to the phenomena that brain mechanisms are responsible for. In this context, correct attitude ascriptions are correlated with relevant neurocognitive systems, despite the fact that such relation between mental attitudes and systems are many-to-many and involve multiple in-between translation steps.

The interpretivist thesis applies to naturalistic philosophy of mind in general. It concludes that such branch of philosophy must take into account the relevant mental attitudes during philosophical investigation. While such broad claims may be sufficient to claim that philosophy

of memory should take the "remembering" attitude into account, they fall short of providing specific insight into the problems that philosophers of memory face today. To fill this gap is the aim of the next subsections. Here I focus on the distinction between epistemic and empirical meanings of the term "remembering" (section 4.5.1), and on (dis)continuism (section 4.5.2) about the processes of memory and imagination[4].

### 4.5.1 Epistemic and empirical remembering

Related to the (dis)continuism debate, philosophers have recently taken interest on the multiple meanings and uses of the term "remembering". According to Craver (2020), *epistemic remembering* is a set of commitments about the accuracy and reliability of a subject's memory that allow it to be categorized as a ground for knowledge about the past. As such, epistemic remembering refers to the epistemic responsibility of claiming to remember: if someone claims to (epistemically) remember that *p*, then we have the right to hold that person's memory up to scrutiny. In the case that her mental state fails to meet the criteria of being a reliable and accurate source about the past, then we can say that she was not, in fact, remembering. Meanwhile, *empirical remembering* refers to people's actual, and sometimes faulty, memories and the mechanisms that support it. The empirical sense of "remembering", following Craver, is the performance of a cognitive system in following the commitments of the epistemic sense.

A similar account, though different in specific and important aspects, is developed by McCarrol, Michaelian, and Nanay (2022), who refer to *normative* and *descriptive* perspectives on episodic memory. Their theory differs from Craver's with regards to accounting for the relation between epistemic remembering (or the normative perspective) and empirical remembering (or the descriptive perspective). Craver (2020, p. 267) claims that there is, in principle, no incompatibility between epistemic remembering and empirical remembering, since the norms of the former are neutral when it comes to the mechanistic details of the latter. His claim is based on the fact that, given the epistemic sense of remembering refers to a certain speech act and the epistemic commitments associated with it, and that the empirical meaning concerns the mechanisms that underlie states of picturing a past event, there is no incompatibility in virtue of them being different terms for different objects. Meanwhile, McCarroll, Michaelian, and Nanay (2022, p. 22) view the normative and descriptive perspectives as referring to one and the same object, namely "memory", thus they cannot be used simultaneously on the same context. This is because it is likely that the normative and descriptive perspectives ask different questions about memory and, hence, will have different theories on how it works. These theories and perspectives will be compatible only in so far as they are applied in different contexts.

Interpretivism provides a middle ground for the discussion on whether epistemic and empirical meanings of remembering refer to the same or separate objects. To elucidate my

[4] Naturally, these is not the only problem about which philosophers of memory currently debate. However, due to space constraints, these are the problems I'll be focusing on the most in this paper.

point, I should first discuss the phenomenon of *misremembering*. Following Robins (2016), a paradigmatic example of misremembering is the Deese-Roediger-McDermott (DRM) effect, in which participants consistently report remembering related but absent words from a previous list (DEESE, 1959; MCDERMOTT; ROEDIGER, 1998). For example, in a list containing *nurse, disease, treatment, diagnosis, vaccine, medicine, hospital*, people tend to incorrectly remember that semantically related words (such as *doctor*) were present in the list. Such experiments are paradigmatic examples of misremembering because they involve partial retention of previous information, but still result in a false recollection (ROBINS, 2016, p. 434). To the extent that epistemic remembering involves the commitment to the reliability of one's memory, and to the extent that the notion of reliability allows for some occasional failures (MICHAELIAN, 2016b), misremembering is a case in which the subject is committed to accuracy and reliability of her memory, despite the fact her memory is false. This indicates that the commitments of epistemic remembering are not entirely in the memory's content, but instead is how we think about that content. In short, the commitments of epistemic remembering are *attitudes* taken towards a memory content.

Taking epistemic remembering as a mental attitude is further warranted in other ways. Particularly, the fact that we sometimes are doubtful as to whether we are epistemically remembering indicates that there are edge cases in which the notion of epistemic remembering is not sufficient to clear cut the boundaries of correct memories. This is because, following interpretivism for mental attitudes, epistemic remembering refers to a pattern of how we think about and create rules over our memories. This pattern, learned on our parents' knee (NELSON; FIVUSH, 2004), is a tendency towards considering our memories as reliable and accurate sources of information about the past, provided we actually experienced the past event or learned about it from reliable testimony (see also COSMIDES; TOOBY, 2000). The notion of pattern is useful here since it allows for degrees of confidence and reliability in our epistemic claims to remember.

Moreover, taking epistemic remembering as pattern of commitments and considerations about one's memory provides an interesting result on how it relates to empirical remembering. As mentioned previously, Craver (2020) defends that they are mostly independent of each other, since they refer to different objects. In contrast, McCarrol, Michaelian, and Nanay (2022) argue that the epistemic and empirical views refer to the same object, and that they are compatible only in so far as they are applied on different contexts.

The interpretivist view, defended in the previous section, is able to dissolve this discussion. Consider again one of the basic tenants of the mechanistic literature: a phenomenon is the behavior of the mechanism (GLENNAN, 2017). As such, the behavior of the mechanism is a manifestation of the coordinated functioning of its component parts (BECHTEL; ABRAHAMSEN, 2005). In this sense, phenomena are dependent on their mechanisms. In another sense, one cannot eliminate the phenomena from a description of the overall system, since talking in a higher-level of abstraction allows us to pick up regularities that we would have missed otherwise

(DENNETT, 1991; KÄSTNER; HAUEIS, 2021). In short, the phenomenon is *metaphysically* dependent on the mechanism, but not *epistemically* so.

Considering epistemic remembering as a mental attitude and, following interpretivism, as a pattern of behavior, it follows that it is epistemically, but not metaphysically, independent on the underlying mechanisms of the empirical view. Epistemic remembering is taken as a pattern that emerges from and, hence, is dependent of the proper functioning of memory mechanisms. In this metaphysical sense, the epistemic view is emergent from the coordinated functioning of elements in the empirical view. However, in the epistemic sense, any description of epistemic remembering is independent of the empirical view, since it stands in a different level of abstraction.

In short, I agree with Craver on the (epistemic) independence of the two meanings of remembering. At the same time, I concur with McCarroll, Michaelian, and Nanay on the intuition that both senses of remembering refer to the same object: it seems that way given the metaphysical dependence of epistemic sense to the empirical sense. That is not to say, however, that one can reduce the former to the latter: the relation of emergence only holds when elements of a mechanism work together in a very specific way, so much so that the overall behavior of the system cannot be reduced to the behavior of a single component.

### 4.5.2 An interpretivist view on (dis)continuism

The interpretivist thesis also has important implications as to whether episodic memory and imagination are continuous or discontinuous. There are several issues at stake here, for there is more than one way of asking for the relation between remembering and imagining. Firstly, there is the question of whether they belong to the same natural kind (WERNING, 2020; ANDONOVSKI, 2018). Secondly, and relatedly, one might ask how (dis)similar are the mechanisms that underlie episodic memory and imagination (PERRIN, 2016; MICHAELIAN, 2016a). Thirdly, one could ask whether the attitudes of remembering and imagining are (dis)continuous (SANT'ANNA, 2021; LANGLAND-HASSAN, 2022).

All of these issues can be approached within the interpretivist framework. Starting with the issue of mental attitudes, interpretivism considers them as overall patterns of behavior. Taking patterns as both a non-random organization of elements and as candidates for recognition (cf. HAUGELAND, 1998), it indicates that patterns are only detectable and, hence, only make sense within a given perspective. In this regard, demarcating mental attitudes of remembering and imagining is a function of which framework we are working on and what characteristics we deem relevant for individuating them. If we focus on how both remembering and imagining are judgments towards mental imagery, then continuism follows. Instead, if we focus on how the attitude of remembering has particular connotations on the accuracy of its content, then discontinuism holds. Whether the former or the latter are correct is, ultimately, making a choice on which perspective is most appropriate for a particular research project.

On the question of whether episodic memory and imagination belong to the same natural

kind, most philosophers tend to adopt the homeostatic property cluster theory of kinds (HPC, for short), made famous by Boyd (1991). This theory binds the natural kinds question with the issue about mechanisms: i.e., the natural kinds of memory and imagination are determined in virtue of their mechanisms. If they are the same, then continuism is true; if they are not, discontinuism follows. Considering how marking the boundaries of mechanisms is top-down constrained on our characterization of the *explanandum* phenomenon (see section 4.4.2), it stands to reason that delineating the mechanisms of memory and imagination will also be so constrained. Moreover, given that such constraint is, by definition, dependent on a research project, any answer (dis)continuism will also particular to the research project and interests at hand.

As De Brigard (2018) claims, to correlate a mental state with a neurocognitive system involves an act of interpretation. Using the concept of patterns, we can see how whichever patterns we can encounter with such interpretation is dependent on the perspective and research project at hand. In this regard, to determine whether episodic memory is (dis)continuous with imagination is, ultimately, a matter of perspective. There is no straightforward fact of the matter on the relation between remembering and imagining. Instead, it depends on what we want to explain. Hence, interpretivism entails *pluralism* about (dis)continuism.

This account advances the diagnosis proposed by McCarroll, Michaelian, and Nanay (2022) on the debate between the causal and simulation theories of memory. According to them, each theory may be best understood from a particular perspective: the causal theory is the better option from a normative perspective, and the simulation theory, from a descriptive perspective. Interpretivism expands their account by allowing for a greater range of possible perspectives that, not only take different stances on causalism *versus* simulationism, but also yield different results on the (dis)continuity between memory and imagination. It could be the case, for example, that researchers in a particular perspective from cognitive neuroscience have more reasons to adopt continuism, given their interest in neurocognitive mechanisms and so on (e.g., ADDIS, 2020), but, from a computational point of view, memory necessitates processes that are not applicable to imagination, thus leading us to adopt discontinuism. McCarroll et al. would probably take neuroscientific and computational frameworks as belonging to the descriptive perspective, but, given how they produce different outcomes for (dis)continuism, I maintain that they should be kept separated.

## 4.6 SUMMARY AND CONCLUSIONS

The (dis)continuism problem asks if episodic memory is continuous with imagination. Given its close proximity with the cognitive sciences, philosophers have traditionally taken this issue as part of a larger naturalistic framework of philosophy of memory. Recently, Robins (2020) argued that such naturalistic methodology necessitates the need for philosophers to also take the attitudes of remembering and imagining into account. However, given that mental attitudes are

folk psychological constructs and, hence, are not intended for scientific explanations of the mind, one could argue that it is unlikely that "remembering" and "imagining" have any significant role in naturalistic philosophy. Such is the *exclusion argument*, which was the main motivation for this paper.

The present article discussed a way of circumventing the exclusion argument and more precisely argue for what roles should mental attitudes have in naturalistic philosophy of memory. Robins herself responds to a similar argument against mental attitudes in naturalistic philosophy, but her account still lacks a more precise characterization of how these attitudes can be individuated, and which role they serve in cognitive science. In this paper, I defended a framework that is able to fill these gaps, called *interpretivism*. It claims that mental attitudes refer to patterns of thought and behavior which, in turn, are supported by cognitive mechanisms. Given that mechanisms are partially demarcated by what they do, there is a top-down constraint between the way in which we characterize the overall pattern, with mental attitudes, and how we go about uncovering mechanisms *for* that pattern. As such, interpretivism highlights how mental attitudes function as guides to mechanistic discovery and, thereby, defends the consideration of "remembering" and "imagining" attitudes into naturalistic philosophy of memory.

Moreover, interpretivism provides a novel account for the relation between epistemic remembering and empirical remembering. As Craver (2020) claims, epistemic remembering consists in the set of norms and implications of claiming to remember. Empirical remembering, on the other hand, refers to the mechanisms active in our day-to-day memories, which are often far from the fulfilment of the epistemic norms of remembering. While Craver defends that these two meanings of remembering refer to different objects, McCarroll, Michaelian, and Nanay (2022) argue that they have only one referent — i.e., memory. Following interpretivism, and taking epistemic remembering as a mental attitude, I explicated how the epistemic view is metaphysically, but not epistemically, dependent on the empirical view. As such, I agree with Craver on their (epistemic) independence, but also concur with McCarroll's et al. intuition that these meanings refer to the same object.

# 5 CONCLUSIONS

In this dissertation, I attempted to explicate the (dis)continuism problem in the context of a mechanistic framework. As discussed in chapter 1, one of the problems of approaching the (dis)continuism problem in the philosophy of memory is how it lacks a precise formulation. When we ask if remembering is imagining, should we formulate the problem as a question about identity (i.e., "is episodic memory identical to imagination?"), about natural kinds and mechanisms (i.e., "do episodic memory and imagination share the same mechanism and natural kind?"), or about the attitudes involved in each mental state (i.e., "what is the relation between the attitudes of remembering and imagining?")? As a consequence of attempting to answer this question via a mechanistic framework, I argued that there is no theory or perspective independent way of answering (dis)continuism. Instead, any proposed relation between episodic memory and imagination should be understood inside a theoretical context that delimits where are the boundaries of these mechanisms. In short, we should be pluralists about (dis)continuism.

Chapter 1 introduced the relevant senses of episodic memory and imagination, as well as the general issues that surround these topics. In that chapter, I also argued that (dis)continuism should not be taken as a question about identity, since this formulation undermines the richness of the current debate in the philosophy of memory.

Chapter 2 argued that the perspectival considerations on mechanistic inquiry do not necessarily entail that antirealism is the case, nor does it pose any problems for the objectivity of the mechanistic enterprise. In this article, submitted at *Filosofia Unisinos*, I claimed that there are empirical bottom-up constraints in mechanistic inquiry, which safeguard, to some extent, the antirealist and subjective objections. Considering how the search for mechanisms is so multiply constrained, I draw some conclusions on the general metaphysics of mechanisms, and how it forms a type of realism that admits a degree of perspectival considerations.

Chapter 3 consisted in a paper, submitted for publication at *Philosophy and the Mind Sciences*, that investigates the (dis)continuism problem as it pertains to the natural kinds of episodic memory and imagination. In this paper, I defended a view called *metaphysical pluralism*, according to which delineating natural kinds and their associated mechanisms are highly dependent on the pragmatic interests of a given research project. Moreover, more than one of such projects are likely to be pick out real differences and relations in nature, thus ontologically and empirically grounding their claims on natural kinds and mechanisms. Therefore, whether episodic memory and imagination belong to the same natural kind depends on the pragmatic and research interests at play in a given conceptual and epistemic context.

Chapter 4 reached a similar conclusion to the previous one, but via an analysis of how mechanistic inquiry is conducted in the cognitive sciences. The paper, submitted at *Synthese*,

approached the issue of whether a naturalist philosopher of memory should take the folk attitudes of "remembering" and "imagining" into account when discussing the (dis)continuism problem. This issue gains more traction under the perspective that, if methodological naturalism compels us to make ontological committments only to those entities present in our best scientific theories, then it should not matter what is the folk conception of the phenomena in which we are interested. Against this line of thinking, I argued that propositional attitudes are serve a crucial epistemic role of picking out patterns and contraining, in a top-down way, our mechanistic inquiry into episodic memory and imagination. Morevoer, given that picking out patterns is inherently a perspective-laden activity, uncovering the mechanisms of cognition, memory and imagination included, is partially and inevitably a project dependent on a given perspective and research project. Chapter 4 thus concludes that interpretivism about propositional attitudes is true, and that pluralism is the best alternative for (dis)continuism.

In sum, the overall argument of this dissertation consists in a simple universal instantiation: all projects of delienating mechanisms are perspective-laden; (dis)continuism is about delineating the mechanisms of episodic memory and imagination; therefore, (dis)continuism is perspective laden. From this conclusion, I defend we should be pluralists on whether memory is (dis)continuos to imagination. This opens several new venues of research — for instance, we can investigate which thoeretical virtues are the most relevant for (dis)continuism, or even how perspectives from different areas of congitive science (such as computer science, or neurobiology) can shed new light on the relation between memory and imagination. Morevoer, the overall mechanistic framework presented can be of significant import for questions about other cognitive processes, such as the relation between perception and hallucinations, the nature of mental disorders, and so on. As such, these considerations are crucial for our understanding of the processes that, ultimately, make up ourselves and our subjective time.

# REFERENCES


ADDIS, Donna Rose. Mental Time Travel? A Neurocognitive Model of Event Simulation. **Review of Philosophy and Psychology**, Springer, v. 11, n. 2, p. 233–259, 2020. DOI: 10.1007/s13164-020-00470-0. Cit. on pp. 37, 66.

ANDERSON, Michael L. **After Phrenology: Neural Reuse and the Interactive Brain**. Cambridge, Massachusetts: The MIT Press, 2014. 385 pp. Cit. on p. 28.

_____. Mining the Brain for a New Taxonomy of the Mind. **Philosophy Compass**, v. 10, n. 1, p. 68–77, 2015. DOI: 10.1111/phc3.12155. Visited on: 5 Nov. 2022. Cit. on pp. 36, 43, 57.

ANDONOVSKI, Nikola. Is Episodic Memory a Natural Kind? **Essays in Philosophy**, v. 19, n. 2, p. 1609, 2018. DOI: 10.7710/1526-0569.1609. Visited on: 15 Feb. 2023. Cit. on pp. 36, 41, 42, 45–48, 65.

ANSCOMBE, G.E.M. **Intention**. Ithaca, NY: Cornell University Press, 1963. Cit. on p. 55.

ARONOWITZ, Sara. Memory Is a Modeling System. **Mind and Language**, v. 34, n. 4, p. 483–502, 4 2019. DOI: 10.1111/mila.12220. Visited on: 2 Jan. 2023. Cit. on pp. 14, 47.

ATANCE, Cristina M.; O'NEILL, Daniela K. Episodic Future Thinking. **Trends in Cognitive Sciences**, v. 5, n. 12, p. 533–539, 2001. DOI: 10.1016/S1364-6613(00)01804-0. Visited on: 27 Apr. 2023. Cit. on p. 15.

ATANCE, Cristina M.; SOMMERVILLE, Jessica A. Assessing the Role of Memory in Preschoolers' Performance on Episodic Foresight Tasks. **Memory**, Routledge, v. 22, n. 1, p. 118–128, 2014. DOI: 10.1080/09658211.2013.820324. pmid: 23889532. Visited on: 13 May 2023. Cit. on p. 50.

BAUMGARTNER, Michael; CASINI, Lorenzo. An Abductive Theory of Constitution. **Philosophy of Science**, v. 84, n. 2, p. 214–233, 2017. DOI: 10.1086/690716. Visited on: 2 Dec. 2022. Cit. on pp. 21, 59.

BECHTEL, William. **Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience**. New York: Routledge, 2008. 308 pp. Cit. on p. 29.

BECHTEL, William; ABRAHAMSEN, Adele. Explanation: A Mechanist Alternative. **Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 36, n. 2, p. 421–441, 2005. DOI: 10.1016/j.shpsc.2005.03.010. Visited on: 20 Jan. 2023. Cit. on pp. 20, 59, 64.

BECHTEL, William; RICHARDSON, Robert C. **Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research**. MIT Press ed. Cambridge, Mass: MIT Press, 2010. 286 pp. Cit. on pp. 18, 19, 21, 25, 26, 33, 51, 53, 55, 59, 60.

BEN-YAMI, Hanoch. Against Characterizing Mental States as Propositional Attitudes. **The Philosophical Quarterly**, v. 47, n. 186, p. 84–89, 1997. DOI: 10.1111/1467-9213.00049. Visited on: 16 Dec. 2022. Cit. on p. 57.

BERNECKER, Sven. **Memory: A Philosophical Study**. [S.l.]: Oxford University Press, 2010. Cit. on p. 50.

BIDERMAN, Natalie; SHOHAMY, Daphna. Memory and Decision Making Interact to Shape the Value of Unchosen Options. **Nature Communications**, v. 12, n. 1, p. 4648, 2021. DOI: 10.1038/s41467-021-24907-x. Visited on: 30 Mar. 2023. Cit. on pp. 19, 30.

BIRD, Alexander; TOBIN, Emma. Natural Kinds, 2022. Available from: <https://plato.stanford.edu/archives/spr2022/entries/natural-kinds/>. Visited on: 27 Feb. 2023. Cit. on p. 35.

BOYD, Richard. Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. **Philosophical Studies**, v. 61, n. 1-2, p. 127–148, 1991. DOI: 10.1007/BF00385837. Visited on: 15 Feb. 2023. Cit. on pp. 35, 36, 39, 43, 66.

_____. Homeostasis, Species and Higher Taxa. In: WILSON, Robert A. (Ed.). **Species: New Interdisciplinary Essays**. Cambridge, Mass: MIT Press, 1999. Cit. on p. 39.

_____. Rethinking Natural Kinds, Reference and Truth: Towards More Correspondence with Reality, Not Less. **Synthese**, v. 198, S12, p. 2863–2903, 2021. DOI: 10.1007/s11229-019-02138-4. Visited on: 15 Feb. 2023. Cit. on pp. 39, 48.

BOYLE, Alexandria. The Mnemonic Functions of Episodic Memory. **Philosophical Psychology**, v. 35, n. 3, p. 327–349, 2021. DOI: 10.1080/09515089.2021.1980520. Visited on: 25 Feb. 2023. Cit. on p. 46.

_____. Episodic Memory in Animals: Optimism, Kind Scepticism and Pluralism. In: SANT'ANNA, André; MCCARROLL, Christopher Jude; MICHAELIAN, Kourken (Eds.). **Current Controversies in Philosophy of Memory**. Abingdon, Oxon: Routledge, 2022. Cit. on p. 47.

BRITANNICA, The Editors of Encyclopaedia. **Tomato**. In: ENCYCLOPAEDIA Britannica. [S.l.: s.n.], 2023. Available from: <https://www.britannica.com/plant/tomato>. Visited on: 23 Apr. 2023. Cit. on p. 17.

CARNAP, Rudolf. **Der logische Aufbau der Welt**. Hamburg: Felix Meiner Verlag, 2017. Cit. on p. 19.

CHAPPELL, Sophie-Grace. Aristotle. In: BERNECKER, Sven; MICHAELIAN, Kourken (Eds.). **The Routledge Handbook of Philosophy of Memory**. New York, NY: Routledge, 2017. Cit. on p. 13.

CHENG, Sen. The CRISP Theory of Hippocampal Function in Episodic Memory. **Frontiers in Neural Circuits**, v. 7, 2013. DOI: 10.3389/fncir.2013.00088. Visited on: 18 Jan. 2023. Cit. on p. 46.

CHENG, Sen; WERNING, Markus. What Is Episodic Memory If It Is a Natural Kind? **Synthese**, v. 193, n. 5, p. 1345–1385, 2016. DOI: 10.1007/s11229-014-0628-6. Visited on: 21 Feb. 2023. Cit. on pp. 36–38, 41, 45, 46.

CHIRIMUUTA, M. Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience. **Synthese**, v. 191, n. 2, p. 127–153, 2014. DOI: 10.1007/s11229-013-0369-y. Visited on: 20 Mar. 2023. Cit. on p. 62.

CHOMSKY, Noam. Problems and Mysteries in the Study of Human Language. In: KASHER, Asa (Ed.). **Language in Focus: Foundations, Methods and Systems**. Red. by Robert S. Cohen and Marx W. Wartofsky. Dordrecht: Springer Netherlands, 1976. v. 43. P. 281–357. DOI: 10.1007/978-94-010-1876-0_15. Visited on: 1 May 2023. Cit. on p. 12.

CHURCHLAND, Patricia Smith; SEJNOWSKI, Terrence J. **The Computational Brain**. Cambridge, Mass.: MIT Press, 1992. 544 pp. (Computational Neuroscience). Cit. on p. 31.

CHURCHLAND, Paul M. Eliminative Materialism and the Propositional Attitudes. **The Journal of Philosophy**, v. 78, n. 2, p. 67, 1981. DOI: 10.2307/2025900. Visited on: 4 May 2023. Cit. on p. 58.

_____. Folk Psychology and the Explanation of Human Behavior. **Philosophical Perspectives**, v. 3, p. 18, 1989. Available from: <https://www.jstor.org/stable/2214269>. Cit. on pp. 51, 58.

COLOMBO, Matteo; PICCININI, Gualtiero. **The Computational Theory of Mind**. Cambridge, UK: Cambridge University Press, Forthcoming. 57 pp. Cit. on p. 31.

COSMIDES, Leda; TOOBY, John. Consider the Source: The Evolution of Adaptations for Decoupling and Metarepresentation. In: SPERBER, Dan (Ed.). **Metarepresentations: A Multidisciplinary Perspective**. [S.l.]: Oxford University Press, 2000. P. 63. Cit. on p. 64.

CRAVER, Carl F. **Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience**. New York: Oxford University Press, 2007. 308 pp. Cit. on pp. 29, 59.

_____. Mechanisms and Natural Kinds. **Philosophical Psychology**, v. 22, n. 5, p. 575–594, 2009. DOI: 10.1080/09515080903238930. Visited on: 27 Feb. 2023. Cit. on pp. 40–42, 48.

_____. Functions and Mechanisms: A Perspectivalist View. In: HUNEMAN, Philippe (Ed.). **Functions: Selection and Mechanisms**. Dordrecht ; New York: Springer, 2013. (Synthese Library, volume 363). Cit. on pp. 20, 21.

_____. Levels. In: METZINGER, Thomas; WINDT, Jennifer M. (Eds.). **Open MIND**. [S.l.]: MIND Group, 2015. Available from: <http://www.open-mind.net/DOI?isbn=9783958570498>. Visited on: 6 Sept. 2022. Cit. on pp. 21, 26, 59.

_____. Remembering: Epistemic and Empirical. **Review of Philosophy and Psychology**, Springer Science and Business Media LLC, v. 11, n. 2, p. 261–281, 2020. DOI: 10.1007/s13164-020-00469-7. Cit. on pp. 52, 63, 64, 67.

CRAVER, Carl F.; DARDEN, Lindley. **In Search of Mechanisms: Discoveries across the Life Sciences**. Chicago ; London: The University of Chicago Press, 2013. 228 pp. Cit. on p. 20.

CUMMINS, Robert. **The Nature of Psychological Explanation**. Cambridge, Mass.: MIT Pr, 1985. 219 pp. (A Bradford Book). Cit. on pp. 19, 21.

D'ARGEMBEAU, Arnaud; VAN DER LINDEN, Martial. Individual Differences in the Phenomenology of Mental Time Travel: The Effect of Vivid Visual Imagery and Emotion Regulation Strategies. **Consciousness and Cognition**, v. 15, n. 2, p. 342–350, 2006. DOI: 10.1016/j.concog.2005.09.001. Visited on: 27 Feb. 2023. Cit. on pp. 46, 50.

DARDEN, Lindley. Thinking Again about Biological Mechanisms. **Philosophy of Science**, v. 75, n. 5, p. 958–969, 2008. DOI: 10.1086/594538. Visited on: 27 Feb. 2023. Cit. on pp. 18, 21, 42, 59.

DE BRIGARD, F. et al. Remembering What Could Have Happened: Neural Correlates of Episodic Counterfactual Thinking. **Neuropsychologia**, v. 51, n. 12, p. 2401–2414, 2013. DOI: 10.1016/j.neuropsychologia.2013.01.015. Visited on: 2 Nov. 2022. Cit. on p. 50.

DE BRIGARD, Felipe. Is Memory for Remembering? Recollection as a Form of Episodic Hypothetical Thinking. **Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science**, Springer Science and Business Media LLC, v. 191, n. 2, p. 155–185, 2013. DOI: 10.1007/s11229-013-0247-7. Cit. on p. 50.

_____. The Nature of Memory Traces. **Philosophy Compass**, v. 9, n. 6, p. 402–414, 2014. DOI: 10.1111/phc3.12133. Visited on: 20 Feb. 2023. Cit. on pp. 35, 37.

_____. What Was I Thinking? Dennett's Content and Consciousness and the Reality of Propositional Attitudes. In: MUÑOZ-SUÁREZ, Carlos; DE BRIGARD, Felipe (Eds.). **Content and Consciousness Revisited: With Replies by Daniel Dennett**. Cham: Springer, 2015. (Studies in Brain and Mind, 7). Cit. on pp. 38, 57.

_____. Cognitive Systems and the Changing Brain. **Philosophical Explorations**, v. 20, n. 2, p. 224–241, 2017. DOI: 10.1080/13869795.2017.1312503. Visited on: 16 Dec. 2022. Cit. on pp. 36, 43.

_____. Memory and the Intentional Stance. In: HUEBNER, Bryce; DENNETT, D. C. (Eds.). **The Philosophy of Daniel Dennett**. New York, NY: Oxford University Press, 2018. Cit. on p. 66.

DEESE, James. Influence of Inter-Item Associative Strength upon Immediate Free Recall. **Psychological Reports**, v. 5, n. 3, p. 305–312, 1959. DOI: 10.2466/pr0.1959.5.3.305. Visited on: 2 Jan. 2023. Cit. on p. 64.

DENNETT, Daniel C. Précis of The Intentional Stance. **Behavioral and Brain Sciences**, v. 11, n. 03, p. 495, 1988. DOI: 10.1017/S0140525X00058611. Visited on: 16 Dec. 2022. Cit. on p. 58.

DENNETT, Daniel C. Real Patterns. **The Journal of Philosophy**, v. 88, n. 1, p. 27, 1991. DOI: 10.2307/2027085. Visited on: 1 Oct. 2022. Cit. on pp. 23, 24, 27, 33, 51, 58, 60, 65.

_____. Intentional Systems Theory. In: BECKERMANN, Ansgar; MCLAUGHLIN, Brian P.; WALTER, Sven (Eds.). **The Oxford Handbook of Philosophy of Mind**. 1. ed. [S.l.]: Oxford University Press, 2009. P. 339–350. DOI: 10.1093/oxfordhb/9780199262618.003.0020. Visited on: 18 Oct. 2022. Cit. on p. 58.

DEWHURST, Joe. Computing Mechanisms Without Proper Functions. **Minds and Machines**, v. 28, n. 3, p. 569–588, 2018. DOI: 10.1007/s11023-018-9474-5. Visited on: 19 Mar. 2023. Cit. on pp. 19, 22.

_____. Folk Psychological and Neurocognitive Ontologies. In: CALZAVARINI, Fabrizio; VIOLA, Marco (Eds.). **Neural Mechanisms**. Cham: Springer International Publishing, 2021. v. 17. P. 311–334. DOI: 10.1007/978-3-030-54092-0_14. Visited on: 1 Nov. 2022. Cit. on p. 57.

DUPRE, John. **The Disorder of Things: Metaphysical Foundations of the Disunity of Science**. 1. paperback ed., Nachdr. Cambridge, Mass.: Harvard Univ. Press, 1996. 308 pp. Cit. on pp. 32, 44.

ELDER, Crawford L. Conventionalism and Realism-Imitating Counterfactuals. **The Philosophical Quarterly**, v. 56, n. 222, p. 1–15, 2006. DOI: 10.1111/j.1467-9213.2005.00425.x. Visited on: 23 Feb. 2023. Cit. on p. 39.

_____. Biological Species Are Natural Kinds. **The Southern Journal of Philosophy**, v. 46, n. 3, p. 339–362, 2008. DOI: 10.1111/j.2041-6962.2008.tb00123.x. Visited on: 23 Feb. 2023. Cit. on p. 35.

FAZEKAS, Peter. Pursuing Natural Piety: Understanding Ontological Emergence and Distinguishing It from Physicalism. **Dialectica**, v. 68, n. 1, p. 97–119, 2014. DOI: 10.1111/1746-8361.12056. Visited on: 8 Apr. 2023. Cit. on p. 27.

FODOR, Jerry A. Propositional Attitudes. In: BLOCK, Ned (Ed.). **The Language and Thought Series**. Cambridge, MA and London, England: Harvard University Press, 1981. DOI: 10.4159/harvard.9780674594722.c5. Visited on: 16 Dec. 2022. Cit. on p. 57.

_____. **Psychosemantics: The Problem of Meaning in the Philosophy of Mind**. Cambridge, Mass.: MIT Press, 1987. Cit. on pp. 54, 57, 58.

_____. **A Theory of Content and Other Essays**. 4. print. Cambridge, Mass.: MIT Press, 1992. 270 pp. (A Bradford Book). Cit. on pp. 38, 58.

FOUCAULT, Michel. **The Order of Things**. [S.l.]: Routledge, 2005–1966. DOI: 10.4324/9780203996645. Visited on: 27 Feb. 2023. Cit. on p. 35.

FRANCKEN, Jolien C.; SLORS, Marc. From Commonsense to Science, and Back: The Use of Cognitive Concepts in Neuroscience. **Consciousness and Cognition**, v. 29, p. 248–258, 2014. DOI: 10.1016/j.concog.2014.08.019. Visited on: 28 Nov. 2022. Cit. on pp. 51, 61.

GLENNAN, Stuart. **The New Mechanical Philosophy**. First edition. Oxford: Oxford University Press, 2017. 266 pp. Cit. on pp. 18, 21, 22, 25, 31, 32, 41, 42, 59, 64.

HALINA, Marta. Mechanistic Explanation and Its Limits. In: GLENNAN, Stuart; ILLARI, Phyllis McKay (Eds.). **The Routledge Handbook of Mechanisms and Mechanical Philosophy**. London ; New York: Routledge, Taylor & Francis Group, 2018. (Routledge Handbooks in Philosophy). P. 213–237. Cit. on p. 29.

HARBECKE, Jens; SHAGRIR, Oron. The Role of the Environment in Computational Explanations. **European Journal for Philosophy of Science**, v. 9, n. 3, p. 37, 2019. DOI: 10.1007/s13194-019-0263-7. Visited on: 10 Sept. 2022. Cit. on p. 22.

HASSABIS, Demis et al. Patients with Hippocampal Amnesia Cannot Imagine New Experiences. **Proceedings of the National Academy of Sciences**, v. 104, n. 5, p. 1726–1731, 2007. DOI: 10.1073/pnas.0610561104. Visited on: 27 Apr. 2023. Cit. on p. 15.

HAUGELAND, J. Pattern and Being. In: HAVING Thought. Essays in the Metaphysics of Mind. Cambridge, Mass: Harvard University Press, 1998. P. 267–290. Cit. on pp. 24, 27, 33, 51, 60, 65.

HEMPEL, Carl Gustav. **Aspects of Scientific Explanation and Other Essays in the Philosophy of Science**. New York, NY: The Free Press, 1965. Cit. on p. 28.

ILLARI, Phyllis. Mechanistic Explanation: Integrating the Ontic and Epistemic. **Erkenntnis**, v. 78, S2, p. 237–255, 2013. DOI: 10.1007/s10670-013-9511-y. Visited on: 8 Apr. 2023. Cit. on p. 29.

KAISER, Marie I.; KRICKEL, Beate. The Metaphysics of Constitutive Mechanistic Phenomena. **The British Journal for the Philosophy of Science**, The University of Chicago Press, v. 68, n. 3, p. 745–779, 2017. DOI: 10.1093/bjps/axv058. Visited on: 23 Nov. 2022. Cit. on p. 21.

KAPLAN, David Michael; CRAVER, Carl F. The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. **Philosophy of Science**, Cambridge University Press, v. 78, n. 4, p. 601–627, 2011. DOI: 10.1086/661755. Visited on: 20 Mar. 2023. Cit. on p. 22.

KÄSTNER, Lena. Integration and the Mechanistic Triad: Producing, Underlying and Maintaining Mechanistic Explanations. In: CALZAVARINI, Fabrizio; VIOLA, Marco (Eds.). **Neural Mechanisms**. Cham: Springer International Publishing, 2021. v. 17. P. 337–361. DOI: 10.1007/978-3-030-54092-0_15. Visited on: 21 Apr. 2023. Cit. on pp. 20, 21.

KÄSTNER, Lena; HAUEIS, Philipp. Discovering Patterns: On the Norms of Mechanistic Inquiry. **Erkenntnis**, v. 86, n. 6, p. 1635–1660, 2021. DOI: 10.1007/s10670-019-00174-7. Visited on: 22 Oct. 2022. Cit. on pp. 22, 24, 29, 30, 33, 51, 60, 61, 65.

KEELEY, Brian. Natural Mind. In: CLARK, Kelly James (Ed.). **The Blackwell Companion to Naturalism**. 1° edition. Hoboken: Wiley Blackwell, 2016. P. 196–208. Cit. on p. 50.

KIMHI, Yael. Theory of Mind Abilities and Deficits in Autism Spectrum Disorders. **Topics in Language Disorders**, v. 34, n. 4, p. 329, 2014. DOI: 10.1097/TLD.0000000000000033. Visited on: 4 May 2023. Cit. on p. 54.

KITCHER, Philip. Species. **Philosophy of Science**, v. 51, n. 2, p. 308–333, 1984. DOI: 10.1086/289182. Visited on: 27 Feb. 2023. Cit. on p. 44.

KORNBLITH, Hilary. **Knowledge and Its Place in Nature**. Oxford : New York: Clarendon Press ; Oxford University Press, 2002. 189 pp. Cit. on pp. 50, 52, 53.

_____. A Naturalistic Methodology. In: OVERGAARD, Søren (Ed.). **The Cambridge Companion to Philosophical Methodology**. Cambridge, UK: Cambridge University Press, 2017. (Cambridge Companions). Cit. on p. 52.

LADYMAN, James et al. **Every Thing Must Go: Metaphysics Naturalized**. Oxford ; New York: Oxford University Press, 2007. 346 pp. Cit. on p. 32.

LANGLAND-HASSAN, Peter. **Explaining Imagination**. First edition. New York: Oxford university press, 2020. 313 pp. Cit. on p. 16.

_____. What Sort of Imagining Might Remembering Be? **Journal of the American Philosophical Association**, Cambridge University Press, v. 7, n. 2, p. 231–251, 2021. DOI: 10.1017/apa.2020.28. Visited on: 27 Apr. 2023. Cit. on pp. 16, 17.

_____. Remembering and Imagining: The Attitudinal Continuity. In: BERNINGER, Anja; VENDRELL FERRAN, Íngrid (Eds.). **Philosophical Perspectives on Memory and Imagination**. New York, NY: Routledge, 2022. (Routledge Studies in Contemporary Philosophy). Cit. on pp. 38, 65.

LEE, Jonny; DEWHURST, Joe. The Mechanistic Stance. **European Journal for Philosophy of Science**, v. 11, n. 1, p. 20, 2021. DOI: 10.1007/s13194-020-00341-6. Visited on: 26 Oct. 2022. Cit. on p. 24.

MACHAMER, Peter; DARDEN, Lindley; CRAVER, Carl F. Thinking about Mechanisms. **Philosophy of Science**, v. 67, n. 1, p. 1–25, 2000. DOI: 10.1086/392759. Visited on: 27 Feb. 2023. Cit. on pp. 18, 20, 25, 31, 33, 41–43, 59.

MAHR, Johannes B. Episodic Memory: And What Is It For? In: SANT'ANNA, André; MCCARROLL, Christopher Jude; MICHAELIAN, Kourken (Eds.). **Current Controversies in Philosophy of Memory**. Abingdon, Oxon: Routledge, 2022. Cit. on p. 47.

MAHR, Johannes B.; CSIBRA, Gergely. Why Do We Remember? The Communicative Function of Episodic Memory. **Behavioral and Brain Sciences**, v. 41, e1, 2018. DOI: 10.1017/S0140525X17000012. Visited on: 22 Sept. 2022. Cit. on p. 47.

MARR, David. **Vision: A Computational Investigation into the Human Representation and Processing of Visual Information**. Cambridge, Mass: MIT Press, 1981–2010. 403 pp. Cit. on pp. 22, 31, 38.

MASSIMI, Michela. **Perspectival Realism**. Oxford, New York: Oxford University Press, 2022. 416 pp. (Oxford Studies in Philosophy of Science). Cit. on p. 32.

MCCARROLL, Christopher Jude; MICHAELIAN, Kourken; NANAY, Bence. Explanatory Contextualism about Episodic Memory: Towards A Diagnosis of the Causalist-Simulationist Debate. **Erkenntnis**, 2022. DOI: 10.1007/s10670-022-00629-4. Visited on: 9 Nov. 2022. Cit. on pp. 52, 63, 64, 66, 67.

MCCULLOCH, Warren S.; PITTS, Walter. A Logical Calculus of the Ideas Immanent in Nervous Activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943. DOI: 10.1007/BF02478259. Visited on: 17 Sept. 2022. Cit. on p. 31.

MCDERMOTT, Kathleen B.; ROEDIGER, Henry L. Attempting to Avoid Illusory Memories: Robust False Recognition of Associates Persists under Conditions of Explicit Warnings and Immediate Testing. **Journal of Memory and Language**, v. 39, n. 3, p. 508–520, 1998. DOI: 10.1006/jmla.1998.2582. Visited on: 2 Jan. 2023. Cit. on p. 64.

MCLELLAND, Victoria; SCHACTER, Daniel L; ADDIS, Donna Rose. Contributions of Episodic Memory to Imagining the Future. In: ADDIS, Donna Rose; BARENSE, Morgan; DUARTE, Audrey (Eds.). **The Wiley Handbook on The Cognitive Neuroscience of Memory**. [S.l.]: Wiley Blackwell, 2015. P. 287–308. Cit. on pp. 37, 50.

MICHAELIAN, Kourken. Is Memory a Natural Kind? **Memory Studies**, v. 4, n. 2, p. 170–189, 2011. DOI: 10.1177/1750698010374287. Visited on: 22 Feb. 2023. Cit. on p. 38.

_____. Against Discontinuism: Mental Time Travel and Our Knowledge of Past and Future Events. In: MICHAELIAN, Kourken; KLEIN, Stanley B.; SZPUNAR, Karl K. (Eds.). **Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel**. New York, NY: Oxford University Press, 2016. Cit. on pp. 50, 65.

_____. **Mental Time Travel: Episodic Memory and Our Knowledge of the Personal Past**. [S.l.]: MIT Press, 2016. Cit. on pp. 15, 35, 37, 46, 49, 50, 64.

_____. Imagining the Past Reliably and Unreliably: Towards a Virtue Theory of Memory. **Synthese**, v. 199, n. 3-4, p. 7477–7507, 2021. DOI: 10.1007/s11229-021-03125-4. Visited on: 15 Aug. 2022. Cit. on p. 37.

MICHAELIAN, Kourken; PERRIN, Denis. Mental Time Travel. In: SVEN BERNECKER; KOURKEN MICHAELIAN (Eds.). **The Routledge Handbook of Philosophy of Memory**. London ; New York: Routledge, Taylor & Francis Group, 2017. (Routledge Handbooks in Philosophy). Cit. on pp. 15, 35, 49, 50.

MICHAELIAN, Kourken; PERRIN, Denis, et al. Mental Time Travel. In: GLĂVEANU, Vlad (Ed.). **The Palgrave Encyclopedia of the Possible**. Cham: Springer International Publishing, 2023. P. 1–8. DOI: 10.1007/978-3-319-98390-5_222-1. Visited on: 7 Nov. 2022. Cit. on pp. 38, 48.

MILLIKAN, Ruth Garrett. **Language, Thought, and Other Biological Categories: New Foundations for Realism**. Cambridge, Mass: MIT Press, 1984. 355 pp. Cit. on p. 21.

NATHAN, Marco J. **Pluralism Is the Answer! What Is the Question?** Philosophy, Theory, and Practice in Biology. 2018. Available from: <https://doi.org/10.3998/ptpbio.16039257.0011.015>. Visited on: 1 Feb. 2023. Cit. on pp. 36, 44.

NELSON, Katherine; FIVUSH, Robyn. The Emergence of Autobiographical Memory: A Social Cultural Developmental Theory. **Psychological Review**, v. 111, n. 2, p. 486–511, 2004. DOI: 10.1037/0033-295X.111.2.486. Visited on: 13 May 2023. Cit. on p. 64.

OKUDA, Jiro et al. Thinking of the Future and Past: The Roles of the Frontal Pole and the Medial Temporal Lobes. **NeuroImage**, v. 19, n. 4, p. 1369–1380, 2003. DOI: 10.1016/S1053-8119(03)00179-4. Visited on: 27 Apr. 2023. Cit. on p. 15.

OPPENHEIM, P.; PUTNAM, H. The Unity of Science as a Working Hypothesis. In: FEIGL, H.; MAXWELL, G. (Eds.). **Concepts, Theories and the Mind-Body Problem**. Minneapolis: University of Minnesota Press, 1958. (Minnesota Studies in the Philosophy of Science, v. 2). P. 3–36. Cit. on p. 19.

PENFIELD, Wilder; MILNER, Brenda. Memory Deficit Produced by Bilateral Lesions in the Hippocampal Zone. **A.M.A. Archives of Neurology & Psychiatry**, American Medical Association, v. 79, n. 5, p. 475–497, 1958. DOI: 10.1001/archneurpsyc.1958.02340050003001. Visited on: 26 Apr. 2023. Cit. on p. 13.

PERRIN, Denis. Asymmetries in Subjective Time. In: MICHAELIAN, Kourken; KLEIN, Stanley B.; SZPUNAR, Karl K. (Eds.). **Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel**. New York, NY: Oxford University Press, 2016. Cit. on pp. 37, 50, 65.

PESSOA, Luiz. **The Entangled Brain: How Perception, Cognition, and Emotion Are Woven Together**. Cambridge, Massachusetts: The MIT Press, 2022. Cit. on pp. 28, 36, 43.

PICCININI, Gualtiero. **Neurocognitive Mechanisms: Explaining Biological Cognition**. First edition. Oxford: Oxford University Press, 2020. 400 pp. Cit. on p. 31.

PICCININI, Gualtiero; CRAVER, Carl. Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches. **Synthese**, v. 183, n. 3, p. 283–311, 2011. DOI: 10.1007/s11229-011-9898-4. Visited on: 21 Apr. 2023. Cit. on p. 31.

POLDRACK, R. Can Cognitive Processes Be Inferred from Neuroimaging Data? **Trends in Cognitive Sciences**, v. 10, n. 2, p. 59–63, 2006. DOI: 10.1016/j.tics.2005.12.004. Visited on: 16 Dec. 2022. Cit. on p. 57.

POTOCHNIK, Angela; OLIVEIRA, Guilherme. Patterns in Cognitive Phenomena and Pluralism of Explanatory Styles. **Topics in Cognitive Science**, v. 12, n. 4, p. 1306–1320, 2020. DOI: 10.1111/tops.12481. Visited on: 14 Dec. 2022. Cit. on p. 24.

POVICH, Mark; CRAVER, Carl. Mechanistic Levels, Reduction, and Emergence. In: GLENNAN, Stuart; ILLARI, Phyllis McKay (Eds.). **The Routledge Handbook of Mechanisms and Mechanical Philosophy**. London ; New York: Routledge, Taylor & Francis Group, 2018. (Routledge Handbooks in Philosophy). Cit. on pp. 26, 27.

PRICE, Cathy J.; FRISTON, Karl J. Functional Ontologies for Cognition: The Systematic Definition of Structure and Function. **Cognitive Neuropsychology**, Routledge, v. 22, n. 3-4, p. 262–275, 2005. DOI: 10.1080/02643290442000095. Visited on: 16 Dec. 2022. Cit. on p. 57.

QUINE, W. V. Epistemology Naturalized. In: GIBSON, Roger F. (Ed.). **Quintessence: Basic Readings from the Philosophy of W.V. Quine**. Cambridge, Mass: Belknap Press of Harvard University Press, 2004. P. 259–274. Cit. on p. 53.

ROBINS, Sarah K. Misremembering. **Philosophical Psychology**, Informa UK Limited, v. 29, n. 3, p. 432–447, 2016. DOI: 10.1080/09515089.2015.1113245. Cit. on p. 64.

_____. Memory Traces. In: BERNECKER, Sven; MICHAELIAN, Kourken (Eds.). **The Routledge Handbook of Philosophy of Memory**. London ; New York: Routledge, Taylor & Francis Group, 2017. (Routledge Handbooks in Philosophy). Cit. on p. 45.

_____. Defending Discontinuism, Naturally. **Review of Philosophy and Psychology**, Springer, v. 11, n. 2, p. 469–486, 2020. Cit. on pp. 37, 49, 50, 54–56, 61, 66.

ROSELLO, Belen et al. Theory of Mind Profiles in Children With Autism Spectrum Disorder: Adaptive/Social Skills and Pragmatic Competence. **Frontiers in Psychology**, v. 11, p. 567401, 2020. DOI: 10.3389/fpsyg.2020.567401. pmid: 33041932. Cit. on p. 54.

RUSSELL, Bertrand. **The Analysis of Mind**. [S.l.]: G. Allen & Unwin, 1922. Cit. on p. 13.

SALMON, Wesley C. Scientific Explanation: Three Basic Conceptions. **PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association**, v. 1984, n. 2, p. 293–305, 1984. DOI: 10.1086/psaprocbienmeetp.1984.2.192510. Visited on: 11 Apr. 2023. Cit. on p. 28.

_____. **Scientific Explanation and the Causal Structure of the World**. Princeton, N.J: Princeton University Press, 1984. 305 pp. Cit. on p. 19.

SANT'ANNA, André. Mental Time Travel and the Philosophy of Memory. **Filosofia Unisinos**, v. 19, n. 1, p. 52–62, 2018. DOI: 10.4013/fsu.2018.191.06. Visited on: 21 Feb. 2023. Cit. on p. 15.

_____. Attitudes and the (Dis)Continuity between Memory and Imagination. **Estudios de Filosofía**, n. 64, p. 73–93, 2021. DOI: 10.17533/udea.ef.n64a04. Visited on: 25 Oct. 2022. Cit. on pp. 38, 65.

SCHACTER, Daniel L; ADDIS, Donna Rose. The Cognitive Neuroscience of Constructive Memory: Remembering the Past and Imagining the Future. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 362, n. 1481, p. 773–786, 2007. DOI: 10.1098/rstb.2007.2087. Visited on: 7 Nov. 2022. Cit. on pp. 15, 37, 46, 47, 50.

SCHROEDER, Timothy. Propositional Attitudes. **Philosophy Compass**, v. 1, n. 1, p. 65–73, 2006. DOI: 10.1111/j.1747-9991.2006.00010.x. Visited on: 18 Oct. 2022. Cit. on p. 54.

SHAGRIR, Oron; BECHTEL, William. Marr's Computational Level and Delineating Phenomena. In: KAPLAN, David Michael (Ed.). **Explanation and Integration in Mind and Brain Science**. First edition. Oxford, United Kingdom: Oxford University Press, 2017. Cit. on pp. 19, 22, 31.

SQUIRE, Larry R. The Legacy of Patient H.M. for Neuroscience. **Neuron**, Elsevier, v. 61, n. 1, p. 6–9, 2009. DOI: 10.1016/j.neuron.2008.12.023. pmid: 19146808. Visited on: 26 Apr. 2023. Cit. on p. 13.

STICH, Stephen P. **From Folk Psychology to Cognitive Science: The Case against Belief**. 5th print. Cambridge, Mass.: MIT Press, 1985. 266 pp. (A Bradford Book). Cit. on pp. 21, 51.

SUDDENDORF, Thomas; CORBALLIS, Michael C. The Evolution of Foresight: What Is Mental Time Travel, and Is It Unique to Humans? **Behavioral and Brain Sciences**, v. 30, n. 3, p. 299–313, 2007. DOI: 10.1017/S0140525X07001975. Visited on: 27 Feb. 2023. Cit. on pp. 15, 47.

SUTTON, John. **Philosophy and Memory Traces: Descartes to Connectionism**. [S.l.]: Cambridge University Press, 1998. Cit. on p. 13.

TAYLOR, Henry. Emotions, Concepts and the Indeterminacy of Natural Kinds. **Synthese**, v. 197, n. 5, p. 2073–2093, 2020. DOI: 10.1007/s11229-018-1783-y. Visited on: 25 Feb. 2023. Cit. on p. 47.

_____. Whales, Fish and Alaskan Bears: Interest-Relative Taxonomy and Kind Pluralism in Biology. **Synthese**, v. 198, n. 4, p. 3369–3387, 2021. DOI: 10.1007/s11229-019-02284-9. Visited on: 25 Feb. 2023. Cit. on p. 47.

TERONI, Fabrice. On Seeming to Remember. In: MICHAELIAN, Kourken; DEBUS, Dorothea; PERRIN, Denis (Eds.). **New Directions in the Philosophy of Memory**. New York: Routledge, Taylor & Francis Group, 2018. (Routledge Studies in Contemporary Philosophy). P. 329–345. Cit. on p. 54.

THOMASSON, Amie L. What Can We Do, When We Do Metaphysics? In: D'ORO, Giuseppina; OVERGAARD, Søren (Eds.). **The Cambridge Companion to Philosophical Methodology**. 1. ed. [S.l.]: Cambridge University Press, 2017. P. 101–121. DOI: 10.1017/9781316344118.007. Visited on: 2 Dec. 2022. Cit. on p. 52.

TULVING, Endel. Episodic and Semantic Memory. In: TULVING, Endel; DONALDSON, W (Eds.). **Organization of Memory**. [S.l.]: Academic Press, 1972. Available from: <https://psycnet.apa.org/record/1973-08477-007>. Cit. on pp. 14, 50.

_____. Memory and Consciousness. **Canadian Psychology/Psychologie canadienne**, Canadian Psychological Association, v. 26, n. 1, p. 1, 1985. Cit. on pp. 14, 50.

_____. What Is Episodic Memory? **Current Directions in Psychological Science**, v. 2, n. 3, p. 67–70, 1993. DOI: 10.1111/1467-8721.ep10770899. Visited on: 21 Feb. 2023. Cit. on p. 37.

_____. Episodic Memory and Autonoesis: Uniquely Human? In: TERRACE, Herbert S.; METCALFE, Janet (Eds.). **The Missing Link in CognitionOrigins of Self-Reflective Consciousness**. [S.l.]: Oxford University Press, 2005. DOI: 10.1093/acprof:oso/9780195161564.001.0001. Visited on: 21 Feb. 2023. Cit. on pp. 15, 37.

VAN LEEUWEN, Neil. The Meanings of "Imagine" Part I: Constructive Imagination. **Philosophy Compass**, v. 8, n. 3, p. 220–230, 2013. DOI: 10.1111/j.1747-9991.2012.00508.x. Visited on: 27 Apr. 2023. Cit. on p. 15.

WERNING, Markus. Predicting the Past from Minimal Traces: Episodic Memory and Its Distinction from Imagination and Preservation. **Review of Philosophy and Psychology**, Springer Science and Business Media LLC, v. 11, n. 2, p. 301–333, 2020. DOI: 10.1007/s13164-020-00471-z. Cit. on pp. 35, 37, 38, 41, 45, 48, 65.

WOODWARD, James. A Theory of Singular Causal Explanation. **Erkenntnis (1975-)**, Springer, v. 21, n. 3, p. 231–262, 1984. DOI: 10.1007/BF00169275. Visited on: 30 Mar. 2023. Cit. on p. 19.

WRINCH, Dorothy. On the Nature of Memory. **Mind; a quarterly review of psychology and philosophy**, JSTOR, v. 29, n. 113, p. 46–61, 1920. Cit. on p. 13.