ARTICLE

The Formats of Cognitive Representation: A Computational Account

Dimitri Coelho Mollo[1] and Alfredo Vernazzani[2]

[1]Department for Historical, Philosophical, and Religious Studies, Umeå University, Umeå, Sweden and
[2]Institut für Philosophie II, Ruhr-Universität Bochum, Bochum, Germany
Corresponding authors: Dimitri Coelho Mollo; Email: dimitri.mollo@umu.se; Alfredo Vernazzani;
Email: alfredo-vernazzani@daad-alumni.de

Abstract

Cognitive representations are typically analyzed in terms of content, vehicle, and format. Although current work on formats appeals to intuitions about external representations, such as words and maps, in this article, we develop a computational view of formats that does not rely on intuitions. In our view, formats are individuated by the computational profiles of vehicles, that is, the set of constraints that fix the computational transformations vehicles can undergo. The resulting picture is strongly pluralistic, makes space for a variety of different formats, and is intimately tied to the computational approach to cognition in cognitive science and artificial intelligence.

1. Introduction

Representation is a central and, arguably, foundational notion in mainstream cognitive science and artificial intelligence (AI) (Burge 2010; Cummins 1989; Neander 2017; Shea 2018). Appealing to representations internal to biological and artificial systems provides us with tools to help explain the relational nature of cognition and intelligence: to be cognitive and intelligent is to behave in such a way as to protect and further the system's own interests, satisfying its needs and preserving its existence (and occasionally that of its group) in interaction with a complex, changing, and often hostile environment. The defining characteristic of representations is their aboutness, that is to say, the fact that representations are about something other than themselves. A map can be about the spatial layout of a region; a sentence can be about the current weather there. Similarly, internal representations are states and processes within biological and artificial systems that are about states, processes, and events beyond themselves, typically in the body and

---

Authors contributed equally.

the environment of the system. What representations are about or refer to are their contents (Shea 2018, 6).[1]

Although representations are primarily characterized by their contents—a representation of the location of my office, a representation of Ursula von der Leyen's face—representations can also be characterized in other terms, typically for somewhat different explanatory purposes. We may be interested in what kinds of physical states and processes carry, or possess, representational contents. And perhaps less obviously, we might be interested, roughly put, in the shape or format a representation takes: Is it a map, a photo, a sentence?

In this article, we will be interested in the latter feature of representations. What are representational formats? What are they good for? We will investigate such questions within cognitive science and AI research. Our exclusive focus will be on the representational states and processes going on in brain areas, layers in artificial neural networks, and the like, which are at the center of the explanatory and modeling endeavors in those fields.

We will advance an account of representational formats in which the main aim is that of capturing the epistemic roles that the notion plays, or can play, in the relevant areas of science and engineering by appeal to the notion of physical computation, that is, computation in physical systems (rather than in mathematical theory). Computational views of representational formats have a long history (Sloman 1978, Larkin and Simon 1987, Fodor 1975). However, such views were often left relatively underdeveloped and/or focused exclusively on specific kinds of format, with the linguistic/iconic distinction drawing most of the attention (Fodor 2008; Sloman 1978). The latter distinction is still among the most discussed (Quilty-Dunn 2019; Quilty-Dunn et al. 2023).[2]

This is unfortunate for at least two reasons. First, extant accounts of formats, including the ones inspired by the computational approach, have typically taken for granted intuitive views about formats modeled on external, public representations, such as words, pictures, and maps. It is debatable, to say the least, that categories applicable to public, external representations can or should be applied to capturing the goings-on in cognitive and computational systems. The focus on intuitive distinctions—such as linguistic/pictorial, analogue/digital—that have marked the literature are a symptom of this (typically implicit) assumption. Second, and relatedly, an account of representational formats should be general and thus able to capture all the formats that are relevant to cognitive (and computational) processing rather than being tailored only to account for a subset of formats.

In this article, we will try to free our understanding of representational formats from its intuitive chains. We will do so by developing a computational view of formats that takes as its starting point the explanatory needs of the cognitive sciences rather

---

[1] Traditionally, it has been preferred to take a nonreferential view of content, individuating contents as conditions of satisfaction instead, which in turn pick out referents. This difference will not matter for our purposes.

[2] The terminology in the debate is rather confusing. The iconic format is sometimes also called "depictive" (Kosslyn et al. 2006), "image-like," "picture-like," or "analogue" (Quilty-Dunn 2019; Beck 2018; Maley 2011; Paivio 1986; but see Clarke 2019 for a distinction between iconic and analogue). The discursive or symbolic format is also called "language-like" (Paivio 1986), "Fregean" (Sloman 1978), or "propositional" (Pylyshyn 1973).

than common intuitions. As a consequence, the resulting account yields formats ill-fitted to the categories traditionally employed in the literature while positing varieties of representational formats that have no analogue in external representations. The standard of success for a theory of representational formats for cognitive science is the epistemic value it has in informing and guiding research, not the extent to which the resulting formats fit our pre-theoretic expectations. The second part of the article will thus be dedicated to illustrating the epistemic value of the resulting computational theory of representational formats.

Here is how we will proceed. After presenting our distinctive perspective on the question of representational formats in section 2.1, we will briefly go through the main extant families of views about their nature, making clear where our own view belongs (sec. 2.2). In section 2.3, we will set out the central explanatory roles played by representational formats in the cognitive sciences, which, together with broader philosophical considerations, make up a set of desiderata for any account of representational formats for those fields. We present and defend the computational view of formats in section 3, whereas section 4 is dedicated to illustrating the account by applying it to two case studies: one from neuroscience (the place cell system) and one from computational modeling (episodic memory recall). Finally, in section 5, we show that the computational view fulfills the desiderata on theories of representational formats in the cognitive sciences.

2. Representational formats: Nature and roles

2.1. Three notions of representational format

Colored pieces of paper, binary code stored in a memory drive, and patterns of neural activation in the brain can all carry representational content: they can all be representations, say, of von der Leyen's face. As carriers of content, these internal states and processes are called *representational vehicles*.

Importantly, vehicles are individuated not purely in terms of their physical properties but rather in terms of those physical properties to which an interpreter or system is sensitive. In a paper map, the vehicles are printed shapes and colors, not the type of paper used; in an electronic computer, the vehicles are, ultimately, voltage ranges that code for 1s and 0s during specific time intervals, irrespective of the continuous values voltages take; in a brain, the vehicles are most likely some aspect(s) of neural activity, such as firing rates, but not neurons' color or smell. Often, different vehicles can carry the same content, thus representing the same thing, and different things can be represented by the same vehicles.

Qualifying the last sentence with an "often" may seem intuitive enough. It seems implausible, or at least very doubtful, that a photo of von der Leyen has the very same content as a verbal description of her facial features. And even if they do have the same content, they seem to represent it in very different ways. They also seem to be more appropriate for different uses: a photo will be better than a verbal description for recognizing von der Leyen in a crowd, whereas a verbal description will be better if we are interested in a specific, less noticeable feature.

It is not always clear how best to try to accommodate these considerations, especially when it comes to examples that rely on intuitions about external, public representations, such as photos, words, and maps. One common attempt is to rely on

the notion of representational format to shed light on those and related differences between representations (Beck 2018; Clarke 2019; Fodor 2007, 2008; Quilty-Dunn 2019). Photos and verbal descriptions, intuition suggests, belong to different formats, insofar as they represent different contents and/or represent them in different ways. Similar considerations apply to cognitive science and AI research. Some kinds of internal representations may have different constraints on what they can and cannot represent and/or on how well or efficiently they can represent what they do.

The general shape that an account of representational formats must take plausibly differs between different domains of application, such as cognitive science and AI, on the one hand, and external, public representations, on the other hand. Even within the former domain, it is likely that there are differences in terms of epistemic needs and tools when it comes to the states that the cognitive sciences discover and investigate and the states that populate our folk psychology. Failing to keep these two domains separate risks generating considerable confusion and unclarity.[3]

Given their importantly different features, it is to be expected that the expression "representational format" captures fundamentally different constructs in the two domains. Indeed, an account of the formats of external, public representations is highly likely to hinge, in complicated ways, on social practices and conventions for the production and consumption of representations, as well as on individuals' goals, intentions, and interpretative abilities. Moreover, in light of the tight connection between the social practices of communication and interpretation and the posits of folk psychology, it is likely that the notion of representational format relevant for folk psychological explanations is closer to the foregoing than it is to the notion central to the states and processes that cognitive science and AI focus on.

An account of representational formats suitable to cognitive science and AI can rely on none of the factors mentioned earlier, on pain of pernicious circularity. This is so because in these sciences, the notion of format at play is much more basic, furnishing part of the representational story that endows systems with the very capacities to engage in social practices and conventions, entertain intentions, interpret, form goals, and so on.

In what follows, we will thereby remain silent on how to account for the representational formats of public representations, as well as those in folk psychology. The computational view of formats we propose is designed to capture solely the notion useful for the scientific study and engineering of cognitive states. Thereby, the standards by which it is to be assessed derive from the epistemic value of appeal to formats within those scientific endeavors.

2.2. Three approaches to formats

There are several ways of carving the space of existing theories of representational format. A popular way of doing so is in terms of the number and kinds of formats that

---

[3] One way to hash this out is in terms of the personal versus subpersonal distinction. As a reviewer helpfully pointed out, the distinction can be spelled out in different ways (Drayson 2014). In the remainder of this article, we shall mainly discuss paradigmatic cases of subpersonal states for ease of exposition, yet our challenge to the mainstream approach to formats applies equally to personal-level states. Our account does not depend on the adequacy or other features of the subpersonal/personal distinction.

different views commit to. Some theories recognize only one kind of format (Pylyshyn 1973), some recognize two (Fodor 2008; Paivio 1986), and others recognize more, but not many more (Haugeland 1998). The most commonly mentioned are symbolic, discursive, iconic, analogue, discrete, and distributed formats. Depending on how each account individuates formats, some of the terms in that list may be considered to be synonymous (e.g., *discrete* and *symbolic*).

Existing theories of representational formats can be grouped into three broad categories, depending on what conceptual component of the notion of representation they take to be central to individuating formats: contents, vehicles, or the function from the former to the latter (Lee et al. 2022).

Some views take representational formats to be tied essentially to the kinds of content a representation can possess (Haugeland 1998; Peacocke 2019). In such views, representations are in different formats insofar as they represent different kinds of contents. In Peacocke's (2019) content-based view, representations in analogue format are those that represent magnitudes (i.e., that have magnitudes as their contents). According to Haugeland's (Haugeland 1998) picture, there are at least three kinds of formats, individuated by the kinds of content they represent: logical or discursive representations, which represent "absolute elements" (i.e., contents that stand by themselves, independently of relations to other elements); iconic representations, which represent "relative elements" (i.e., contents intrinsically tied to relations to other contents); and distributed representations, which represent "associative elements" (i.e., contents associated by similarity or by stimulus-response patterns).

A more common family of views takes formats to depend on the properties of representational vehicles (Beck 2019). For instance, if a representational system (only) employs representational vehicles that come in discrete types, such as the digits/voltage ranges in digital computers, then that system has a discrete format. If, on the other hand, it employs vehicles typed in terms of continuous variation across one or more dimensions, as in a mercury thermometer, the system has an analogue format.

The third family of views, the function-based account, is often conflated with the former two and especially with the vehicle-based one. This account has it that representational formats are individuated by the function that maps vehicles into contents (Lee et al. 2022). A view along these lines might, for example, identify a type of format in terms of vehicles structurally resembling, or mirroring, their contents (Beck 2018).

The debate is still open as to which of these approaches, if any, is most adequate. Challenges have been moved against all of them, typically taking the shape of examples in which they seem to yield counterintuitive results, such as categorizing a format as analogue that actually seems to be digital (Shimojima 2001). Often, such disputes are evaluated in terms of intuitions about public representations, as in pictures and sentences, or about the nature of our conscious states, such as in perception and thought. We will not delve into those discussions.

Our purposes here are exclusively constructive—namely, to detail and defend a version of the vehicle-based approach motivated and shaped by the computationalist framework in mainstream cognitive science—and aimed at producing a notion of representational format that can be useful and fruitful for cognitive science and AI research. Accordingly, our standards of evaluation for accounts of representational

formats rely not on intuitive judgments about specific cases but rather on the potential of such accounts to capture the epistemic roles and needs of cognitive science and AI and to point toward fruitful avenues of research. This distinguishes our proposal from most other approaches to formats—including vehicle-based ones—given the latter's reliance on intuitions rather than on explanatory needs, and their failure to keep folk psychological considerations separate from those most relevant to the cognitive sciences, which, as we have pointed out, are likely to involve rather different factors and constraints.

We must therefore look more closely at the roles that representational formats play and can play in the explanatory practices of the cognitive sciences in order to shed light on the nature of representational formats in biological and artificial cognitive systems. In other words, why are representational formats important for explaining cognition and intelligence? Why aren't contents and vehicles enough?

2.3. The role of formats

There is widespread agreement that representational formats play key explanatory roles in the cognitive sciences. Both early (Sloman 1978, 1994, 2002; Larkin and Simon 1987) and more recent proponents (Fodor 2008) of computation-based approaches to formats have often characterized formats in analogy to public representations. Sloman (1978, 144–76) discusses Fregean (discursive) and analogical representational structures (or "symbolism" in his jargon), such as pictures and diagrams. Fodor (2008, 171–73) distinguishes between *discursive representations*—modeled on sentences in natural languages—and *icons*, understood as akin to pictures (see also Quilty-Dunn 2019; Quilty-Dunn et al. 2023).

Analogies to public representations provide an intuitive grasp of why some explanations need to appeal to formats. As noted earlier, we use different kinds of external representations depending on what we want to achieve with them: a city map is a more immediate and flexible means to convey information about spatial layout than a series of sentences.

Let us examine in detail an example of this sort of analogical appeal to formats in science, more specifically in animal cognition research, discussed by Camp (2009). Some baboon species live in troops of varying size, sometimes comprising several dozen members, in which there are separate hierarchies of dominance-subalternity relations for males and females. There are dominance-subalternity relations between females belonging to different families, forming a hierarchy of high-status, mid-status, and low-status families. Within families, there is also a hierarchy dictated by age, with younger mature females having higher status than older sisters (with some complications; see Lea et al. 2014). Female baboons are very capable of navigating this complex, two-tiered hierarchy, behaving according to their ranks across and within families, both when they are directly involved in a dispute and when only a kin member is. They also seem to show surprise when experimenters play calls associated with encounters in which lower-ranking females challenge higher-ranking ones (Cheney and Seyfarth 2007).

The behavior of female baboons indicates that they can represent single individuals, relations of dominance between individuals and families, and occasional changes in the hierarchy. We can safely assume that the representational vehicles are

certain features of neuronal activity in the baboon nervous system. More must be said, however. How are those contents represented, such that appropriate behavior is produced, for instance, when there are changes in the dominance relations that call for prompt adaptation to a partially different social environment?

Some degree of discreteness seems to be required, such that each individual can independently come to occupy a different place in the represented social hierarchy. Similarly, some degree of combinatoriality is needed, such that changed social status changes an individual's represented dominance-subalternity relations to other individuals and families. Finally, and more tentatively, it might be expected that the relevant representations of social hierarchy be in some sense holistic, in the sense that when an individual's represented place in the hierarchy changes, all of its represented relations to other members of the group change in one go, as it were.

In light of considerations along these lines, Camp (2009) hypothesizes that the format that the social hierarchy representational system takes in those female baboons is somewhat akin to that of a tree diagram, similar to the genealogical trees that some humans are quite keen on cobbling together. Indeed, tree diagrams can represent individuals and their hierarchical relations; they have combinatorial properties; and when an individual's position in the tree changes, their relations to all other individuals change automatically, as it were.[4] (Compare: if such representations were somewhat similar to linguistic representations, then for each change in dominance relation, a large number of single representations would have to be updated—X is now higher in the hierarchy than Y; X is now higher than Z, etc.—which is arguably inefficient and cognitively taxing).

Another explanatory virtue of appealing to tree diagrams or similar formats in this case study is that it helps explain not only what female baboons can do but also what they cannot. If we were to ascribe language-like representations to baboons, insofar as they are also discrete and combinatorial, we would be left with the puzzle of why they can use such a powerful and flexible representational system to represent social hierarchy while their behavior in other tasks indicates less powerful representational capabilities.[5]

Putting aside whether it is appropriate to frame the discussion in terms of analogies to public representations, this case study illustrates that questions about the nature of the representations employed remain even after determining (or assuming) that the content and vehicle questions have been answered. These remaining questions are questions about representational format.

In brief, we need to appeal to representational formats in cognitive science and AI because they play distinctive epistemic roles: they allow us to identify distinctive features of cognition and intelligence that call for treatment in ways that are not exhausted by appeal to contents and vehicles. More specifically, representational formats are useful in cognitive science and AI, at least in large part, insofar as they fulfill the following explanatory roles:

---

[4] In a similar vein, Boyle (2019) suggests that mind-reading in apes may be underlain in some cases by yet another format, namely, *map-like* representations (see also Camp 2009).

[5] For this reason, Camp (2009) rejects Cheney and Seyfarth's (2007) claim that in light of their ability to navigate such a complex social hierarchy, baboons must thereby make use of language-like representations.

Transformation-based explanations: help explain the workings and behavioral effects of cognitive states and processes in terms of the specific kinds of transformation or manipulation available and performed over such states and processes

Efficiency-based explanations: help explain why certain cognitive states and processes are more (or less) adequate for a specific task in terms of certain sets of transformations/manipulations being more efficient, powerful, less taxing, and/or temporally advantageous[6]

In addition, a theory of representational formats should be epistemically fruitful (**epistemic fruitfulness**). First, in light of the ambiguity of many of the appeals to representational formats in the literature, a theory of representational formats should identify a clear, motivated domain of questions that can or should be tackled by such an appeal. Second, such a theory of formats needs to strike a balance between overly coarse-grained and overly fine-grained individuation of formats so as to secure a distinctive explanatory role for representational formats and avoid conflating them with contents or vehicles. Should such an attempt fail, we would have grounds to be eliminativists about formats, insofar as their job description could be filled by appeal to contents and vehicles, thus voiding their explanatory purchase. Third, a theory of representational formats for cognitive science and AI should provide insight into the nature of representational formats as explanatory posits in those sciences. It should, in other words, clarify how formats fit with other posits in the cognitive sciences, including, therefore, the related notions of representational content, representational vehicle, and computational process.

These considerations are both a job description and a list of desiderata on theories of formats suitable for the cognitive sciences. Such theories are to be evaluated in terms of the extent to which they satisfactorily provide an account that fits that job description.

With this description of the job to do, it is time to put together a job application.

3. The computational view of formats

3.1. Computational vehicles, functions, and formats

Views about the nature of representational formats in cognitive systems have typically relied heavily on the notions of computation, computational process, and computational transformation. These notions, in contrast to their use in mathematics and computability theory, are to be understood in concrete, physical terms: they are meant to capture the physical systems that are computational and carry out computational processes, such as laptops, smartphones, artificial neural networks, and plausibly, nervous systems.

Appeal to computation makes it possible to explain the behaviorally adequate transitions between and transformations of representations in a purely mechanical way—in terms, that is, of following computational rules that are appropriate to the task at hand. Computational rules are regularities in a physical system that capture the systematic transitions from inputs (and internal states) to internal states and outputs.

---

[6] For a recent account of mechanistic efficiency-based explanations, see Fuentes (2023).

Computational vehicles are individuated by their computational roles, not by the physical details of their implementation. Individuation of computational systems and the computations they perform abstracts away from implementational details almost completely: computational vehicles and processes may be equivalent in their roles and effects while differing, even radically, in what kinds of physical states and processes implement them—voltages, neuronal spike rates, or beads in an abacus. Facial identification can be achieved by means of the computations performed by populations of neurons in the fusiform gyrus of the mammalian brain, as well as, arguably, by matrix operations performed by an electronic computer, as in artificial neural networks. Only those properties that allow physical states to perform their computational roles are relevant to their computational properties. These are the degrees of freedom, or dimensions of variation, of the subset of physical properties of the physical vehicles that are computationally relevant (Piccinini 2015, 2020; Miłkowski 2013; Fresco 2014; Coelho Mollo 2018, 2019).

An important, albeit occasionally rejected (Dewhurst 2018), feature of computational systems is that they can miscompute (Fresco and Primiero 2013). They can fail to compute what they are supposed to, or in other words, they can fail to perform the computations it is their function to perform. An old pocket calculator, say, due to some dust in a transistor, may generate the wrong values, or no value at all, for an arithmetic operation it gets as input. Functions to compute may derive from design— or from human-independent processes in the case of biological systems (Piccinini 2015, Coelho Mollo 2019).

To be a computational system, therefore, is just to be a physical system of a type that can perform transformations over physical vehicles according to medium-independent rules and that has the designed or natural function to do so. Similarly, to be a computational vehicle or a computational operation is just to be a physical state or process in a computational system individuated in terms of its contributions to its computational nature.[7]

According to the mainstream representational-computational approach to cognition, representational systems in cognitive systems, be they biological or artificial, are composed of computational states and processes, some of which are also carriers of content and thus representational vehicles (Colombo and Piccinini Forthcoming). Computational vehicles are individuated by means of theories of computational individuation—such as the one briefly presented in this section— whereas representational vehicles and contents are individuated by means of theories of cognitive representation (Shea 2018; Neander 2017; Millikan 2017).[8]

Cognitive systems are regimented so that the transitions between and transformations of computational vehicles mirror the behavioral, semantic, or rational constraints relevant to the contents they carry. That is, parts of representational

---

[7] For more detail on and detailed defense of this approach to the individuation of physical computation, see Piccinini (2015).

[8] There is ongoing debate about how best to individuate computation, especially in ways that avoid computations becoming indeterminate (Fresco et al. 2021; Shagrir 2001, 2022; Piccinini 2015). Such a debate is beyond the scope of the article, but see Coelho Mollo (2018, 2019) for defense of the foregoing view against indeterminacy worries. At any rate, for our purposes, any theory of computational individuation that avoids the indeterminacy problem would be suitable, be it the one hinted at here or a different one.

systems can be manipulated computationally in ways that are appropriate to their contents. In explaining and building cognitive systems in cognitive science and AI, computation and representation typically go together, each playing a distinctive explanatory role.

3.2. Individuating representational formats computationally

Representational systems can vary considerably in their computationally relevant dimensions of variation, depending on the computational vehicles of which they are composed. Such computational vehicles can have a host of different computationally relevant properties. They can vary in the number and nature of the values they might take across multiple dimensions of variation, and changes in the values they take across one or more dimensions may lead to constraints on the values that other computational vehicles may take. We call the limitations over available values across one or more computationally relevant dimensions of variation of computational vehicles their *inner constraints*, and the mutual constraints between computational vehicles in a representational system are their *outer constraints*.[9]

In artificial systems, these constraints typically stem from design choices, as well as engineering convenience and technical limits. In biological systems, on the other hand, they likely stem from contingent features of evolutionary and developmental history, as well as the limitations imposed by the "wetware."

As an illustrative analogy, consider action figures, a popular kind of toy. One important feature that distinguishes between action figures is which parts of the toy can be moved (e.g., arms, legs, head) and how independent their movements are. Some action figures, often the cheapest ones, are fully rigid, and none of their parts can be moved. Sophisticated ones, on the other hand, have several mobile parts (legs, arms, neck, etc.), which can typically be moved independently of the others. Moreover, their limbs may move fluidly and stop at any specific position, or less satisfyingly, they may move in jerks and have predetermined stop points. Some less sophisticated ones, to great frustration, have more constrained movements: moving a forearm is impossible without moving the whole arm, or moving a leg also makes the other leg move.[10] The parts that can be moved are what we may call, with quite some stretch, the "vehicles." The number of relative positions the moving parts can occupy and the relations between variations over them (e.g., one leg also making the other leg move) are their inner and outer constraints, respectively.

We can categorize different types of action figures in terms of their moving parts, the values that those moving parts can take (which positions they can occupy relative to the body and to each other), and the relations between variations over those parts (e.g., whether moving a leg also leads to moving the other leg or, rather, an arm), and

---

[9] Outer constraints bear some similarities to what Lande (2021, 651) calls *distributional properties*, that is, the properties of a mental state that "characterise how states of that type can, cannot, or must co-occur in a particular system with mental states of other types" (2021, 651). In contrast to Lande's account, however, we focus exclusively on the relevant computational features of cognitive states and processes.

[10] Incidentally, talk of degrees of freedom is not extraneous to talk of action figures: indeed, given the former's correlation with quality (and fun), advertisements for these toys often mention their degrees of freedom explicitly.

we can do the same with computational vehicles. How many parts of the vehicle can be computationally wiggled, and what values can they take? How does wiggling one part affect the possible values of another part? And how does wiggling the values of a vehicle affect (or not) other computational vehicles—that is, does changing the values taken by one vehicle affect the values of the others?[11]

We can thus type representational systems in ways not unlike how we type action figures, that is, in terms of their computational (moving) parts, the values (positions) those parts may take across multiple dimensions, and the mutual constraints between values of the parts of different vehicles. Representational systems that differ in these respects differ in what we call their *computational profiles*.

Computational profiles are individuated by the inner and outer constraints of the computational and representational vehicles in a representational (sub)system. These factors determine what computations are available to representational systems and thus which kinds of transformations of representations are available to tackle a certain behavioral task. To type representations in this way, per the foregoing computational view of formats, is to type them in terms of their representational format.[12]

In sum, we hold that the proper way of characterizing the computational view of formats is in terms of the following set of claims:

> **T1**: Representational formats are the computational profiles of representational (sub)systems in cognitive systems, be they biological or artificial.

> **T2**: Computational profiles, in turn, are individuated by the inner and outer constraints of computational vehicles—that is, the values they can take and their mutual constraints.

It is a corollary of the view that different representational formats have different computational properties. In most cases, different formats will be best suited to solving different tasks and will require different amounts of processing steps—and thus, in real-time systems, of time—than other task-appropriate formats.

To illustrate how our proposed computational view of formats can tackle relevant questions in the cognitive sciences, we will apply it to a couple of case studies coming from the cognitive sciences, namely, the place cell system in the mammalian brain and computational models of episodic memory recall. We will show that this purely computational approach to the individuation of representational formats makes analogies to public representations explanatory redundant and, at best, of heuristic value (sec. 5).

---

[11] It is important to keep in mind that only the degrees of freedom that are computationally relevant are to be considered here (see sec. 3.1). For instance, even though physical vehicles in electronic computers can take continuous voltage values, downstream systems are only sensitive to those values falling within two specific voltage ranges. Therefore, in such a case there is only one computationally relevant degree of freedom (i.e., the voltage range is either "0" or "1").

[12] Of course, computations and representations are ultimately implemented by neural computations in brains or symbolic or numerical computations in AI systems. However, as pointed out earlier, the relevant kind of individuation for our purposes is medium independent—that is, computational and representational—rather than implementational.

4. Representational formats in the cognitive sciences

4.1 The case of place cells

Place cells are neurons found in the hippocampus of several mammals, which have a very interesting property: they fire when the animal occupies specific points in space (O'Keefe and Dostrovsky 1971; Grieves and Jeffery 2017). Together, they form a sort of array, with different (groups of) cells firing when the animal occupies different points in space. Due to this property, place cells are believed to be part of the "cognitive map" system comprising the entorhinal cortex and hippocampus and including other kinds of cells relevant to spatial cognition, such as grid cells and head-direction cells. In light of its activation properties, it seems natural to treat this system of brain areas as forming a mechanism for representing spatial locations and spatial relations in the immediate environment of mammals, given their abstract similarity to how public maps represent their content.

However, place cells are not spatially arranged in a way that corresponds to the spatial locations they respond to: there is no map-like correspondence between the relative spatial locations of place cells in the hippocampus and the relative spatial locations of points in the environment. The crucial feature of this system is the coactivation relations cells have with each other: cells that represent a certain location tend to produce activation in cells that represent nearby locations, both in online and offline tasks (Shea 2018; Diba and Buzsáki 2007; Dragoi and Tonegawa 2013).

Let us forget for a second that place cell activation correlates with spatial location and that cells that are more likely to be coactivated correlate with nearby spatial locations. Let us look purely at the computational properties of the vehicles themselves, that is, the populations of cells and their firing patterns. These patterns constitute a structure of activation relations, which can be described in terms of probabilistic coactivation relations: if cell A has firing rate a, then cells B, C, D . . . N will have firing rates in the range $x–y$ with probabilities $p, q, r, \ldots u$. Taken together across the whole system of place cells, these activation relations constitute a relational structure of computational vehicles.

The computational view allows us to examine the place cell system purely in terms of its computational features. We have a set of computational vehicles that can vary across one dimension and whose values are equivalence classes of firing rates that are treated as the same by downstream processes. The possible values depend on which and how many such equivalence classes there are, which in turn hinge on the physiological properties of the cell and those of the cells it feeds its output to—and are thus to be empirically determined. These are the inner constraints of the place cell system.

The outer constraints are more interesting in this case. If each cell probabilistically modulates the activity of the cells it is strongly connected to, then in computational terms, each vehicle's value stochastically constrains the values a subset of the other vehicles in the system may take. If vehicle V has value H (a high value, say), then vehicles C, D can take values in the range, say, M (medium) to H, with specific probabilities assigned to each downstream vehicle and possible value.[13] In other

---

[13] This is, of course, a simplification, for the sake of ease of illustration.

words, we have, roughly, a partially connected stochastic array of computational vehicles.

In brief, a description of the computationally relevant features of the place cell system comprises the following:

- A set of computational vehicles A, ..., N, implemented by the place cells
- Their inner constraints: the values that each vehicle may take, that is, the set of discrete values a, ..., n implemented by different firing rates (assuming that firing rates are what is computationally relevant)
- Their outer constraints, captured by a probabilistic function from values a, ..., n of vehicles A, ..., N to values a, ..., n of vehicles A, ..., N – 1

These computational features determine which kinds of representational roles place cells can adequately play: any representational task that involves representing concrete or abstract points in a concrete or abstract space of relations should be a good candidate. Such a computational profile seems well suited to be employed by representational systems tasked with solving spatial cognition tasks. But there is evidence suggesting that this system is also employed to solve other kinds of tasks, having to do with "distance" relations in abstract conceptual spaces (Constantinescu et al. 2016), as well as other behavioral tasks (Aronov et al. 2017; Miłkowski 2023; Mok and Love 2019; Whittington et al. 2020). Place cells may not always, nor even often, be about places. In light of the foregoing computational view, this is to be expected because the computational profile of that representational system makes it adequate for a variety of nonspatial tasks.

According to the computational view, these computational features together— that is to say, the computational profile of the place cell system as a partially connected stochastic array of vehicles—constitute a representational format. Analogies with public maps are misleading for at least three reasons.

First, as noted, there is no spatial-to-spatial correspondence relation between place cells and what they represent, as in maps. Second, the place cell system has strongly stochastic features that maps do not have. Third, the analogy to public maps erroneously suggests that the place cell system is only about space. To talk of the place cell system as having a map-like format—and thus as helping to form "cognitive maps"—is thereby misguided: the analogy with maps is very partial, and overreliance on it obscures important computational and representational features of the system.

4.2 The case of episodic memory

Episodic memory is a type of declarative memory that concerns, roughly speaking, stored information about experienced episodes, such as our memory of whom we met yesterday and in what context (Cheng and Werning 2016). Growing evidence suggests that episodic memory retrieval and recollection are generative processes of scenario construction (Cheng and Werning 2016; Lackey 2005). This means that memories are reconstructed at each retrieval through the complex, dynamic interaction of different functional areas that encode different memory traces or engrams (Sekeres et al. 2018).

The main neural locus for episodic memory is the hippocampus and its subregions, although other brain areas are involved as well (Rolls 2018; Scoville and Milner 1957).

The anterior hippocampus (aHPC) encodes the memory trace about the gist of the episode, that is, essential features like the "story elements" that are central to plot coherence (Sekeres et al. 2018).

For instance, this could be the "storyline" of your 10th birthday party—that there were other children, it was in the afternoon, and so on. The posterior hippocampus (pHPC) and the neocortex encode the memory trace with fine-grained perceptual-like details, such as the shape and color of your birthday cake (Collin et al. 2015; St.-Laurent et al. 2016). Finally, the aHPC has been shown to interact with the medial prefrontal cortex (mPFC), which stores the schema engrams, that is, networks of knowledge structures extracted from multiple similar experiences (Robin and Moscovitch 2017). In our example, this would be information about birthday parties in general.

The exact nature of the computations relevant for memory recall in the brain is still largely unknown (Cheng 2013; Rolls 2018). According to plausible theories about what is involved in recall, however, we can identify four different components: rich representations of perceptual and semantic information; a representation of the gist of the episode; an even less informationally rich memory trace that can reactivate the relevant episodic gist; and the output representation, namely, the reconstructed detailed memory that is eventually recalled, where the informational detail left out in the gist is "filled in" by recourse to rich representations of perceptual and semantic information. In other words, we have, basically, a process of lossy compression followed by a process of decompression that includes generative elements (Fayyaz et al. 2022).

There have been promising recent attempts to model this process in a biologically plausible way by means of artificial neural networks—for instance, by combining a variational autoencoder with a convolutional neural network and simple attentional selection mechanisms (Fayyaz et al. 2022). The details of such models will not concern us here; what matters for our purposes is that they provide a computational story through which the process of recall, as described earlier, may be implemented in brains and/or artificial systems. And that story plausibly involves transitions between different representational formats.

In order to shed further light on this case, it is helpful to introduce the notions of vehicular density and inner repleteness.[14] Roughly, a representational system can be more or less dense depending on whether it admits, for each pair of vehicles, a third vehicle between them or not, and a further one between this third vehicle and another one, and so on. In turn, a vehicle can be more or less replete, depending on the range of computationally relevant dimensions of variation it possesses. A vehicle may be able to take a range of values in one dimension (like a line), in two dimensions (like a shape), in three dimensions (like a solid), and so on.

A potential way to build episodic gists from perceptual information is by means of forcing rich perceptual information into a "vehicular funnel" before storage. That is to say, the system must move from a format with a high density of relatively replete representational vehicles—which, due to these features, are able to represent the fine-grained details of an episode—to a format with a rather low density of vehicles with relatively low repleteness, which encodes only the gist of the episode and thus

---

[14] These notions are inspired by Goodman (1976).

requires less storage space and may be less energetically expensive to access and reactivate. Because information is lost, this is a lossy compression process.

Memory recall, in turn, may involve a transformation from a low-density, low-repleteness format, with its highly compressed representations (the gist), into a higher-density, higher-repleteness format, marked by a qualitatively higher availability of vehicles and a larger range of possible values and mutual constraints between them. Because the compression process involves information loss, recall is partly a generative process. Gist information can provide pointers to access information stored elsewhere, for instance, in semantic memory, to fill in the information lost during compression (Fayyaz et al. 2022).

This case illustrates that in computational models and possibly in cognitive systems, vehicular density and inner repleteness are computationally relevant properties that help distinguish different formats because they involve qualitative differences in the computational profiles of representational systems. Given the lack of detailed knowledge about the specific features of the vehicles and processes underlying episodic memory recall, the foregoing computational view of formats can only provide pointers rather than a precise specification of the formats involved. However, these pointers can be precious because they help identify some of the likely features that the underlying vehicles and processes possess and thus the processing signatures that might be expected from their employment (e.g., more or less sparse connectivity, higher or lower ranges of values). Moreover, they help identify the features that need yet to be discovered so that we can have a fuller picture of the workings of episodic memory recall.

At this juncture, it is worthwhile to point out that small differences in vehicular density and inner repleteness may be overly fine-grained for the individuation of different formats. For many explanatory purposes, we may wish to generalize over formats, which would be hindered by an overproliferation of formats, leading to the near impossibility of two representational systems sharing the same format.

In the life and cognitive sciences, it is often the case that there are no sharp boundaries fixed by our explanatory concepts. For many explanatory purposes, representational formats, like other cognitive and biological concepts, should be seen as coarser grained and as having fuzzy boundaries: formats are thereby more or less well-defined clusters of computational profiles that are significantly similar in their computational capacities to be considered as identical without explanatory loss. On the other hand, some explanatory purposes may require finer-grained individuation of formats, say, if one wants to examine small but relevant computational differences between two place cell–like formats.

The computational view is thus pluralist in more than one sense. It is pluralist insofar as it recognizes a large variety of different representational formats (instead of the few intuition-based ones typically discussed in the literature), and it is pluralist insofar as it recognizes that formats may be individuated in more or less fine-grained ways depending on the explanatory aims at hand. It is likely that no immediate analogy can be made to the formats of public representation, but this is no impediment (or guide) to providing an epistemically useful notion of representational format for the cognitive sciences.

5. The explanatory roles of representational formats

5.1. Satisfying the job description without public representations

The foregoing case studies illustrate that reference to public representational formats—such as words, pictures, and maps—does not play explanatorily relevant roles and is, at worst, misleading. We contend that a purely computational approach to formats can fulfill the explanatory roles identified in section 2.3—*transformation-based explanation*, *efficiency-based explanation*, and *epistemic fruitfulness*—without any appeal to public representations. Let us look at each explanatory role in turn.

It should be quite clear that the foregoing computational view is well positioned to meet the role of transformation-based explanation. After all, it individuates representational formats by appealing to some of their computational properties, that is, their computational profiles. And the notion of computation in cognitive science and AI has as its chief role that of allowing explanations of internal state transitions that are rule based and able to respect semantic, coherence, and rationality constraints—and all that in naturalistically acceptable ways. The main innovation of the cognitive revolution was not the vindication of the notion of internal representation, which has a long history in philosophy and science, but rather the discovery of that of computation and its ensuing application to explaining how transitions between representational states can lead, mechanically, to behaviorally adequate outcomes (Fodor 1975; Haugeland 1981).

The computational view has it that the proper way of understanding formats is in terms of the computational transformations that representational systems can undergo, which are determined in turn by the nature of the computationally individuated vehicles that compose them, and the constraints they pose in light of their computational properties. By capturing such computational properties, the notion of representational format opens the way to explaining how computational goings-on in representational systems go along, or map onto, goings-on in the subject matter represented, such as to lead to adequate behavior. Therefore, the role of transformation-based explanation is satisfied: representational formats capture the computational operations available to representational systems, which have important consequences for the behavioral appropriateness of their outputs.

There are typically many different possible solutions to one and the same problem. The same applies to behavioral problems and the representational and computational states and processes that can solve them. That, of course, does not mean that every solution is equally desirable. There are better or worse, quicker or slower, more or less efficient ways of solving problems. Rube Goldberg machines, for instance, do solve problems, but in absurdly, unnecessarily complicated ways. Something similar applies to formats: some computational solutions to a behavioral problem can be more or less efficient in terms of resources employed, such as (metabolic) energy and time. The more appropriate the computational profile of a certain representational (sub)system to a task, the fewer or less expensive the computations to reach the solution will be.

In brief, by capturing the relevant computational properties of representational systems, the computational view of formats allows us to explain why certain representational formats are better suited to specific kinds of tasks—such as spatial navigation—than others. More appropriate formats will typically involve fewer,

less complex, and less expensive computations than less appropriate ones. In consequence, the view satisfies the role of efficiency-based explanation.

Moreover, this sort of consideration can be of quite some epistemic value: even though natural selection does not typically lead to optimal outcomes, it is, in any case, to be expected that it will have led to representational formats that approximate to some extent the most adequate one for a certain task. Thereby, we can try to reverse-engineer the representational format at work in a certain behavioral task by trying to find the best computational solutions to that task, and then we can assess whether behavioral, psychological, neuroscientific, or explainable AI techniques suggest that something similar is taking place in the cognitive system at hand. This is one of the aspects that makes the computational view also fulfill the third and final part of the job description, namely, epistemic fruitfulness.

5.2. Format pluralism and the fate of public representations

Approaches to formats that are modeled on public representations are typically saddled with dichotomies, such as the much-discussed one between propositional and pictorial formats. However, once we have freed the computational view of formats from the shackles of intuition, a more pluralistic perspective opens up, in which there are many different varieties of formats—many more than typically discussed. For instance, in the case of episodic memory, we have shown that vehicular density and inner repleteness are computationally relevant properties. Both density and inner repleteness are dimensions of variation that admit different degrees. There can be computational structures that are more or less dense and more or less replete, and these features can be combined in different ways in different systems.

Although we still lack a good understanding of the computational workings of cognitive systems, a purely computational view of formats sheds light on what we should be looking for when we look for representational formats in cognitive systems, be they biological or artificial, namely, computational profiles. There are no a priori limitations on what sort of computationally relevant dimensions of variation may be discovered.

The resulting picture is thus highly pluralistic because it envisages the following:

- Multiple computationally relevant dimensions that must be empirically discovered
- Graded computationally relevant dimensions rather than only all-or-nothing ones
- Multiple possible combinations of such dimensions

It is clear that this pluralism about formats goes well beyond the frequently discussed formats based on public representations.

Before we conclude, let us briefly return to public representations and look at what epistemic roles, if any, they can still play. Consider once again the case of baboon social navigation. Camp's reasoning that the format of baboon social cognition is more diagrammatic than pictorial or language-like may be construed as a *heuristic*. On the basis of observable behavior, we can put forward conjectures about what sorts of properties the underlying representational structures should possess, such that they

can explain the capacities observed, as well as the capacities that are not displayed by the system under investigation. Such initial conjectures may helpfully tap into analogies with the behavioral capacities we display when we use specific types of external, public representation.

When used as heuristic tools, public representations work as *format schemas*, that is, sketchy, tentative hypothetical models of the computational profiles that internal representational systems might possess (Machamer, Darden, and Craver 2000). Such tentative models can then be improved and adjusted in light of more fine-grained information (behavioral, psychological, neuroscientific, etc.) about the cognitive system at hand. This process is likely to generate more advanced explanatory models that depart considerably from the initial format schemas based on public representations as the heuristically useful analogies break down.

6. Concluding remarks

In this article, we have shown that the computational theory of representational formats—targeted at capturing an explanatorily useful notion of format for cognitive science and AI research—not only does not require but can actually be hindered by overbearing analogies to public representations. This computational view offers an account of what representational formats are: the computational features of physical vehicles that capture the kinds of transformation/manipulation they can undergo. We have dubbed such features *inner* and *outer constraints*, which, together, come to form computational profiles. Per the computational view, representational formats are just the computational profiles of representational (sub)systems.

Representational formats can be individuated in coarser- or finer-grained ways, depending on the explanatory purposes at hand. The computational view also detaches the question of what formats are, and how many there are, from intuitions based on public representations. We have taken some preliminary, speculative case studies from current neuroscience and computational modeling to illustrate the type of analysis that the computational view provides and the lines of further empirical investigation that it invites, both in biological organisms and artificial systems.

References

Aronov, Dmitriy, Rhino Nevers, and David W. Tank. 2017. "Mapping of a Non-Spatial Dimension by the Hippocampal-Entorhinal Circuit." *Nature* 543 (7647):719–22. doi: 10.1038/nature21692

Beck, Jacob. 2018. "Analog Mental Representation." *WIREs Cognitive Science* 9 (6):e1479. doi: 10.1002/wcs.1479

Beck, Jacob. 2019. "Perception is Analog: The Argument from Weber's Law." *Journal of Philosophy* 116 (6):319–49.

Boyle, Alexandria. 2019. "Mapping the Minds of Others." *Review of Philosophy and Psychology* 10 (4):747–67.

Burge, Tyler. 2010. *Origins of Objectivity*. New York: Oxford University Press.

Camp, Elisabeth. 2009. "A Language of Baboon Thought." In *The Philosophy of Animal Minds*, edited by Robert W. Lurz, 108–27. New York: Cambridge University Press.

Cheney, Dorothy, and Robert M. Seyfarth. 2007. *Baboon Metaphysics*. Chicago: University of Chicago Press.

Cheng, Sen. 2013. "The CRISP Theory of Hippocampal Function in Episodic Memory." *Frontiers in Neural Circuits*. doi: 10.3389/fncir.2013.00088

Cheng, Sen, and Markus Werning. 2016. "What Is Episodic Memory if It Is a Natural Kind?" *Synthese* 193 (5):1345–85. doi: 10.1007/s11229-014-0628-6

Clarke, Sam. 2019. "Beyond the Icon: Core Cognition and the Bounds of Perception." *Mind & Language* 37 (1):94–113. doi: 10.1111/mila.12315

Coelho Mollo, Dimitri. 2018. "Functional Individuation, Mechanistic Implementation: The Proper Way of Seeing the Mechanistic View of Concrete Computation." *Synthese* 195 (4):3477–97.

Coelho Mollo, Dimitri. 2019. "Are There Teleological Functions to Compute?" *Philosophy of Science* 86 (3):431–52.

Collin, Silvy H. P., Branka Miliovojevic, and Christian F. Doeller. 2015. "Memory Hierarchies Map onto the Hippocampal Long Axis in Humans." *Nature Neuroscience* 18 (11):1562–64. doi: 10.1038/nn.4138

Colombo, Matteo, and Gualtiero Piccinini. Forthcoming. *The Computational Theory of Mind*. New York: Cambridge University Press

Constantinescu, Alexandra O., Jill X. O'Reilly, and Timothy E. J. Behrens. 2016. "Organizing Conceptual Knowledge in Humans with a Gridlike Code." *Science* 352 (6292):1464–68.

Cummins, Robert. 1989. *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

Dewhurst, Joe. 2018. "Computing Mechanisms without Proper Functions." *Minds and Machines* 28 (3):569–88.

Diba, Kamran, and György Buzsáki. 2007. "Forward and Reverse Hippocampal Place-Cell Sequences during Ripples." *Nature Neuroscience* 10 (10):1241–42.

Dragoi, George, and Susumu Tonegawa. 2013. "Distinct Preplay of Multiple Novel Spatial Experiences in the Rat." *Proceedings of the National Academy of Sciences* 110 (22):9100–5.

Drayson, Zoe. 2014. "The Personal/Subpersonal Distinction." *Philosophy Compass* 9 (5):338–46.

Fayyaz, Zahra, Aya Altamimi, Carina Zoellner, Nicole Klein, Oliver T. Wolf, Sen Cheng, and Laurenz Wiskott. 2022. "A Model of Semantic Completion in Generative Episodic Memory." *Neural Computation* 34 (9):1841–70.

Fodor, Jerry. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, Jerry. 2007. "The Revenge of the Given." In *Contemporary Debates in Philosophy of Mind*, edited by Brian McLaughlin and Jonathan Cohen, 105–16. New York: Blackwell.

Fodor, Jerry. 2008. *LOT2*. Cambridge, MA: MIT Press.

Fresco, Nir. 2014. *Physical Computation and Cognitive Science*. Berlin: Springer.

Fresco, Nir, Jack Copeland, and Marty Wolf. 2021. "The Indeterminacy of Computation." *Synthese* 199 (5–6):12753–75.

Fresco, Nir, and Giuseppe Primiero. 2013. "Miscomputation." *Philosophy and Technology* 26 (3):253–72.

Fuentes, Jorge Ignacio. 2023. "Efficient mechanisms." *Philosophical Psychology*. doi: 10.1080/09515089.2023.2193216

Goodman, Nelson. 1976. *Languages of Art*. Indianapolis, IN: Hackett.

Grieves, Roddy M. and Kate J. Jeffery. 2017. "The Representation of Space in the Brain." *Behavioural Processes* 135:113–31.

Haugeland, John. 1981. *Semantic Engines*. Cambridge, MA: MIT Press.

Haugeland, John. 1998. *Having Thought*. Cambridge, MA: Harvard University Press.

Kosslyn, Stephen, William L. Thompson, and Giorgio Ganis. 2006. *The Case for Mental Imagery*. Oxford: Oxford University Press.

Lackey, Jennifer. 2005. "Memory as a Generative Epistemic Source." *Philosophy and Phenomenological Research* 70 (3):636–58.

Lande, Kevin. 2021. "Mental Structures." *Noûs* 55 (3):649–77.

Larkin, Jill H., and Herbert A. Simon. 1987. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11 (1):65–100.

Lea, Amanda J., Niki H. Learn, Marcus J. Theus, Jeanne Altmann, and Susan C. Alberts. 2014. "Complex sources of variance in female dominance rank in a nepotistic society." *Animal Behaviour* 94:87–99.

Lee, Andrew Y., Joshua Myers, and Gabriel Oak Rabin. 2022. "The Structure of Analog Representation." *Noûs* 57 (1):209–37.

Machamer, Peter, Lindley Darden and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1):1–25.

Maley, Corey. 2011. "Analog and Digital, Continuous and Discrete." *Philosophical Studies* 155 (1):117–31.

Miłkowski, Marcin. 2013. *Explaining the Computational Mind.* Cambridge, MA: MIT Press.

Miłkowski, Marcin. 2023. "Correspondence Theory of Semantic Information." *British Journal for Philosophy of Science* 74 (2):485–510. doi: 10.1086/714804

Millikan, Ruth G. 2017. *Beyond Concepts.* New York: Oxford University Press.

Mok, Robert M., and Bradley C. Love. 2019. "A Non-Spatial Account of Place and Grid Cells Based on Clustering Models of Concept Learning." *Nature Communications* 10 (1):1–9.

Neander, Karen. 2017. *A Mark of the Mental.* Cambridge, MA: MIT Press.

O'Keefe, John and Jonathan Dostrovsky. 1971. "The Hippocampus as a Spatial Map." *Brain Research* 34: 171–75.

Paivio, Allan. 1986. *Mental Representations.* New York: Oxford University Press.

Peacocke, Christopher. 2019. *The Primacy of Metaphysics.* New York: Oxford University Press.

Piccinini, Gualtiero. 2015. *Physical Computation.* New York: Oxford University Press.

Piccinini, Gualtiero. 2020. *Neurocognitive Mechanisms.* New York: Oxford University Press.

Pylyshyn, Zenon. 1973. "What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery." *Psychological Bulletin* 80 (1):1–24.

Quilty-Dunn, Jake. 2019. "Perceptual Pluralism." *Noûs* 54 (4):807–38.

Quilty-Dunn, Jake, Nicolas Porot, and Eric Mandelbaum. 2023. "The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis across the Cognitive Sciences." *Behavioral and Brain Sciences* 46:e261. doi: 10.1017/S0140525X22002849

Robin, Jessica, and Morris Moscovitch. 2017. "Details, Gist and Schema: Hippocampal-Neocortical Interactions Underlying Recent and Remote Episodic and Spatial Memory." *Current Opinion in Behavioral Sciences* 17: 114–23.

Rolls, Edmund T. 2018. "The Storage and Recall of Memories in the Hippocampo-Cortical System." *Cell and Tissue Research* 373 (3):577–604.

Scoville, William B., and Brenda Milner. 1957. "Loss of Recent Memory after Bilateral Hippocampal Lesions." *Journal of Neurology, Neurosurgery & Psychiatry* 20 (1):11–21.

Sekeres, Melanie J., Gordon Winocur, and Morris Moscovitch. 2018. "The Hippocampus and Related Neocortical Structures in Memory Transformation." *Neuroscience Letters* 680:39–53.

Shagrir, Oron. 2001. "Content, Computation and Externalism." *Mind* 110 (438):369–400.

Shagrir, Oron. 2022. *The Nature of Physical Computation.* New York: Oxford University Press.

Shea, Nicholas. 2018. *Representation in Cognitive Science.* New York: Oxford University Press.

Shimojima, Atsushi. 2001. "The Graphic-Linguistic Distinction." *Artificial Intelligence Review* 15 (1/2):5–27.

Sloman, Aaron. 1978. *The Computer Revolution in Philosophy.* Hassocks, England: Harvester Press.

Sloman, Aaron. 2002. "Diagrams in the Mind?" In *Diagrammatic Representation and Reasoning*, edited by Michael Anderson, Bernd Meyer, and Patrick Olivier, 7–28. London: Springer.

Sloman, Aaron. 1994. Toward a general theory of representations. In D. M. Peterson (ed.), Forms of representation: an interdisciplinary theme for Cognitive Science. Exeter: Intellect Books. pp. 118–140.

St-Laurent, Marie, Morris Moscovitch, and Mary P. McAndrews. 2016. "The Retrieval of Perceptual Memory Details Depends on Right Hippocampal Integrity and Activation." *Cortex* 84:15–33.

Whittington, James, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E. J. Behrens. 2020. "The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation." *Cell* 183 (5):1249–63.