Preprint of the final version to appear in: Gualtiero Piccinini (ed.) (forth.) *Neurocognitive Foundations of Mind*. Routledge

Frames of Discovery and the Formats of Cognitive Representation

Dimitri Coelho Mollo & Alfredo Vernazzani

Abstract: Research on the nature and varieties of the format of cognitive representations in philosophy and cognitive science have been partly shaped by analogies to external, public representations. In this paper, we argue that relying on such analogies contributes to framing the question of cognitive formats in problematic, potentially counterproductive ways. We show that cognitive and public representations differ in many of their central features, making analogies to public representations ill-suited to improving our understanding of cognitive formats. We illustrate these points by examining two case studies in which analogies to public representations may have had a negative impact on research: the 80's-90's debate about compositionality and cognitive architecture between symbolicists and connectionists; and contemporary discussions about the nature of visual demonstratives. Finally, we outline an alternative, computational account of formats that does not share the shortcomings of appeal to public representations.

1. Introduction

A central posit in contemporary cognitive science, broadly understood to include AI, is that cognition is at least in large part a matter of producing and using internal representations (Miłkowski 2013; Neander 2017; Shea 2018). Representations allow cognitive systems to store information about the environment and to manipulate it internally, so as to inform their behaviour in the world. What representations are about, what they represent, is their content. A map of Umeå represents the spatial layout of a nice little university town in Northern Sweden, a description of the city's winter on a tourist guide is about the features of Umeå winters. The physical states and events that realise representations are representational vehicles: specific combinations of colour and shapes on a piece of paper are a common vehicle for maps of Umeå, a sequence of written symbols a vehicle for the description of its winters.

In addition to contents and vehicles, representations also have formats. Public, external representations can have different formats: a map represents Umeå differently from how a picture does, let alone a textual description of the city. It is widely believed that the same applies to cognitive representations: they can also come in different formats. Accounts of the formats of cognitive representations have tended to focus on analogies with public representations, typically contrasting putative language-like internal representations with putative picture-like representations, and trying to establish which cognitive functions rely on one or the other sort of format (e.g., Sloman 1978; Fodor 2008; Kosslyn 1980, 1994; Paivio 1990).

Our critical target in this paper is such an approach to formats: investigating representational formats in cognitive science in terms of analogies to public, external representations.

Here is how we proceed in what follows. In §2, we briefly go into the explanatory role of representational formats in cognitive science. In §3, we outline how scientific discovery is shaped by scientific frames, and how philosophers and cognitive scientists have used public representational formats as frames for tackling questions around the format(s) of cognitive representations. In §4, we turn to the shortcomings of such a framing, illustrating them by analysing two case-studies: the compositionality debates of the late 80s-early 90s that pitched symbolicists against connectionists; and recent discussion about the nature and explanatory roles of visual demonstratives in understanding human visual perception. Finally, in §5 we outline a computational approach to cognitive formats that does away with analogies to public representations, and which we take to offer a more promising approach for improving our understanding of cognitive formats.

2. Representational formats in cognition

The explanatory value of representational formats in cognitive science and AI is due to the fact that some questions remain unanswered even after we have established what the relevant representational vehicles are in a representational system, and what contents they carry. Elsewhere, we have argued that the explanatory purchase of representational formats comes from two main questions that they help answer (Coelho Mollo & Vernazzani 2023):

- **Transformation-based explanations** help explain the workings and behavioural 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.

In other words, cognitive representational formats capture properties of representational vehicles that are relevant to providing explanations concerning the available transformations of representational vehicles, and the extent to which they contribute to making the solution to cognitive tasks more or less difficult and/or resource-consuming.[i]

It may turn out that there are no principled grounds for grouping representational vehicles in ways relevant to those explanatory roles: it may be, for instance, that representational vehicles have properties that radically vary with context in unsystematic ways, or that are overly fine-grained, such that any grouping into formats ends up being ad hoc, or explanatorily unhelpful. Should that be the case, eliminativism about representational formats would be warranted.

Another possibility, which has played an important role especially in older debates about cognitive formats, is that there is one or at most a few cognitive formats that cover all or most of cognition. The traditional dichotomous approaches to cognitive formats have largely taken this path, pitting language-like vs iconic formats, digital vs analogue, discrete vs continuous, and so on (e.g., Fodor 2008; Sloman 1978).

Yet another possibility, the one that we favour and that has received growing attention in recent years (see, e.g., Quilty-Dunn, Porot & Mandelbaum 2022) is a rather liberal pluralism about cognitive representational formats. Representational vehicles may turn out to have a variety of sets of shared properties that yield a large number of representational formats. Moreover, the properties of vehicles that are relevant for typing formats may vary depending on the explanatory aims and needs of specific research endeavours. In other words, different categorisations of formats, focusing on different sets of properties representational vehicles have, can be scientifically legitimate in light of differing explanatory targets.

Past and ongoing debates about cognitive formats have focused on analogies with the formats of the public representations we are familiar with in our everyday lives, such as sentences, pictures, and maps. This sort of analogy may help motivate monism or a limited pluralism about representational formats: natural language is analogous to the language of thought (LOT); pictures, photos, and paintings are analogous to so-called iconic formats in perception and memory; and cartographic maps are analogous to cognitive maps in entorhinal cortex. As those are the dominant formats we use in everyday life, their cognitive analogues are likely to be the main (if not the only) formats at play in cognition too.

This line of reasoning underpins a specific sort of framing of research on cognitive formats, one that takes as its starting point analogies to public representations. As we will argue, such a framing has several shortcomings, as the two domains being analogised are too different to be mutually informative.

3. Formats and Frames of Scientific Discovery

3.1. Scientific Inquiry and Frames of Discovery

The process of scientific discovery does not happen in a theoretical vacuum. Scientists inquire into phenomena guided by different epistemic concerns and goals (cf. Vernazzani ms), which derive in part from the state-of-the-art of the specific field of research, and in part from the adoption of specific theoretical frames that provide a characterization and structure to the process of discovery.[ii]

A theoretical or scientific frame can be understood as a (set of) conceptual tools that, under an intended interpretation, provides a cohesive and unitary schema for understanding a particular target phenomenon (Camp 2019). As such, frames differ from Kuhnian paradigms (Kuhn 2012)—at least under one sense of 'paradigm'—in being less general and coherent: they mould research by suggesting hypotheses, possible interpretations, and explanations, and by embodying specific norms for scientific inquiry within a particular field. A frame used for the purposes of scientific discovery can be seen as a lens or template that directs research, even though a frame may later be dropped for another one, or coexist with other frames considered to be useful for the same domain of research.

We can construe past and ongoing research on the formats of cognitive representations as essentially guided by different frames of inquiry. According to what we may call—not without some simplification—the 'received view', the formats of cognitive representations are modelled on the formats of public representations, most paradigmatically on natural languages (such as English or French), and on icons, a broad category including pictures, paintings, maps, and photographs.

Our aim in this paper is to show that the traditional, public representation-based frame to understanding formats is deeply problematic, as it can lead our research efforts into unfruitful directions. An alternative, computation-based account of formats (Coelho Mollo & Vernazzani 2023), we will argue, is preferable.

3.2. Public Representations as Frames

Current research on the format of cognitive representations has often used public representations as frames of discovery. Public representations include a broad range of representational artefacts: paintings, photographs, maps, linguistic scripts and many other kinds of signs. Philosophers and cognitive scientists typically identify two broad categories or families of public formats: icons and language-like representations (e.g., Fodor 2008; Kosslyn 1980, 1994; Paivio 1990). Applied to cognitive representations, this distinction has been taken to undergird important differences in cognitive architecture. For example, Carey (2009), Burge (2010), and Block (2023) argue that the difference between perception and cognition is to be cashed out in terms of different formats. More specifically, they argue that while perceptual representations are typically iconic or image-like, conceptual cognitive representations are instead language-like.

The strategy typically adopted by philosophers and cognitive scientists using public representations as frames of discovery is to hypothesise functional and/or structural similarities between relevant properties of a public format and the properties of the cognitive representations putatively at play in a cognitive function of interest. This is often motivated by apparent similarities between the kinds of tasks we solve by using external representations and the tasks cognitive systems have to tackle (Coelho Mollo 2022), as well as by considerations regarding the required expressive power of representational systems, be they external or internal, for them to be appropriate to a task. Such similarities are then used to suggest that the sorts of operations and transformations of vehicles made available to users by a public representational format are analogous to the computations made available to cognitive (sub)systems by its analogous cognitive format. It will be instructive to discuss one example of this strategy.

Fodor (2008) contends that pictures, just like sentences, have a compositional semantics. However, he maintains that pictures and natural languages obey different principles of compositionality. Specifically, he argues that pictures obey what he calls the 'Picture Principle': if P is a picture of X, then parts of P are pictures of parts of X. In Fodor's view, this principle undergirds the fact that pictures, unlike sentences, do not admit of a canonical decomposition, each part of a picture of a person represents a part of a person, whereas a sentence can only be decomposed into semantically meaningful units according to specific rules.

A direct application of Fodor's Picture Principle to cognitive representations can be found in Quilty-Dunn (2016).[iii] Quilty-Dunn argues that the segmentation of the visual scene into discrete objects requires a canonical decomposition into distinct constituents, each requiring a distinct vehicle. This is apparent in the case of multiple object tracking (MOT) (see also below, §4), where test subjects are shown different items on a display, among other confounds, and are tasked to keep track of a selected number of objects. Studies indicate that subjects are typically capable of easily tracking up to four or five distinct objects (Pylyshyn 2007). This limitation may be due to the fact that the visual system has a limited number of simultaneously active vehicles to represent and keep track of objects. Now, if the visual scene were represented iconically—so Quilty-Dunn argues—it would be impossible for the visual system to represent objects as segregated from other objects and the scene; this because the segmentation of objects requires different representational vehicles, which in turn is strongly indicative of a canonical decomposition into discrete objects. However, this is the sort of thing that should be impossible were the representation iconic, since, in virtue of the Picture Principle, icons do not admit of canonical decomposition.

This sort of case nicely illustrates the strategy discussed above. The operating assumption is that there is a structural similarity between some alleged properties of icons (public representations), i.e., their non-canonical decomposition, and the cognitive representations that undergird visual object perception and scene perception. The kind of compositional semantic operating in the case of object perception does not seem to obey the Picture Principle, hence, allegedly, i such representations cannot be iconic.

Although we are about to criticize adopting public formats as frames of discovery for the format of cognitive representations, such a strategy is not unmotivated. Arguably, one reason for adopting public formats as frames of discovery is the widespread use of analogies in science (Hesse 1966). A well-known example is the wave equation, which can be used to study mechanical waves such as water and sound waves, but also electromagnetic waves. The underlying assumption is that all these different phenomena display a wave-like dynamic that can be aptly captured by the same equation.

A similar line of thinking may underlie the use of public formats to model the format of cognitive representations. On the assumption that the two families of representations may have interesting analogies, one may study the computational properties of public formats to gather insights into the format of cognitive representations. In doing so, however, it is crucial to keep in mind the abstractions and idealisations involved in such a process, as the two domains being analogised are likely to have crucial differences, and only partially share interesting features. Failure to recognise such abstractions and idealisations may lead to improper extension of features of one domain to the other, leading researchers astray (Coelho Mollo 2022). As we will argue, framing cognitive formats in terms of public formats exemplifies such a risk.

4. The limits of the public representation frame

Our strategy to assess the adequacy of taking public representations as a frame to better understand the formats of representational formats in cognition is two-pronged. First, we will focus on theoretical arguments about what public and cognitive representations have in common, and what sets them apart. As we will argue, the similarities between these two domains of representational phenomena are limited, while the differences are substantial. This suggests that using one to help understand the other is epistemically risky.

Second, we will briefly examine two case studies that, we believe, illustrate the perils of relying on the public representation frame when investigating cognitive formats: the debate in the late 80's and early 90's about compositionality and cognitive architecture between Fodor & Pylyshyn (1988), Smolensky (1987), and others; and contemporary work on the nature and roles of 'visual demonstratives' in visual perception (Burge 2018, 2023; Green & Quilty-Dunn 2021; Pylyshyn 2007).

4.1. Public vs Cognitive Representations

Let us take a closer look at what public and cognitive representations have in common, and what distinguishes them. Properties that they share include the following:

1. being about, or meaning, or having as content something, be it concrete or abstract, vague or definite, actual or imaginary;
2. having some physical vehicle(s) that implement the representation, i.e., that 'carry' the content, be they a sheet of coloured paper, strings of digits in a solid-state memory drive, or the activity of neurons in brains;
3. standing in appropriate, often complex sets of relationships to something, as described in 1., which make it so that the vehicle, as described in 2., comes to have the content that it does. (Shedding light on what these relations are is the core mission of the project of naturalising representation, see Neander 2017, Millikan 2017, Shea 2018);
4. playing a representational functional role, i.e., being used by some system to represent something else.

While these are important similarities, they are features that are essential to being a representation at all. Public representations and cognitive representations, in brief, have contents, are implemented by physical vehicles, and come to be representations courtesy of both appropriate relationships between vehicles and what they represent, and appropriate functional roles. Insofar as they share these features, both phenomena of interest count as representational. If any of them lacked one or more of those features, they would not count as representations. In other words, the features that public and cognitive representations have in common are the minimal ones that make them both cases of representation.

A closer look at their more specific features, over and beyond those that make them into representations, reveals crucial differences.

First, there are important differences between the sorts of vehicles used for public representations—ink on paper, light on screens, paint on surfaces, etc.—and the sorts of vehicles thought to be relevant for cognitive representations, which ultimately involve aspects of the activity of brain cells, especially neurons in the neocortex. While this may seem like a relatively trivial difference, having to do 'merely' with implementational details, it is not so. Implementational details constrain the sorts of transformation that representations can undergo. Ink on paper cannot typically be erased or overwritten without loss of clarity, while Unicode in computer memory can. Moreover, implementational details help determine which systems can access and make use of a representation: paint on the surface of neurons is unlikely to be something that downstream cognitive systems can use, as they are not sensitive to the colour of neurons.

More generally, the implementation of public representations hinges on communicative intentions, and complex socio-historical processes that led to the maintenance of certain conventions over others, with one guiding constraint being interpretability by the humans target of the representation—which in its turn depends on their familiarity with the chosen conventions.[iv] The implementation of cognitive representations, on the other hand, is guided and constrained by rather different factors. They depend on mind-independent evolutionary and developmental constraints, as well as on the fact that the consumer systems are not full-fledged interpreters, but ultimately 'dumb homunculi' (Dennett 1991).

Second, the relationships between vehicles and what they are about that partly endow representations with content are fundamentally different between external public representations and internal, cognitive ones. In the former, a combination of individual propositional attitudes (e.g., one's intention to convey a certain content to an interlocutor); socially-established conventions (e.g., crosses for Christian churches in city maps); and social practices (e.g., Gricean norms of communication) all play a role in establishing the contents of a representation. None of these factors, however, are part of (naturalistic) accounts of cognitive representation. Cognitive representations are supposed to underlie propositional attitudes, and, more indirectly, conventions and social practices. Thus, on pain of explanatory circularity, cognitive representations must stand in other kinds of relation to what they represent.

Naturalistic theories of content have settled on a combination of causal-informational and teleofunctional relations as the ones that determine the contents of cognitive representations (Millikan 2017, Neander 2017, Shea 2018, Coelho Mollo 2022). These are natural relations that do not involve nor presuppose mental states and/or social conventions or institutions. While some theorists take such naturalistic theories to also cover or at least help explain public representations as well (e.g., Millikan 2017), others are sceptical that there is any straightforward extension of an account of representation that applies to cognitive states to one that also applies to public representations, and vice versa (Neander 2017, Shea 2018, Coelho Mollo 2015).

Finally, when it comes to the functional roles of public and cognitive representations, we also find substantial differences. As we pointed out, public representations are typically means of communication between two or more full-fledged intelligent, cognitive agents (including the same agent at different times, as in personal notes and reminders). In most cases of public representation, there are intentions to communicate something to someone else, often with a specific valence, and relying on an assumption that the communicative intention and the representation's 'code' are grasped by the receiver.

Typical cases of cognitive representation, on the other hand, are infra-agent, and do not involve communicative intentions. They are produced by cognitive subsystems and are 'read' or 'consumed' by other cognitive subsystems. Such a reading involves no process of interpretation—at least in its common sense—but merely a mindless, 'mechanical' responsiveness by downstream subsystems to certain features of the representation.

In brief, public and cognitive representations share only those features that make them both representations. When it comes to the central details about how they come to be representations, the stories that need to be told are very different. The kinds of vehicle, content-fixing relations, and functional roles at play in public and cognitive representations importantly differ. This suggests that public representations are a poor frame for understanding the formats of cognitive representations.

4.2. Two Case-Studies

Far from being of merely theoretical interest, framing cognitive representational formats in terms of analogies to public representations can actively hinder progress in understanding cognition. We will briefly present two such cases: one in which analogies to public representations led to overstrict views about how a certain capacity need be implemented (the compositionality of thought debate between classicists and connectionists); and one in which allegiance to the public representation frame led to positing cognitive components that may not really exist (visual demonstratives).

The compositionality debates

In the late 80's and early 90's, the connectionist, neural network-based approach to artificial intelligence and cognitive science gained in (short-lived) popularity with the growth of what came to be known as the PDP (Parallel Distributed Processing) programme (Rumelhart, McClelland & PDP Research Group 1986). Some of its proponents argued that connectionism offered an alternative account of cognitive architecture to the classical paradigm. The latter posits that cognitive processes are a matter of symbolic computations over atomic representations (or molecular compositions thereof), with a syntax similar to that of natural (and artificial) languages (Fodor 1975)[v]; a hypothesis that has come to be known as the Language of Thought Hypothesis, or LOTH. The former, on the other hand, has it that cognitive computation is non-symbolic, being rather numerical, or at most 'subsymbolic'; and that representations are distributed and partially overlapping across nodes in the network, instead of being discrete (e.g., Smolensky 1987).

Fodor & Pylyshyn (1988), central proponents of LOTH, provided a detailed attack of connectionism as a candidate cognitive architecture. Their core argument was simple. Thought is productive—i.e., it can produce an indefinite number of different thoughts—and systematic—i.e., being able to think certain thoughts is intrinsically tied to the ability to think other thoughts: a mind that can think 'John loves Mary' can also think 'Mary loves John'. But productivity and systematicity require thoughts to have constituent structure, that is, discrete parts that retain their identity throughout and that can be combined in rule-governed ways. Thought must thereby be compositional. However, representations in connectionist networks lack constituent structure, and their computations are based on stochastic relations, rather than strict symbolic rules that govern how representations can and cannot be combined. Therefore, connectionist networks cannot underpin a compositional representational system, and thus cannot account for the productivity and systematicity of thought. Connectionism thereby fails as a claim about cognitive architecture.

Interestingly, Fodor & Pylyshyn's (1988) attack against connectionism is almost purely theoretical—they make little to no recourse to empirical evidence, and they occasionally rely on intuitions about what kinds of minds can or cannot exist.[vi] Crucially, they endorse two central undefended assumptions: a) sentences in natural language express thoughts; b) sentences and thoughts have similar structures (the underlying idea being arguably that b) is needed if sentences are to adequately express thoughts).

Unsurprisingly, given such assumptions, Fodor & Pylyshyn's arguments move seamlessly between features of language and putative features of thought. As natural languages are productive, systematic and compositional, so must thought be, as the former expresses the latter.

Assumption b), however, indicates a serious 'lack of imagination' on the part of Fodor & Pylyshyn, as Chalmers (1993) puts it. While natural language provides a paradigmatic case of a productive, systematic, and compositional representational system, other kinds of representational system may also realise, or at least approximate, those features so as to lead to appropriate behaviour.[vii] Those representational systems may however have no analogue among the public representations we are familiar with. Indeed, connectionist networks have been shown to have the representational means to, as a minimum, approximate compositional competence, despite lacking the sort of explicit constituent structure and symbolic composition rules that Fodor & Pylyshyn take as characteristic of language-like formats (Smolensky 1987, 1990; Lepori, Serre & Pavlick 2024; see Coelho Mollo 2024 for extended argument and illustration).[viii]

At any rate, Fodor & Pylyshyn (1988) display a lack of imagination also because they do not consider that expressing thoughts via natural language may involve a transformation of format from a potentially non-LOT cognitive format to the natural language format. Natural language, rather than being the analogue of the cognitive format, may actually be just an approximate means to express thoughts that do not share the same strictures of a language-like format (Smolensky 1987). Classical compositionality may be an artefact of language use, rather than a feature of the cognitive architecture of minds.[ix]

In sum, framing cognitive representational formats in analogy with public representational formats seems to have worked as a limit on imagination for the defenders of LOTH within that old debate. The constraints on possible cognitive formats are different from the historical, socio-cognitive constraints on public representation systems. To frame or model the former in terms of the latter risks constraining our scientific imagination, dismissing options that are worth exploring, such as the possibilities allowed by large numbers of numerical computations in artificial neural networks.[x]

Visual Demonstratives

A *Perceptual Object Representation* (or POR, for short) is the representation of a single visual object, a coherent visual whole (e.g. Feldman 2003; Green 2018a; Rubner & Vernazzani, ms; Vernazzani 2022). PORs are postulated to explain different feats of our visual system. Firstly, they are posited as a solution to the synchronic binding problem, i.e., the problem of unifying different surface and spatiotemporal features (like colour, shape, trajectory, etc.) into a single coherent whole. Secondly, they are supposed to provide a solution to the correspondence problem, i.e., the problem of explaining how the visual system re-identifies an object as the same in a dynamic context, despite featural and spatiotemporal changes (Flombaum et al. 2003; Quilty-Dunn & Green 2023; Scholl 2007).

It is generally agreed that PORs include representations of an object's features. In addition, some researchers contend that an additional component, a *visual demonstrative* or *pointer* must be added (e.g. Alvarez & Franconeri 2007; Echeverri 2017; Pylyshyn 2003, 2007; Scholl et al. 2001). The main motivation for the introduction of a demonstrative component in PORs comes from studies on dynamic object tracking, as exemplified by experiments on multiple object tracking (MOT).

MOT studies require test subjects to lock reference to a limited number of targets—e.g., one or more white circles—displayed on a screen while the targets move among other figures, which work as confounds (Pylyshyn & Storm 1988). In some studies, the targets do not simply move, but also change some of their surface features, such as form and colour. Typically, subjects can easily target four or five such objects (Pylyshyn 2007; Green 2018b; cf. also Alvarez & Franconeri 2007). Since objects are tracked despite spatiotemporal and featural changes, it has often been argued that reference is secured to the object itself via a demonstrative-like component and not via the object's properties:

> Unless at some point one can think the equivalent of '*This* has property P', one cannot refer to a particular object token in a way that will allow it to be bound to the arguments of a visual predicate or to serve as the basis for action (e.g. to

> point toward or grasp the token object in question) (Pylyshyn 2003, p. 254; cf. also 2007, p. 95).

The visual demonstrative may be construed in different ways. Burge construes it as a complex demonstrative embedded into a proto-propositional structure such as 'This *F*' (2018, 2023), where the property *F* is not predicated of the object but merely attributed to it. Alternatively, one may combine the complex demonstrative account with a sentential and predicative structure—such as 'This *F* is *G*' (e.g., Quilty-Dunn & Green 2023)—or cash out the visual demonstrative as a simple demonstrative plus a predicate, such as 'This *a* is *F*' (e.g. Fish 2009; Fodor & Pylyshyn 2015).

Another way of understanding the visual demonstrative component of PORs is modelled after pointer architecture, a computational structure that represents a reference to a feature stored in memory (Green & Quilty-Dunn 2021). This latter way of construing the visual demonstrative may be coupled with object files, where the pointer plays the twofold role of anchoring reference to a worldly object, while also relating it to information stored in short-term memory (Kahneman et al. 1992).

For our present purposes, it is critical to see that, however visual demonstratives are construed, they suggest a language-like format of PORs. This is very explicit in Pylyshyn's account of the structure of visual objects within his broadly Fodorian approach, which that embraces both modularity and LOTH (Fodor 1983). More recently, Green & Quilty-Dunn (2021) have similarly suggested that object files have a propositional structure encoding discrete symbols that can be syntactically combined in different ways.[xi]

In this context, the format of a public representational structure, i.e., language, works as a frame for scientific inquiry and discovery, structuring and guiding our understanding of PORs' nature and organisation, as well as of the way they refer to worldly objects. Furthermore, adopting such a frame of discovery makes some questions more salient: e.g., how should we understand visual demonstratives, as complex or simple demonstratives? Such research questions invite a way of treating the problem that draws on the resources provided by linguistics and the philosophy of language, at the same time downplaying other possible ways of thinking about it in non-linguistic terms.

As we have seen, visual demonstratives are typically introduced to explain dynamic object tracking and to provide a solution to the correspondence problem. However, there are reasons to be sceptical about whether they are needed for these explanatory tasks. In work in progress, Rubner & Vernazzani (ms) argue that there can be a satisfactory solution to the correspondence problem without appeal to visual demonstratives. The argument unfolds in two steps.

First, they point out that empirical evidence strongly suggests that properties of objects are all that is needed to keep track of objects in a dynamic context. Studies have shown that spatiotemporal properties, such as topological properties or an object's trajectory play an important role in object tracking (Mitroff & Alvarez 2007; Zhou et al. 2010). More recently, it has been shown that surface properties such as colour and shape are also critical for object tracking in some contexts (Jiang 2020; Hein & Moore 2012).

There is thus ample empirical evidence suggesting that properties are all that is needed in keeping track of items, where some properties of the object are used for the purpose of reference guidance or tracking, while other properties take a subordinate role, as they belong to the POR although they are not used for the purpose of tracking (Quilty-Dunn & Green 2021). Which properties the visual system uses for tracking depends on context. For instance, Tremoulet et al. (2000) show that infants track objects through shapes and colours when spatiotemporal information does not provide sufficient criteria of individuation. Thus, the visual system seems to adopt flexible strategies to the problem of object tracking.

Once we acknowledge the critical role of properties in object tracking, we can articulate a solution to dynamic object tracking tasks without resorting to visual demonstratives. As we have seen, visual demonstratives are introduced to account for the fact that the visual system can keep track of an object despite featural change. Supporters of visual demonstrative have taken this as evidence that object tracking is not achieved via properties. Rubner & Vernazzani, however, advance an alternative solution to the correspondence problem that does not require the postulation of visual demonstratives, relying on what they call the Continuity Principle:

> For any objects O_1, O_2, sets of properties F_1, F_2, and times t_1, t_2, where t_1 and t_2 occur sequentially, if F_1 *guides reference to* O_1 and F_2 guides reference to O_2, then O_1 at t_1 may be perceived as the same as O_2 at t_2 only if $F_1 \cap F_2 \neq \emptyset$.

The key to solving the correspondence problem according to the Continuity Principle is that the visual system does not track only one single property of the object, but rather a set of properties. As long as the intersection of an object's properties used for tracking at two different times is non-empty, reference is secured and the correspondence problem solved, without any need to introduce visual demonstratives in PORs architecture.[xii]

It is helpful to illustrate how the Continuity Principle works in practice. Take a case of dynamic object tracking with featural change, e.g., an object moving from the left-hand side of the visual scene at t1 to the right-hand side of the scene at t2 while changing colour (say, from red to green) and shape (say, from rectangular to triangular). According to the Continuity Principle, the visual system is able to keep track of the object and solve the correspondence problem, i.e., to identify the object as the same in spite of featural change, because some of the object properties from t1 to t2 are preserved, in this case, the spatial trajectory, where the spatial trajectory belongs to the set of properties F that the visual system uses for reference guidance or tracking. Notice that the Continuity Principle also enables us to make empirically testable predictions, for it follows from it that object tracking would be disrupted just in case from t1 to t2 all properties used for tracking were lost and replaced by other properties. In real-life contexts, this condition virtually never obtains, and this illustrates how the Continuity Principle captures the robustness and flexibility of object tracking.

Accepting the Continuity Principle provides an empirically sound solution to the correspondence problem without resorting to (natural) language as a frame for modelling object tracking.[xiii] A first, direct implication is that the Continuity Principle makes the postulation of visual demonstratives unnecessary. By removing visual demonstratives from the structure of PORs, further questions directly linked with their posited architecture are removed as well: there is no need to draw on the research on linguistic demonstratives to best capture the nature of visual demonstratives, and no need to think of them as simple or complex demonstratives.

Second, once the need for visual demonstratives is removed, one reason for thinking of PORs as articulated in a language-like format (see Green & Quilty-Dunn 2021) is undermined as well. There is no need to construe them as either proto-attributional structures—as in Burge (2023), 'That *F*' where *F* is a visual attributive—or predicational—as in 'That *F* is G' or 'That *a* is *F*' (Fish 2009).

We do not wish to suggest that the entire question of the language-likeness of the format of PORs rests on visual demonstratives. Rather, our discussion was meant to highlight two important lessons. First, defenders of visual demonstratives read the empirical studies and interpret the correspondence problem through the lens of the public representational format frame, thus inviting specific further research questions, e.g., about the nature of visual demonstratives and their roles in PORs architectures. But an empirical and conceptual solution to the correspondence problem does not necessitate a language-like format for PORs.

This brings us to the second lesson. Framing the issue through the public representational format frame downplays alternative ways of interpreting the evidence and construing the correspondence problem. This may obfuscate other empirically viable approaches, such as the one offered by the Continuity Principle. Embracing the Continuity Principle and dismissing visual demonstratives as a viable solution to the correspondence problem leads to a reconfiguration of the problem, highlighting some issues—e.g., whether there is a hierarchy of property-types for purposes of tracking—while discarding others as conceptual artefacts—e.g., how best to construe visual demonstratives.

In brief, the public representational format frame leads to hypotheses and lines of research that may be less fruitful than alternatives, but that seem natural due to the explicit or implicit adoption of that intuitive frame. In the next section, we put forth an alternative frame for thinking about the formats of cognitive representations that draws, instead, on foundational notions in the cognitive sciences.

5. A computational account of formats

In a recent paper, we have put forward and defended a view of representational formats that takes as its starting point the explanatory needs of the cognitive sciences (Coelho Mollo & Vernazzani 2023). Such a starting point casts suspicion on the idea that categories that are useful for understanding the formats of external, public representations should also be applicable to internal, cognitive representations. Indeed, while in both cases we are dealing with representations, they are of such different types that it is implausible to think that a single account, or even two very similar accounts, can capture them both.

If public representations, as we have argued, are indeed ill-suited as a frame for understanding the formats of cognitive representations, we must look for alternatives that are more likely to suit the explanatory needs of cognitive science. We have developed a theory of representational formats for the cognitive sciences that relies on an approach that has proven to be enormously scientifically fruitful: namely, the use of the notion of concrete computation in accounting for cognitive processing—a foundational insight of mainstream cognitive psychology and cognitive neuroscience that underpins much of the progress these fields have made in the past few decades (Miłkowski 2013; Piccinini 2020).

The resulting computational view of formats that we suggest identifies cognitive representational formats with what we call 'computational profiles'. Computational profiles are features of cognitive vehicles. More specifically, they are the set of features of computationally individuated cognitive vehicles that constrain and govern which computational operations and transformations vehicles can undergo. When computational vehicles in cognitive systems are also representational vehicles, their computational profiles are their representational formats.

We distinguish two types of constraints on vehicles, which we dub 'inner constraints' and 'outer constraints'. Briefly, inner constraints capture the ways in which parts of a computational-representational vehicle can vary: how many of its parts play computational roles, and which values those parts can take. External constraints, on the other hand, are those constraints that come from other computational-representational vehicles within the same representational (sub)system: that is to say, how values taken by parts of vehicles V2,..., Vn in the same (sub)system influence the values that can be taken by a vehicle V1.

A helpful, though rather rough, analogy is with action figure toys. If we take each limb of the toy to be a vehicle, the inner constraints are the positions each limb part can take, which tend to vary with the design and quality of the toy. The outer constraints are instead the mutually constraining relations that may exist between limbs: for instance, in some such toys having an arm at a certain position constrains the positions that the other arm can take (in the simplest case, both arms must always be in the same relative position).

While the inner and outer constraints in action figure toys help determine their degrees of freedom—an important selling point—the computational profiles of representational vehicles in cognitive systems help determine which computational transformations those vehicles can undergo, and thus how cognitive representations can be manipulated. This, we claim, is just what cognitive representational formats are, and why they are scientifically useful notions: they allow us to capture features of cognitive processing that shape how representations can be transformed by downstream systems to contribute to furthering the systems' goals.

The computational view of formats makes use of conceptual tools that are not only widely employed, but also foundational to cognitive psychology, neuroscience, and cognition-focused AI. This makes it possible to inform and assess hypotheses about representational formats by appeal to theoretical and empirical results from those fields—as our case studies illustrate. This stands in stark contrast with the extra-scientific origin of the public representation frame, which comes from everyday representational practices of human individuals and societies, rather than the posits of the cognitive sciences.

While extra-scientific concepts and analogies can be heuristically useful in some cases, employing them comes with risks (Coelho Mollo 2022). Indeed, as we have pointed out in section 4, the public representation frame has several shortcomings, some theoretical and some that can be seen in how certain debates in the field have been (unhelpfully) structured. Looking at the case studies we examined, does the computational view fare any better? We believe that it does.

With regard to the issue of the compositionality of thought that exercised Fodor, Pylyshyn, Smolensky and many others in the 80's and 90's, and which has seen a rebirth in recent times with the advent of Large Language Models, the computational view does not share the limits of the public representation frame. As it is not wedded to analogies to natural language, the computational view does not invite a view of compositionality that hinges on how language comes to have that feature. It thus opens the way to seeing the compositionality of thought in a different light, as a phenomenon that may, but need not, be akin to linguistic phenomena. Moreover, by directing focus toward computational profiles and computational transformations of representational vehicles, the computational view invites the idea that similar capacities may be computationally realised in different ways. As much as there are different algorithms for division, there can be different computational means to realise behaviourally appropriate compositionality of thought, with Smolensky's suggestion being one such alternative, and current research on Large Language Models revealing yet others (e.g., Lepori, Serre & Pavlick 2024, Merullo et al. 2024).

Similarly, positing visual demonstratives hinges on a rather clear analogy to the role played by demonstratives in natural language. Indeed, the notion comes from researchers committed to the idea that cognitive systems employ mostly or exclusively language-like formats. The computational view works under no such assumptions. Cognitive formats are likely not to resemble anything we are familiar with when it comes to public representations, for the reasons suggested in section 4. A language-like demonstrative is one potential way in which the correspondence problem in visual perception can be tackled, but far from the only one. We should therefore actively investigate alternatives, such as the one sketched above and in (Rubner & Vernazzani ms), which may be more in line with what we know about the workings of the brain.

Conclusions

In this chapter, we have presented two competing strategies for tackling the nature and roles of the format(s) of cognitive representations. The received view uses a scientific frame that draws from the formats of public representations, such as maps, pictures, drawings, and natural languages. Researchers adopting such a framing typically start by singling out some computational properties of the former, and then they compare such properties to the properties of a target cognitive system or subsystem. We have questioned the fruitfulness of this strategy, pointing to its theoretical limits, and illustrating its shortcomings with two case studies. We can make headway into thinking about the format of cognitive representations once we sever, or at least considerably weaken, associations with public representational formats, and focus instead on the computational profiles of representational vehicles, and the explanatory roles they can play in our theories of mind and brain.

References

Alvarez, George A. & Steven L. Franconeri (2007) “How many objects can you track? Evidence for a resource-limited attentive tracking mechanism” Journal of Vision 7(13), pp. 1-10.

Barack, David (2024) “What is foraging?” *Biology & Philosophy* 39. https://doi.org/10.1007/s10539-024-09939-z

Block, Ned (2023) “Let’s Get Rid of the Concept of an Object File” in B. McLaughlin & J. Cohen (eds.) *Contemporary Debates in Philosophy of Mind* (pp. 494-516). New York: John Wiley.

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

Burge, Tyler (2018) “Iconic representation: maps, pictures, and perception” in Shyam Wuppuluri & Francisco Antonio Doria (eds.) *The map and the territory* (pp. 79-100). Cham: Springer.

Burge, Tyler (2023) *Perception: First Form of Mind*. New York: Oxford University Press.

Camp, Elisabeth (2009) “A Language of Baboon Thought.” In R.W. Lurz (ed.) *The Philosophy of Animal Minds* (pp. 108–27). New York: Cambridge University Press.

Camp, Elisabeth (2019) “Perspectives and Frames in Pursuit of Ultimate Understanding” in S. Grimm (ed.) *The Varieties of Understanding* (pp. 17-45). New York: Oxford University Press.

Carey, Susan (2009) *The Origin of Concepts*. New York: Oxford University Press.

Chalmers, David J. (1993) “Connectionism and Compositionality: Why Fodor and Pylyshyn were wrong” *Philosophical Psychology* 6(3), pp. 305-319.

Clark, Austen (2000) *A Theory of Sentience*. New York: Oxford University Press.

Coelho Mollo, Dimitri (2015) “Being Clear on Content - Commentary on Hutto and Satne” *Philosophia* 43, pp. 687–699.

Coelho Mollo, Dimitri (2022) “Deflationary Realism: Representation and idealisation in cognitive science” *Mind & Language* 37(5), pp. 1048-1066.

Coelho Mollo, D. (2024). AI-as-Exploration: Navigating intelligence space. arXiv:2401.07964.

Coelho Mollo, Dimitri & Alfredo Vernazzani (2023) “The Formats of Cognitive Representations: A Computational Perspective” *Philosophy of Science*. doi:10.1017/psa.2023.123

Colombo, Matteo & Gualtiero Piccinini (2023) *The Computational Theory of Mind*. New York: Cambridge University Press.

Cohn, Neil (2013) *The Visual Language of Comics*. London: Bloomsbury.

Dennett, Daniel C. (1991) *Consciousness Explained*. Boston: Little, Brown & co.

Echeverri, Santiago (2017) “Visual Reference and Iconic Content” *Philosophy of Science* 84(4), pp. 761-781.

Green, E.J. (2018a) “A Theory of Perceptual Objects” *Philosophy and Phenomenological Research*. doi: 10.1111/phpr.12521

Green, E.J. (2018b) “What Do Object Files Pick Out?” *Philosophy of Science* 85, pp. 177-200.

Green, E.J. & Jake Quilty-Dunn (2021) “What is an object file?” *British Journal for the Philosophy of Science* 72(3), pp. 665-699.

Feldman, Jacob (2003) “What is a visual object?” *Trends in Cognitive Science* 7(6), pp. 252-256. doi: 10.1016/s1364-6613(03)00111-6

Fish, William (2009) *Perception, Hallucination, and Illusion*. New York: Oxford University Press.

Fodor, Jerry (1975) *The Language of Thought*. New York: Thomas Crowell.

Fodor, Jerry (1983) *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. (2001). *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press

Fodor, Jerry (2008) *LOT2*. New York: Oxford University Press.

Fodor, Jerry & Zenon Pylyshyn (1988) “Connectionism and Cognitive Architecture: A Critical Analysis” *Cognition* 28 (1-2), pp. 3-71.

Fodor, Jerry & Zenon Pylyshyn (2015) *Minds Without Meanings*. Cambridge, MA: MIT Press.

Haziz, Eliran (2023) “Norms of Inquiry” *Philosophy Compass*. DOI: 10.1111/phc3.12952.

Hesse, Mary (1966) *Models and Analogies in Science*. Indiana: University of Notre Dame Press.

Hein, Elisabeth & Cathleen M. Moore (2012) “Spatio-temporal priority revisited: The role of feature identity and similarity for object correspondence in apparent motion.” *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), pp. 975–988.

Jiang, Hao (2020) “Effects of transient and nontransient changes of surface feature on object correspondence” *Perception* 49(4): pp. 452-567. DOI: 10.1177/0301006620913238.

Kahneman, Daniel, Anne Treisman & Brian J. Gibbs (1992) “The reviewing of object-files: object specific integration of information” *Cognitive Psychology* 24(2), pp. 175-219.

Kaplan, David M. (ed.) (2017) *Explanation and Integration in Mind and Brain Science*. New York: Oxford University Press.

Kosslyn, Stephen M. (1980) *Image and Mind*. Cambridge, MA: Harvard University Press.

Kosslyn, Stephen M. (1994) *Image and Brain*. Cambridge, MA: MIT Press.

Kuhn, Thomas S. (2012) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Ladyman, James, Don Ross, David Spurrett & John Collier (2007) *Every Thing Must Go*. New York: Oxford University Press.

Lepori, Michael, Serre, Thomas & Pavlick, Ellie (2024) “Break it down: Evidence for structural compositionality in neural networks” *Advances in Neural Information Processing Systems*.

Merullo, J., Eickhoff, C., Pavlick, E. (2024). Circuit component reuse across tasks in Transformer Language Models. ICLR.

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

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

Mitroff, Stephen R. & George A. Alvarez (2007) “Space and Time, Not Surface Features, Guide Object Persistence” *Psychonomic Bulletin & Review* 14(6), pp. 1199-1204.

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

Paivio, Allan (1990) *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Piccinini, Gualtiero (2020) *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford: Oxford University Press.

Pylyshyn, Zenon (2003) *Seeing and Visualizing*. Cambridge, MA: MIT Press.

Pylyshyn, Zenon (2007) *Things and places*. Cambridge, MA: MIT Press.

Pylyshyn, Zenon & Ron W. Storm (1988) “Tracking Multiple Independent Targets: Evidence for a Parallel Tracking Mechanism” *Spatial Vision* 3(2), pp. 179-197.

Quilty-Dunn, Jake (2016) “Iconicity and the Format of Perception” *Journal of Consciousness Studies* 23(3-4), pp. 255-263.

Quilty-Dunn, Jake, Nicolas Porot & Eric Mandelbaum (2023) “The best game in town: the reemergence of the language-of-thought hypothesis across the cognitive sciences” *Behavioral and Brain Sciences* 46(e261): pp. 1-75. doi:10.1017/S0140525X22002849.

Rubner, Andrew & Alfredo Vernazzani (ms) “A Bundle View of Perceptual Object Representations.”

Rumelhart, David E., McClelland, James & the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Scholl, Brian J. (2007) "Object persistence in Philosophy and Psychology." *Mind & Language* 22(5), pp. 563-591.

Scholl, Brian, Zenon Pylyshyn & Jacob Feldman (2001) "What is a visual object? Evidence from target merging in multiple object tracking." *Cognition* 80, pp. 159-177.

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

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

Smolensky, Paul (1987) "The constituent structure of connectionist mental states" *Southern Journal of Philosophy Supplement* 26, pp. 137-160.

Smolensky, Paul (1990) "Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems" *Artificial Intelligence* 46, pp. 159-216.

Sterelny, Kim (2010) "Minds: Extended or Scaffolded?" *Phenomenology and the Cognitive Sciences* 9(4), pp. 465-481.

Tremoulet, P., Leslie, A. M. & D.G. Hall, D. G. (2000). "Infant individuation and identification of objects." *Cognitive Development* 15, pp. 499–522.

Vernazzani, Alfredo (2022) "Do We See Facts?" *Mind & Language*. DOI: 10.1111/mila.12336

Vernazzani, Alfredo (ms) "Many Roads Lead to Scientific Understanding"

Wilson, Bret (2003). "The Artistic Tower of Babel: Inextricable Links between Culture and Graphic Development." *Visual Arts Research* 29(57), pp. 25-38.

Zhou, Ke, Huan Luo, Tiangang Zhou, Yan zhou & Lin Chen (2010) "Topological change disturbs object continuity in attentive tracking" *PNAS* 107(50), pp. 21920-21924.

---

[i] Other approaches to representational formats focus instead on types of contents (e.g., Peacocke 2019), or content-encoding schemes (e.g., Lee et al. 2023). We are open to the idea that different approaches to representational formats can be useful for different explanatory purposes. At any rate, we believe that the main point of this paper, namely the claim that there are serious risks with framing cognitive representational formats in terms of public representations, applies to those other approaches as well. Detailed defence of this is beyond the scope of this paper. We thank an anonymous referee for suggesting this clarification.

[ii] See Haziz (2023) for an overview of the epistemology of inquiry; and see Barack (2024) for a biologically-inspired perspective on scientific inquiry.

[iii] Quilty-Dunn (2016) is interested in what he dubs "Iconicity," whose formulation is very close to Fodor's Picture Principle.

[iv] An illustration of the role of socio-historical conventions in public formats is provided by Wilson (2003), which found that children's drawings vary significantly across cultural backgrounds. For instance, over 70% of childrens' drawings from Middle-East Islamic countries tend to represent human torsos with a rectangular shape, whereas only less than 4% of children's drawings from the USA, Australia, European countries, and New Guinea exhibited that feature. English children in the late 1800s tended to represent bottle-shaped bodies (Wilson 2003; Cohn 2013). According to Wilson (2003), this indicates that children's drawings are not directly drawn from perception, but rather rely on culture-specific schemata.

[v] Despite being one of the most forceful proponents of LOTH, Fodor accepted the existence of iconic representations, understood in terms of his Picture Principle mentioned above, even though ascribing to them rather limited roles in comparison with LOT representations. It is also worthwhile to point out that Fodor was rather sceptical of the capacity of LOTH to explain global, non-modular cognitive processes (see Fodor, 2001, 2008).

[vi] A recent defence of LOTH, instead, prefers to focus on empirical evidence, and downplays Fodor & Pylyshyn's (1988) compositionality-based arguments (Quilty-Dunn, Porot & Mendelbaum 2022).

[vii] This latter option makes recourse to the competence/performance distinction in defending LOTH border on being question-begging (cf. Fodor & Pylyshyn 1988; Quilty-Dunn, Porot & Mendelbaum 2022).

[viii] There is ongoing research on how current Large Language Models (LLMs) tackle tasks that involve systematicity and compositionality. Some of this incipient research indicates that LLMs may be fruitfully described, with some idealisation, as approximating some of those features by non-symbolic means (see, e.g., Merullo et al., 2024). This seems to be along the lines of Smolensky's (1987) suggestion that LOT may be a useful abstraction, but does not capture the nature of the non-discrete representations and non-symbolic computations at play in cognition. As such, LOT should rather be seen as a highly idealised scientific model of linguistic cognition, rather than as a hypothesis about the underlying computational-representational architecture. We thank an anonymous reviewer for pressing us on this point.

[ix] There are other live options that Fodor & Pylyshyn (1988) fail to consider: formats other than language, like maps and tree diagrams allow for compositionality (Camp 2009), so that the putative compositionality of thought can be accounted for by appeal to non-LOT representations; and the compositionality of human thought may depend on the 'cognitive scaffolding' enabled by learning natural languages (cf. Sterelny 2010). Moreover, Fodor & Pylyshyn seem to maintain that connectionist networks can implement and/or simulate compositional processes, as they agree that connectionist networks can implement LOT architectures, which are compositional. They seem not to appreciate the import of this point for the tenability of their arguments against connectionism, as Chalmers (1993) notices.

[x] Contemporary proponents of LOT have indeed abandoned both the earlier claims that LOT is the only cognitive format, and that it should be modelled on natural language (Quilty-Dunn, Porot & Mendelbaum 2022, pp. 2-3).

[xi] Green & Quilty-Dunn (2021) contend that a POR just is an object-file, while Block (2023) argues that PORs are different from object-files (see also Rubner & Vernazzani ms).

[xii] The Continuity Principle is here phrased in terms of properties, but one may phrase it in terms of property-instances or tropes.

[xiii] One may have other reasons to posit a demonstrative, however. For instance, Green & Quilty-Dunn (2021) contend that the demonstrative should be more appropriately understood as a pointer—a notion familiar from computational architectures—and that the pointer plays a twofold role in PORs: connecting an object file to the worldly target, and referring to information stored into a multiple-slot memory structure (see Rubner & Vernazzani ms).