

Article

Not peer-reviewed version

---

# Compositionality in Visual Representations and the Hardwiring of Our Visual System

---

[Athanasios Raftopoulos](#)\*

Posted Date: 24 January 2025

doi: 10.20944/preprints202501.1819.v1

Keywords: iconic representations; operational constraints in vision compositionality; mereotopology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

# Compositionality in Visual Representations and the Hardwiring of Our Visual System

Athanassios Raftopoulos

Professor of Epistemology and Cognitive Science, Department of Psychology, University of Cyprus, raftop@ucy.ac.cy

**Abstract:** It is widely accepted that visual representations are iconic and differ from the symbolic representations of propositional attitudes. Iconic representations compose differently from symbolic representations. Symbolic compositionality is canonical (it conforms to a set of rules determined by symbolic logic or by some grammar). Iconic representations display a whole/parts compositionality, in the way parts of objects combine to form whole objects. In this paper, I take recourse to Mereotopology as it applies to the compositionality of parts to form wholes to explain, first, the compositionality of icons, and to argue, second, that the hardwiring of our visual system is such as to reflect the basic compositional rules of Mereotopology.

**Keywords:** iconic representations; operational constraints in vision compositionality; mereotopology

---

## 1. Introduction

Cognitive and visual perceptual states are held by many researchers (Ayers 2019; Beck 2012, 2018, 2019; Burge 2010, 2018; Burnston 2017; Carey 2009; Crane 2003; Cummins and Roth 2012; Dretske 1981, 2000; Fodor 2007; Goodman 1976, 1978, 1984; Haugeland 1981, 1987; Heck 2007; Jackendoff 1987; Kosslyn 1994; Peacocke 1986, 2019; Sainsbury 2005) to be cast in different representational formats, namely, symbolic (digital) and iconic or analog formats respectively. Among those who hold that visual perception (or early vision for those of us who think that only early vision is unaffected by concepts and is a possible candidate for having purely iconic states) does not have symbolically structured format but an iconic or analog format, there is considerable disagreement as to the nature of iconicity in two respects.

First, among the researchers who hold this distinction there is talk of iconic or analog representations. The relation between iconic and analog representations is complicated and there exists a lot of confusion concerning these terms. For this reason, in this paper, I do not attempt to explicate the differences between iconic and analog representations and confine myself to discussing iconic representations for two reasons. First, those who endorse the view that visual representations are analog also hold that iconicity is one of the properties of analog representations. In other words, all analog representations are iconic (but not vice versa), which means that the traits of iconic representations are also traits of analog representations. Second, one of the problems I discuss in this paper concerns the compositionality of iconic representations and with respect to that problem, whether a representation is analog or iconic is immaterial since both types are thought in the literature to compose in the same way.

Second, there is disagreement concerning the definition of iconicity. Three main theses have been proposed in the history of the debate. (i) Iconic representations are those representations that physically resemble what they represent, i.e., their referent; (ii) Iconic representations are those representations that are context dependent; (iii) Iconic representations are those representations that map in an appropriate way on what they represent. The first and third views are the more dominant and I concentrate on these two when I discuss the definition of iconicity. The reader should notice, however, that under both views of iconicity, iconic representations compose the same way, and, thus,

which definition of iconicity one adopts is orthogonal to the problem of the compositionality of iconic representations.

In the first section, I examine the two main definitions of iconicity and explain why the mapping definition is prevalent among, at least, philosophers.

In section Two, I discuss the problem of the compositionality of iconic representations, which I distinguish from the compositionality of symbolic representations. I compare compositionality of iconic representations with the compositionality of symbolic representations, and argue that the rules that govern iconic compositionality are those of Mereotopology and not a set of rules determining syntax, as is the case with symbolic representations.

In the third, and last, section I concentrate on the compositionality of visual iconic representations and discuss the basic principles of Mereotopology that express the mathematical structure of the compositionality of visual representations. Then, I argue that iconic compositionality in vision is realized by neural mechanisms that are based on the synchronization of oscillating neurons and explain why, by discussing the synchronization of neurons involved in visual tasks. Finally, I propose the thesis that the assumptions on which our visual system operates so as to be able to represent the world, which in previous work (Author XXX) I have called “operational constraints”, express basic principles of Mereotopology that are hardwired in our visual perceptual system.

In the concluding discussion I stress the importance of the arguments in the other section of the brain to say a few words concerning the way in which the environment imposes its structure to our minds. This is important because arguments that support the view that our minds impose structure onto the world, which is correct to some extent, tend to overlook the fact that before that the world has imprinted in many respects its own structure in our minds, rendering them more adept at perceiving and understanding it.

## 2. Iconicity

### 2.1. *Iconicity as Resemblance*

The first construal of iconicity is that iconicity relies on resemblance. Peirce (1977; 1903), for example, among the first to study referential signs, writes that icons are non-arbitrary representational signs that resemble (are similar to) what they stand for or represent. They physically evoke through their resemblance what they refer to; icons physically resemble their referents. In icons there is a non-arbitrary, natural linkage between the form and the meaning, that is, between the iconic sign and its referent. Thus, syntax parallels semantics and, in fact, it is the referent that determines the form of the sign since the latter must resemble in some way the former. This resemblance ranges from strong visual resemblance to resemblance with respect to some abstract structure. This idea gains tract if one considers realistic pictorial representations, but, also cubic paintings; in both cases the drawing resembles what it represents but the resemblance ranges from faithful depiction to capturing abstract structure in the represented domain. The idea that underlying iconic representations is some sort of similarity between some aspects of the picture and the represented item that is experienced by the viewer, which entails that the representation resembles in some respects what it represents, is found in many accounts of pictorial or iconic representations.

If, arguably, the first iconic representations are the mental representations formed in visual perception, Peirce is right that icons resemble what they refer to because perception is designed to ensure that we extract from the environment and represent the information required for our successful interaction with it (evolution has seen to it), and this can be achieved only if perception depicts accurately with an acceptable margin of error the physical attributes in the environment with which we have interacted in our evolutionary past. Thus, iconic perceptual representations resemble in qualitative ways many of the properties (including spatial properties) of the objects in the visual scenes they represent, as these properties look from the perspective of the viewer. This view can be extended to other pictorial, external representations. According to Ayer (2019, 77-78), pictorial in general representations convey information about what is represented ‘by exploiting the content of

the experience of seeing it, that is, how the object or scene looks from a particular point of view'. This is a good example of what Peirce meant by writing that icons 'physically evoke through their resemblance what they refer to.'

Similarity is clearly stated in Kosslyn's (1994, 5) discussion of depictive (depictive representations are a subset of iconic representations) representations as a necessary condition for iconicity.

In a depictive representation, each part of an object is represented by a pattern of points, and the spatial relations among those patterns in the functional space correspond to the relations among the parts themselves. Depictive representations convey meaning via their *resemblance* (my emphasis) to an object, with parts of the representation corresponding to parts of the object.

Jackendoff (1987, 181-2) explicates "resemblance" in iconic representations as follows 'In a geometric representation [Jackendoff's term for 'iconic representations'] objects are necessarily represented in terms of their shapes and apparent sizes . . . In a geometric representation multiple objects under simultaneous consideration are necessarily spatially related in distance and orientation.' Jackendoff also points out that in geometrical representations distances can be measured directly off the image without any specification of units. To the extent that objects in iconic representations are necessarily represented in terms of their shapes and apparent sizes, it follows that the relations among aspects within the representation are not arbitrary but depend on the relations between the aspects of the visual scene onto which they map. In this sense, an iconic (or an analog in the minimal sense) representation may be said to approximate the representatum (I borrow the term 'approximate' from Peacocke (1986)).

Hyman (2006) and Hopkins (1995; 1998) make this proposal more specific. Hyman (2006, 148-149) thinks that a pictorial iconic representation represents some object by eliciting the recognition of this object on account of the fact that the picture allows viewers to get in touch with the occlusion shapes of the depicted objects through the shapes of the parts of the picture that purports to represent the object; 'If O is a depicted object and P is the smallest part of a picture that depicts O, the general principle can be stated as follows: the occlusion shape of O and the shape of P must be identical.' (Hyman 2006, 81). Hopkins (1995, 443) uses the notion of outline shape and argues that if a part P of a surface resembles an object O in outline shape, appropriate observers see O at P; thus, pictures are experienced as resembling things in outline shape to various degrees of determinateness.<sup>1</sup> The reader should be reminded, however, that the sort of similarity required by iconic representations in general may concern purely relational attributes. An arrangement of matches in sketch may represent a column of tanks if the spatial arrangement of the matches maps one to one (and in this sense it similar) to the spatial arrangement of the tanks. Since, in general, resemblance is always defined relative to some quality or set of qualities, if a picture P represents O with respect to quality Q, this means that P depicts or represents O as being Q, which, in the case of the tank column, means that the matches represent this column as having the spatial arrangements that the set of matches does.

The discussion on experienced similarities as underlying the property of iconic representations to depict their objects is in line with Ayers' (2019, 77-78) view that pictorial representations, which are a subset of iconic representations, convey information about what is represented 'by exploiting the content of the experience of seeing it, that is, how the object or scene looks from a particular point of view'. Almost fifty years earlier, Gombrich (1951, 220) had claimed that pictorial representations are records of visual experiences that act as substitutes for one's perceptual encounters with the depicted items. Many authors that work on imagistic representations agree with this view. Recognition accounts of pictorial representation, for example, hold that pictures represent their objects the way they do because they enable viewers to recognize the objects the pictures depict in a way similar to the one the viewers would use to recognize these same objects in ordinary perception; in other words, a picture P depicts an object O if P elicits viewers' capacities to recognize O. This is a clear case in which iconic representations convey information about their objects by exploiting the content of the visual perceptual experience, which renders them imagistic iconic representations. It is also true, however, that on several occasions one can see what a picture depicts without necessarily

being able to recognize the *depictum*. It has been proposed, therefore, that a picture P depicts O if P elicits in viewers a visual experience as of O (which differs from recognizing O). Thus, iconic representations are directly tied to the experience of seeing. Even in the case of the set of matches that represents iconically a column of tanks by preserving the spatial arrangement of the tanks, this spatial arrangement is visually perceived.

According to Carey (2009), in iconic representations their parts represent parts of what the representation as a whole represents, a condition that is analysed by Fodor (2007, 173) as the *Picture Principle*, which states that if P is a picture of X, then parts of P are pictures of parts of X. In all these views, iconic representations have a whole/part structure, that replaces the compositional/combinatorial structure of digital representations (I will return to the whole/part structure of iconic representations later). Quilty-Dunn (forthcoming, 4) calls it the 'Parts principle', according to which 'parts of icons correspond to parts of what they represent', whereas elsewhere (Quilty-Dunn 2016, 47) calls it 'Iconicity'.

Despite its initial appeal, the view that iconicity should be viewed as a physical resemblance between a sign and its referent has lost most of its tract and it is commonly held that resemblance is neither a sufficient nor a necessary condition for iconicity (Crane, 2003; Goodman 1972, 1976; Greenberg 2013; Newall 2011). Notice, however, that in the modern literature there is discussion of a kind of similarity different from "physical similarity" that characterized iconicity. In this view of similarity, a class of iconic representations resemble their referents only in the sense that there is a 2nd order structure mapping between the sign and the referent and there is no requirement that there be any physical resemblance (that is, any resemblances concerning their qualitative properties). This is a variation of Shepard's (1976) criterion of 2nd order isomorphism satisfied by iconic representations, according to which relations among elements of the computational processes in the represented domain should be mapped onto relations among elements of the physical processes in the representing system.

## 2.2. Iconicity as Mapping

Since physical resemblance lost its appeal as a necessary and sufficient condition for iconicity, most recent discussions define iconicity as some sort of appropriate mapping from the representation onto the represented entity. This mapping is usually described as an isomorphism or homomorphism between the representation and the represented domain. An isomorphism between two domains A and B holds in general if the structure of A corresponds to the structure of B, which means that the relations among the elements of domain A can be put in a one-to-one correspondence to relations among elements in the domain B (in case of a homomorphism the relation can be one-to-many). Mapping two structures does not require that the elements in the two domains be the same or similar, in which case the two structures share some abstract property (think of the points in the number line and the natural numbers the line represents). In other cases, the two structures may share elements, in which case the isomorphism also concerns concrete elements (think of visual perceptual representations). Although detailed accounts of the exact nature of the sort of mapping that renders a representation iconic differ from one author to another, as we shall see all accounts more or less agree on the following: an iconic representation (an individual representation and not a representational system) satisfies the following conditions:

(i) the representing symbol maps onto the entity it represents in a natural, non-arbitrary way; what "natural" means has many interpretations that we shall address when discussing the various accounts of iconicity or analogicity, with a special emphasis on what this natural, non-arbitrary is when it comes to perceptual iconic representations;

(ii) the spatial and semantic structure of the representing symbol maps onto, or corresponds to, the internal structure of the represented entity in such a way that the representation preserves (at least functionally) the spatial structure (that is, spatial relationships) in the represented domain. This means that the components of the representation are spatially arranged so that this arrangement mirrors the spatial arrangement of the components of the represented entity. For some

representations the semantic structure, that is, the semantic relationships among the parts of the represented entity (such as relations of size or orientation), in the represented domain are preserved in the representation. In other iconic representations, some of the surface properties of the components of the representation (such as shapes or colors) may correspond to the represented properties on the represented entity.

The second condition needs further elaboration. While mapping of spatial structure is clear, one need explain what is the semantic structure of the represented domain that the representation must preserve and why is the qualification “to a certain extent” is needed. By semantic structure I mean, first, the distribution of qualitative features of items in the represented domain. If the represented domain is a visual scene, these items are the objects in the scene and the qualitative features are their properties. A perceptual representation of this scene, if accurate, preserves this aspect of the semantic structure to a certain extent. This last qualification is needed for the following reasons. The representation represents only those properties that are suitable for the representing organism’s interaction with the environment. Humans and bats, for example, differ with respect to some of the properties of objects they extract from the environment. Moreover, a visual perceptual representation is from the perspective of the viewer and, thus, the properties represented are as perceived from that perspective. Thus, a visual representation of a visual scene from one perspective may be different from a representation of the same visual scene from another perspective. In addition, representations of the same scene, say drawings, may omit several properties in the scene depending on the intentions of the artist. Finally, other representations, say a sketch of the arrangement of the tanks in a tank column on a piece of paper may omit all the qualitative features of the tanks and preserve only their spatial arrangement with respect to each other and the elements of the terrain and, perhaps, the number of the tasks.

Semantic structure, second, comprises the relations among the features of the items in the represented domain, such as relative sizes, for instance. On certain occasions, the semantic structure also includes, third, the numerosity of the items in the represented domain, even in the form of a one-to-one correspondence between these items and the objects files that perceivers create for the items in a perceived visual scene. This makes sense in the case of perceptual representations, but, obviously, is irrelevant in other iconic representations that represent magnitudes, such as temperatures or time. Fourth, and this is as we shall see very important for characterizing the representation as iconic, the representation preserves the semantic structure of the represented domain in the sense that state transformations in the latter (say, an increase in size or in temperature) must be captured by analogous transformations in the representing medium (say, an increase in the height of mercury in regular thermometers), so that the mappings between elements in the representation (at the syntactic/vehicle level) and elements in the represented domain be preserved after the change in the represented domain. As we shall see, this demand imposes restrictions on the syntax, that is, the physical properties, of the representation, since they must allow state transformations that map onto transformations in the represented domain.

The reader should notice that the representations whose iconicity is discussed are individual representations, as opposed to representational systems. This is important, because isomorphism at the systemic level does not warrant the conclusion that the representations involved are iconic. Recall our discussion of the digital thermometer; it is a representational system whose structure is isomorphic to the structure of temperatures in the environment since it preserves the ordering and arithmetic relations among the temperatures. This isomorphism withstanding, the representational system is a digital representation. It is individual representations that must have the requisite structure to map appropriately onto their *representanta* so that they may be deemed to be iconic representations. Depending on whether the representation is physical or mental, this structure may be at the level of the representational vehicle or representational content. In physical representations that, as such, do not have representational contents and whose contents are the represented domains, the structure at issue concerns the representational vehicle. An iconic individual representational

vehicle (whether it be a map or a picture) is isomorphic to the entity it represents; Lyons (2022) calls it “vehicle isomorphism.”

A representation is isomorphic to what represents if both represents and representatum instantiate relational structures that are isomorphic to each other. The representation has a number of components and those components bear various relations to each other, and all these relations together specify a relational structure. A similar situation holds for the representatum. The representation and the represented entity are isomorphic if and only if their relational structures are isomorphic. Since it is not required that every part of the *representatum* has to map to every part of the representation, *or vice versa*, in order for *the* representation to represent the entity, a homomorphism may also suffice for iconicity. On certain occasions, both the representation and the representatum have common some of their elements (as in photos where the colors and shapes in the photo may be the colors in the pictured entity, but this is not required, as in black and white photos.

In mental representations, in order for the representation to be iconic in the mapping sense, it has to have the requisite structure both at the vehicle and representational content level. Both the representational vehicle and the representational content must map appropriately onto the represented domain. This means that in iconic mental representations both the vehicle and the content must have internal structure. I expand on this problem when I discuss the problem of the internal structure of perceptual iconic representations.

Independently of whether an iconic representation is physical or mental, the following are true. First, each individual representation maps naturally on its represented item. Second, a system of such representations forms a representational scheme that appropriately and naturally maps onto the represented domain that consists of the individual entities that the individual representations represent. A mental representational perceptual state maps onto a visual scene both at the level of representational vehicle and the level of representational content. It does that because both vehicle and content have an internal structure that is mappable onto the internal structure of the represented visual scene. In this vein, individual iconic perceptual representations map naturally map onto their represented entities (through, say, similarity of shapes or colors, etc.). Moreover, they form a system of iconic representations that iconically represents a worldly states of affairs, say a visual scene. Similarly, an amount of liquid (which is a representational vehicle) in an analog thermometer iconically represents a certain temperature because it is a magnitude that can be naturally mapped onto a temperature *qua* magnitude in the sense that when the temperature increases so does the amount of liquid in the thermometer in a systematic way. Moreover, a system of such vehicles, which is an analog thermometer, iconically represents the domain of temperatures of some entity.

It may seem to strain the meaning of iconic representation, but an individual magnitude is an iconic representation of some temperature. An analog thermometer is a system (and not a mere collection, which means that the components of the system are interrelated in a systematic way) of such individual representations. Being a representational system and not an individual representation, an analog thermometer does not represent anything but it is a representational determinable. The specific volumes of Hg comprised in the system represent specific temperatures; they are determinate representations. As I argued above, a specific volume of Hg is an iconic representational vehicle. The thermometer is an analog representational system on account of the fact that the individual representations of which it consists are iconic (remember that mappings and even isomorphisms at the scheme level do not suffice to establish the iconicity of the representation), and on account of the way the individual representations are interrelated in the representational system. This way ensures that the representational system is isomorphic to the domain of temperatures that it purports to represent.

The mapping account of iconicity presented in the preceding paragraphs is unproblematic with physical or external representations that are instantiated in some publicly accessible physical medium because such representations are spatially arrayed and, thus, have a spatial structure that can be mapped onto the spatial structure of the represented entity. Venn diagrams, for example, are semi-iconic representations of logical relationships among classes of objects, because by using them

one exploits the spatial relationships between shapes, in this case, circles, to represent logical relationships between classes of objects. Let me first explain the qualification “semi-iconic”. Venn diagrams include a non-iconic representative element, namely the shading of an area, which signifies that that area is empty. In order to be able to understand and use Venn diagrams to represent Aristotelian logic, one must know the convention that a shaded area is empty; this is a convention as there is no natural correspondence between a shaded area and an empty area and, thus, shading is not an iconic representation of emptiness. It follows that Venn diagrams are hybrid representations that involve both iconic and digital symbols. In fact, the shading of an area, or a cross that indicates that there exists an item in some area, are not the only digital elements in Venn diagrams; each circle must be accompanied by a label indicated the type of class the circle represents and this label bears a conventional relation to the indicated class.

Physical, non-mental iconic representation, such as photographs, pictures, paintings, etc., involve space essentially, since that they are themselves spatially arrayed. This does not mean, of course, that there is a spatial-structure preserving mapping from the representation to the represented entity concerning all spatial properties involved (objects in pictures, for example, are smaller than the objects in the pictured scene, and the distances between the objects in the picture are scaled down with respect to the actual distances), but, still, the spatial arrangement of the objects in the picture is the same (and does not merely correspond to) the spatial arrangement of the objects in the pictured scene. This means that the isomorphism involved is a “strong isomorphism” in the sense that some of the spatial relations among the components of a representation will not be merely isomorphic to the spatial relations among the components in the pictured scene, but the spatial relations among the components in the representation are literally shared with the components of the pictured scene; so, this not case of a mere one-to-one correspondence, but a case of sharing the same (spatial) property. Since physical representations are typed and individuated only by their vehicles, it follows that being spatially arrayed is a property of representational vehicles. Significantly, this also applies to neural states, which are the vehicles of mental representations, although the spatial relations involved are, as we shall see in the next paragraphs, functionally defined.

However, the spatial-mapping condition in the definition of iconicity poses problems for internal, mental representations, such as perceptual representations because these representations purport to represent spatial arrangements (such as visual scenes) and, yet, they are not spatially arrayed, being mental entities. We will see how theoreticians attempt to account for, and surmount, this problem.

The first clause of the mapping condition requires that the mapping between the symbol and the represented entity be *natural*. Although I shall explain in which way the mapping is natural with respect to visual perceptual representations, let me say here that, in general, a mapping is natural if understanding the representation does not rely on some set of conventional rules that are learned and are, thus, the result of enculturation (such as linguistic systems), but on ‘the degree to which human nature—including relatively universal aspects of cognition, physiology, social behavior, and environmental interaction—rather than enculturation, makes that system easy to internalize and use.’ (Giardino & Greenberg 2015, 8)

When, later on, I explain the naturalness of perceptual iconic mappings, I base my arguments on the physiology of the human visual perceptual system and the interaction with our environment to claim that the mappings between the constructs of our perceptual systems and the visual scenes in the world are natural because of the aforementioned two factors, and do not rely on any conventions. In this case, the mapping between a perceptual iconic representation and the represented visual scene is natural because it does not rely on some learned convention but, rather, on the way our perceptual systems are structured and on the interactions with our environment.

Some iconic representations, however, do require some sort of enculturation in order to be understood and used. A sign with an airplane on the road may signify the direction to the airport, but this is understood only by those viewers who have experience of airplanes or have learned about airplanes even though they have never seen one. The relation between the sign and airplanes is

natural in that the sign depicts the shape of an airplane, while, in contrast, the term 'airplane' is conventionally related to airplanes, but understanding and, hence, using, the sign is learned. It is true that by depicting an airplane, or, rather, the shape of a typical airplane, the sign is easy to internalize and use. Accounts of iconic representations that are based on some notion of a natural mapping between the representation and the representatum, therefore, should consider the role of convention in establishing the correspondence between the sign and the signified item in parallel with the role of enculturation in understanding and using the sign. In view of the trade-off between "naturalness" and "enculturation" one should talk of degrees of iconicity; some representations are more iconic than others. In view of the fact that several iconic representations, such as diagrams, are in reality hybrid representations because they necessarily involve digital information elements to denote relations, types of magnitudes etc., it makes sense to claim that a diagram with less digital elements is more iconic than a diagram with more digital components. Talk of degrees of iconicity, on the other hand, does not make sense in the case of pure iconic representations where all information is conveyed solely by the mapping of the representational vehicle and content onto the represented domain.

Returning to the view that a representation is iconic owing to the way it maps onto its represented domain, a way to discriminate between iconic representations and notational representations, such as sentential representations, which is based on the mapping of the representational scheme onto the represented domain is offered by Haugeland (1987, 88-91). Haugeland assesses Palmer's (1978) distinction between notational systems (such as set of propositions) and iconic representations, according to which iconic representations represent properties and relations intrinsically, whereas digital systems represent them extrinsically. For example, in a digital representation of a domain in which an object is taller than another, the relation 'taller' must be explicitly represented by a distinct symbol. Imagistic representations (the term Haugeland uses for pictorial or quasi-pictorial representations), on the other hand, just show this relation and they do not need to represent it by importing an extrinsic element, such as a digital symbol. The reason is that in imagistic representations 'the representing relation has the same inherent constraints as the represented relation. That is, the logical structure required of the representing relation is intrinsic to the relation itself rather than imposed from outside.' (Palmer 1978, 271) Extending Palmer's views, one could say that the intrinsic logical structure of the representing relations in iconic representations, which mirrors the structure of the represented relations, entails that all information in the represented structure is explicitly shown in the representation. This is not the case with quasi-linguistic, sentential representations in which the fact that relations are represented extrinsically entails that there is a clear distinction between what is represented explicitly and what is represented implicitly, since inference is needed for that information to be extracted from the explicitly represented information.

Consider a map that depicts the spatial relations among three cities A, B, and, C and shows their distance in terms of magnitudes. Now think of a set of sentences that express the same information, namely, A is x miles north of B and y miles northwest of C. In this system, the relation between B and C can also be determined but this would take some geometrical calculations and in this sense the relation between B and C is implicit in the system and must be extracted from the explicit information that A is x miles north of B and y miles northwest of C. In the map, however, no calculations are necessary; the relation between B and C is shown as a result of representing in the map the relation between A and B, and A and C. The distinction between explicit and implicit representations in the map collapses because just looking at the map one cannot tell which are the original representations and which is the resulting one. Haugeland (1987, 90) calls this property of imagistic representations 'complicity'.

Heck (2007, 127) points to the same direction noting that in iconic representations a change in the representation is automatically a change in what is represented, and that the effects of changes in the represented domain automatically show as changes in the representation, whereas in propositionally structured representations one has to make these changes manually. Underlying this

difference is the fact that in a propositional representation of a map one first forms the propositions that describe parts of the map, by assigning properties and objects that the map purports to represent to specified locations in the map that denote locations of the mapped entity, and then one conjoins these propositions to get the propositional representation of the entire map. The conjunction of the propositions, however, fails to show the spatial relations among the assigned properties and objects. These relations must be added by using additional representational resources and by inferential derivation. Rescola (2009) calls the process by means of which propositional representations of maps are constructed 'predicative analysis' and uses the failure of propositional representations to show automatically the spatial relations among the elements of the map to argue that maps do not have predicative structure. Kulvicki (2015) thinks that maps predicate properties and objects to locations, but maps organize those predicates in a fundamentally different way than sentential vehicles, partly because maps exploit spatial structure and, thus, show the spatial relations among the representational elements of the map and there is no need to add distinct representations of these relations, as the predicative analysis of propositional representations of maps requires. I revisit the distinction between predication in digital representations and iconic representations later.

The map in the example that Haugeland uses is iconic because it satisfies the mapping or mirroring criterion, since it mirrors what it represents in that it bears a structure-preserving relation toward what it represents, which in our case is the spatial structure of the cities A, B, C. Haugeland shows that a representation such as a map that bears a structure preserving relation to the represented domain differs from a series of descriptions that expresses the structure preserving relation because the representational format of the map allows it to update automatically its structure when a change occurs. Rescola (2009 197-198), echoing Palmer (1978,) relates the property of automatic updating to the fact that iconic representations show directly the relations between the elements of the representation and do not need a distinct representational symbol to represent the relation as propositional representations do; 'A map's geometric structure is not just another element to be listed alongside its markers and coordinates . . . the markers and coordinates *stand in* [italics in the text] geometric relations, and those relations have representational import.' So, when these relations change in the map, which is the representational vehicle and, thus, these changes are syntactic changes, they automatically correspond to the relevant change to how the mapped entity is represented as being. Therefore, there is no reason to perform some sort of inference to arrive at the conclusion that describes the change occurred. As Shin (2003) puts it, an active inference from premises to a conclusion in a predicate calculus is transformed to a "free ride" in iconic representations such as maps.

Kulvicki (2006, 2014) also holds the view that imagistic representations are those representations that preserve the structure of what they represent. In this sense, they present their *representatum* to us in manifesting aspects of their *representata* structure. He notes that structure-preservation differs from resemblance, even though those representations that systematically resemble in some respect their *representata* make use of isomorphism (Kulvicki 2014, 105), because structure preservation could concern relations between represented elements and not their qualities, which is along the lines of Shepard's (1976) criterion of 2nd order isomorphism required in analog representational systems. According to it, relations among elements of the computational processes in the represented domain should be mapped onto relations among elements of the physical processes in the representing system, that is, that there must be an isomorphism or a homomorphism between syntax and semantics. We find again the idea that since iconic representations map appropriately onto their represented domain, they must necessarily depict, or correspond to, at least some aspects of this domain. This means that the form of the iconic representation must allow the representation to depict these aspects, for example, shapes and apparent sizes, and this demand put restrictions on the form of the representation; semantics constrains syntax. In contrast, in digital representations where digital symbols refer by convention and need not bear any relation to their *representata*, semantics puts no constraints on syntactic forms.

### 3. Compositionality of Symbolic and Iconic Representations

I shall argue that iconic representations do not have the canonical compositionality of symbolic, discursive representations, but are characterized by a part/whole compositionality. Let us start by briefly discussing the canonical or discursive compositionality of symbolic representations to contrast it to the compositionality of iconic representations.

#### 3.1. Compositionality of Symbolic Representations

Initially, discussions about symbolic representations, following Goodman's (1976, 1978, 1984) pioneering work, focused mainly on discreteness and reference by convention as the defining properties of symbols. However, the last twenty years or so there has been a shift of focus to the compositionality of the class of representations that are symbolic, with the main line of thought being that a representational system is symbolic if and only if its representations compose by means of the combination of their parts, which conforms to the principle of "semantic constituency", and the composition is discursive. This means that the composition of symbols is governed by a set of rules that constitute a syntax. Complex symbols are, thus, syntactically structured. A clear example of this is the various natural languages and symbolic logic. In both cases, complex symbols are formed by putting together individual symbols following the rules of the grammar of the language, or the rules of symbolic logic. This sort of compositionality is discursive since a complex symbol thus formed may be used as an input symbol for the formation of a more complex symbol.

Haugeland (1998, 174) defines symbolic, or logical as he calls them, representations those that have a discursive compositional semantics in the sense that complex digital structures have a recursively specifiable syntax and determinate atomic constituents. This means that the semantic significance of the whole is determined by its syntax and the semantic significances of its constituents; this type of semantics is called compositional semantics. The semantic significances of the constituents are fixed arbitrarily in the representational scheme, as are the rules of the syntax (they depend on the language at hand).

According to Fodor (2007), who echoes Haugeland (1998), symbolic, representations are discursive, logical, or, equivalently, propositional representations, that are compositional in a logical sense. A representation is *syntactically compositional* if and only if its syntactic analysis is exhaustively determined by the grammar of the relevant language together with the syntactic analyses of its lexical primitives. In other words, a representation is syntactically compositional if its formation follows the rules of some formal logic. A representation is *semantically compositional* if and only if its semantic interpretation is exhaustively determined by its syntax together with the semantic interpretations of its lexical primitives. Differently put, the meaning of a complex term is a syntax-dependent function of the meanings of its syntactic parts. More recently, Planer (2021, 237) expresses this as follows:

On this view, to token a thought is to token a mental expression—a representation couched in "mentalese"—where the content of the whole is determined by the content of the expression's parts, together with their mode of composition.

Assuming that one is concerned with formal systems or with languages that allow for no ambiguity, metaphors etc., in which semantic compositionality holds, the semantic compositionality of symbolic representations entails that their truth values are determined by the way their constituents are combined together according to their syntax, and the truth values of the constituents. This is usually defined as the "principle of semantic constituency". Some philosophers, following Fodor, hold that a representation is digital iff it satisfies this principle (Werning 2012). The principle states that if a representational vehicle  $s$  is a syntactic part of a representational vehicle  $p$  (representational vehicles are situated at the syntactic level of analysis of a representation), then the meaning of  $s$  is a semantic part of the meaning of  $p$ . It follows that the compositional structure on the syntactic level is mirrored on the semantic level, the famous Fodorian parallelism between syntax and semantics that underscores all discussions of the *Language of Thought* (LOT).

Having discursive or logical syntactic structure means that some parts of the digital representation are constituents and others parts are not, depending on whether the decomposition that yields the parts follows the rules of the system. Since complex discursive symbolic

representations consist of distinguishable discrete parts that are composed according to some rules of good formation, as dictated by some formal logic or some set of rules, for any decomposition to yield constituents, it must follow the rules. ' $\Phi$ ' is a constituent of the representation ' $\Phi(a)$ ' but ' $\Phi$ ' ('is not a constituent. This is so for two reasons, one pertaining to syntax, and the other to semantics. ' $\Phi$ ' ('is not a constituent of ' $\Phi(a)$ ' because it is not formed according to the rules of good formation as determined by formal logic (which dictate that there should be no unmatched parentheses). ' $\Phi$ ' ('is not a constituent because it does not have a semantic interpretation that contributes to the semantic interpretation of ' $\Phi(a)$ '). This shows that discursive representations are subject to canonical decomposition.

All these have repercussions concerning the compositionality afforded by symbolic representations, but, most importantly, they mark some basic distinctions between symbolic and iconic representations. First, through the use of discrete symbols, any discursive representation can be analyzed into distinct parts each of which can represent a unique object or property without representing anything else. Through concatenation, such representations can be combined to represent, say, a whole visual scene. As we shall see, iconic representations do not have such a property, although they consist of parts too. Second, symbolic but not iconic representations support logical connectives.

Let me close the discussion of the compositionality of symbolic representations by returning to mental propositional representations to highlight their main characteristics as they emerge from our discussion thus far, adding a new twist. Mental representational states have representational contents. Those mental states that are sentences or sentence-like entities have propositional contents that form propositional structures. Propositional structures are complexes of representational contents that are formed by means of various combinations of atomic propositional units that are performed by means of specific compositional mechanisms. The simplest and most fundamental such mechanism is predication, where an unsaturated, predicative part of thought is completed by a term that denotes some entity in the world (such as a proper name, for example) and becomes a complete part of thought, that is, a proposition that has some truth value (Frege 1892, 54). (Recall that according to Frege predicates or concepts are functions that take as inputs objects or denoting terms and output a truth-value). Rescola (2009, 177) defines predication as 'a compositional mechanism whereby terms fill the argument-places of a predicate that carries their denotation into a truth-value.' For example, the proper names "John" and "Mary" are denoting terms that when they fill the two argument-places of the 2nd order (relational) predicate "Love" form the complete term or proposition 'John loves Mary', which is true or false depending on whether John loves Mary or not.

Propositional structures are, thus, characterized by the kinds of operations that can be performed on them. Camp (2018, 25-26) argues that the operations that underwrite propositional structures share four basic features.

1. The operations are highly digital, which means that they take a small number of discrete elements as inputs;
2. They are universal, or highly general, in that they can combine a wide variety of inputs. Predication, for instance, is a highly general operation taking many kinds of objects and predicates and delivering truth-values;
3. The combinatorial operations are asymmetrical. This means that one element must be of a type that enables it to serve as input for the other; only a denoting term can fill the argument-place of a predicate but not the other way around, and this makes the predicate and the denoting term asymmetrical. This asymmetry means that the same terms combined differently yield different propositions. "John" "Loves", and "Mary" may be combined to yield 'John loves Mary', or 'Mary loves John'. The head of a phrase and its appendage constitute another asymmetry. The grammatical ending "d" may be used differently with two content words and depending on which is used as an ending creates a hierarchy between them, making the one the head and the other the appendage. For example, a principled compromise is not a compromised principle because the heads and the appendages (the terms to which 'd' is post-fixed) in the two phrases

are different; “d”s use creates a hierarchy. Finally, hierarchical structure makes embeddings possible (embedding is placing one element inside another). For example, one may take ‘my mother’ and place it inside the phrase ‘my mother’s second cousin’.

4. The operations are recursive, which means that its outputs can serve as inputs to the same operation. Differently put, one can have sentences or phrases within sentences or phrases, or, in case of cognitive abilities in general, one can have thoughts within thoughts and, thus, one can think thoughts about thoughts. The recursive application of an asymmetrical relation produces an intricate hierarchical structure that contains nested iterations of the same type of representation. In linguistics, examples of propositions with nested clauses that exemplify the recursiveness of languages abound. For many researchers, deep hierarchical recursion is a defining trait of propositional structures.

In sum, advanced linguistic propositional structures, or formally defined systems are digital, asymmetrical, hierarchical, and recursive. In these systems, syntax has the upper hand in determining meaning, since, given the meanings of the constituents of a complex representation, it is the way these constituents are put together, that is, the syntactic rules, that determine the meaning of the whole. In other words, grammar acts like a semantic filter that guides the interpretation of the complex representation (Chomsky 1965). I think this analysis of propositional structure is partly successful because all other formal properties of propositional structures that we have discussed thus far are made possible by operations that satisfy the preceding conditions. Systematicity, combinability, digitalness, discreteness, all derive from the properties of the operations that underwrite propositional representations.

One should note the emphasis on the formal structure of propositional representational structures. Despite the fact that propositions are the representational contents of sentential-like representational vehicles, in characterizing propositional structures emphasis is put in the formal properties of the structures, how they do compose, for example, rather than in the semantics of the propositions. Of course, propositions, *qua* representational contents, are abstract entities and, thus, operations do not apply to them but to the representational vehicles that carry them. Still, these operations impose a certain formal structure on the propositional structure, a structure that endows these propositional systems with systematicity, digitalness, discreteness, and combinability, all characteristics of thought.

### 3.2. Compositionality of Iconic Representations

Let us return to the problem of the compositionality of iconic representations and especially of visual iconic representations. If iconic visual perception purports to represent the environment, the accuracy conditions of visual representations are fixed by the way the properties and objects in the represented scene are bound together to form the parts of the visual representation of the scene, and by the way the parts compose to form the whole scene. The accuracy conditions of a visual perceptual representation are determined by the accuracy of the representation of its parts and by the accuracy of the way these parts compose to form the whole representation. The problem is that if this composition is not discursively determined by some syntax, as in the case of symbolic representations, one should explain how the parts of iconic representations compose to form the whole. Merely conjoining parts does not suffice to yield the whole, because the atomic parts must be placed together in a way that the spatial and semantic relations in the represented scene be preserved in the representation. To answer this question, one must turn, I argue, to the mathematical theory of Mereotopology, because iconic representations do not compose discursively but holistically, that is, they display a parts/whole compositionality.

Let us start by discussing the structure of the content of the visual perceptual representations of objects and of their properties in a visual scene. Since this content consists in objects and their properties arrayed in space, one could render this representational content in the form of a set of subject-predicate structures. If this is correct, it follows that visual representations have a rich semantic structure that is to a certain extent similar to the subject-predicate structure of linguistic

representation, or of all representations. I refer to this structure as the basic relation R of noun phrases. Let me elaborate here on R. The rich semantic structure of an iconic representation of a visual scene (be it a perceptual representation or a picture or a drawing of the scene insofar as the latter exploits perceptual capacities to interpret the scene) consists in the compilation of various R-relations binding objects to predicates, a compilation that composes the whole out of its parts (the R-relations). As I argue in the next section, the compositionality involved in visual iconic representations is not the discursive compositionality grounded in hierarchical relations and recursive structures that characterizes symbolic representations, but the compositionality befitting part-whole structures that involve different kinds of restrictions as to what counts as a proper part of the representation, and engage different combinatory mechanisms that underwrite the compilation of parts to form wholes, mechanisms described by Mereotopology.

To put it differently, what counts as a well-formed construction in symbolic representations is determined by the formal rules of the relevant grammar; the restrictions are at the syntactic level, and they fix in turn the semantics of the representation, that is, they determine which parts are meaningful or not. Things are different in iconic representations, especially visual icons, where what counts as a well-formed part is determined by semantics and pragmatics, the latter in the form of the operational constraints at work in perception, and not by some syntax. While in symbolic representations the compilation of parts to form wholes is fixed by a set of formal rules, in iconic representations, the compilation of parts to form a whole is not determined syntactically. As we shall see when we discuss Mereotopology, what counts in iconic representations as a well-formed, natural whole is determined geometrically and not formally. The compilation, thus, is not the discursively analyzed composition of symbolic representations.

All these have implications for the nature of R-relation as it is figures in iconic representations and distinguish it from the R-relation in symbolic representations. First, predication works differently in the two sorts of representation. Predication is a combinatorial operation or compositional mechanism that connects predicates to either subjects in thought/language, or to objects in perception. The objects fill the argument-places of a predicate to form a full or saturated expression that has a truth-value. In a propositional representation of a visual scene, one forms atomic sentences that have the R structure that represent parts of the scene, and conjoins them to form the representation of the whole scene. The construction of such "atomic" sentences is done as follows: one identifies an object and predicates to it the predicates that represent the properties of the object in the visual scene. Then, one conjoins the atomic propositions to get the propositional representation of the entire scene. In this construction, the spatial relations between the objects in the scene are not represented directly, but they can be derived if the properties predicated to the objects include spatial information. This entails that changes in the spatial relations in the represented visual scene are not automatically instantiated in the sentential vehicles of the symbolic representation. In an iconic representation, in contrast, the spatial relations among the objects in the scene are instantiated in the iconic representation (both vehicle and content in perceptual representations, and in the vehicle in maps, pictures or drawings), so that any spatial change in the scene automatically updates the representation.

It follows that iconic representations organize their predicative structures (the R-relation) in fundamentally different ways from sentential representations, exactly because they exploit spatial structure and place objects and properties at locations. Not only in the sense that in an iconic representation an object and its properties as they are perceptually represented are tied to a specific location (and one could add time), which explains the well-known situatedness of visual representations and ties them to a specific spatio-temporal context, but also in the sense that spatial information is primarily used to individuate objects in the scene (Author) and is part and parcel of the visual representations of the object. In addition, iconic representations do not allow the hierarchical, asymmetrical structure of propositional representations; there are, for example, no nested perceptual representations within other perceptual representations. In this specific sense, representations have a flat structure (Camp, 2018; Deutscher, 2005; Everett, 2017). This means that

the combinatory mechanisms binding atomic predicative representational components differ between iconic and symbolic representations. In symbolic representations there is a digital, hierarchical, universal, and asymmetrical combinatory mechanism, while in iconic representations the mechanism is holistic, spatially bound, and allows automatic updating in the representational vehicle when spatial relations in the represented domain change (Camp 2018, 41).

Sterelny (2020, 778), starting from a distinct perspective reaches the same conclusion, arguing that

Iconic representations, such, maps, photos, diagrams, and the like can convey a lot of complex information asynchronously and in a more or less language-independent ways. Moreover, in contrast to other powerful graphic codes, the expressive power of these depictions does not depend on the recombination and recursive interpretation of a much smaller number of standardized signals.

Let me emphasize that in the preceding paragraph I wrote 'in this specific sense', because there is another sense of "flat" that does not apply to iconic representations. There are wholes that are flat in the sense that they have no internal organization. Mere sums of elements are a typical example. There are also whole that have an internal, syntactic or semantic, structure and, thus, there is a division into levels of organization. Iconic representations such as images of any kind are such examples. The representation of a house clearly consists of parts that are hierarchically organized into parts of parts, etc.

The semantic interpretation of an iconic representation of a visual scene is exhaustively determined by its whole/part relations together with the semantic interpretations of its parts, that is, the objects and properties to which the analogues in perception of the singular terms in perceptual contents and the attributives in the representation refer. Thus, visual iconic representations are semantically structured. They have the syntactic structure of the basic relation R cast not in a discursive form but in a whole/part form that constraints the sorts of relations that hold between the syntactic parts of the representations.

Let me elaborate on the semantic structure of visual representations. According to Crane (1992), even though the nonconceptual content of visual perception, which is iconic, does not have the structure of judgeable content, it still represents a manifold of objects, properties and events. Peacocke (2001, 241) argues that the nonconceptual content of perceptual experience represents things, events, or places and times in a certain way, as having certain properties or standing in certain relations, "also given in a certain way." I have argued (Author) that early vision, the stage of perception whose states have nonconceptual, purely iconic contents, represents a structural description of the visual scene that contains information about 3D shapes as viewed from the perceiver, spatio-temporal and surface properties, color, texture, orientation, and motion of objects, in addition to the representations of objects as bounded, solid entities that persist in space and time. The numerosity of objects is also represented. The contents of early vision are context-bound and demonstratively applied (Author) in the sense that these contents are inextricably tied to the context of the perceptual encounter and to their *de re* application to the objects and properties in the visual scene.

In this vein, Burge (2014, 574-575) argues for the existence of 'preconceptual cognitive representations', where 'pre-conceptual' intends to convey that these representations are not propositionally structured, and also claims that these pre-cognitive, perceptual as we shall see, representations have the same structure as noun phrases, the difference being that they do not employ concepts as linguistic noun phrases do.

Another type of (cognitive) pre-conceptual representation is formed through learning or other processing in long-term memory (modal or amodal). Even certain operations on working visual memory count as cognitive. Like perceptual-level representation, the foregoing cognitive representations are pre-conceptual in lacking propositional representational format.

These pre-conceptual cognitive representations are perceptual because

Like perception, these types of pre-conceptual cognitive representation have the same structure as noun phrases constituted of contextual-determiner-dominated attributives—the structure of that

F or those Fs. When representation occurs, the representational types are applied in a demonstrative-like manner. In all perceptual-level and most pre-conceptual cognitive level representations, such determiner-governed attributions are part of a complex iconic array. Visual perception consists in a rich, topographical array of demonstratively applied attributives, at various levels of specificity.

On a more general note, and returning to the idea that iconic representations have the syntactic structure of the basic R relation that they share with noun phrases, and thus, that iconic representations can express semantic relations that R makes possible, it should be added that visual iconic representations in particular and iconic representations in general share two characteristics with linguistic representations that use symbols, on account of the fact that they both exemplify the basic R relation. The first is that iconic representations are constructed on the basis of a set of construction rules. This is readily shown in the case of visual perception; the perceptual system constructs the percept, which is an iconic representation at the level of representational content, first, on the basis of a set of transformations that gradually transform the light pattern impinging onto the retina to the full-blown percept. Marr (1982), for example, was the first to give an account of such transformations from light patterns to primal sketches, 2D, and finally 3D visual representations. This set is essentially a set of mathematical transformations that allow the perceptual system to extract from the pattern of light that is the input to the visual system the spatial and semantic configuration of the entities in the visual scene from which the light emanates.

This does not suffice, however, for the construction of the percept. As many have argued, these transformations must be supplemented by a second set of rules that express both the geometry of the environmental space and the basic physical regularities of the interaction between light and our perceptual systems, which I have called operational constraints (Author). I discuss these constraint in detail in the next section. These two sets of rules that operate in, and render possible, visual perception are construction rules that parallel the syntactic rules of linguistic systems with the major difference, however, that the latter are conventionally established and culturally learned systems, while the former are neither of these. Thus, visual representations are syntactically decomposable, albeit in a non-discursive way, and have a syntactic, albeit non-discursive structure. It is important to note that the operational constraints function by restricting the solutions available to perceptual processing, solving thus the problem of the underdetermination of perceptual processing from the input. Their role is to ensure that perception retrieves from the environment and represents information that is useful for the organism in its interactions with the environment. In other words, these rules contribute to retrieving from the environment and representing (when representations are required) information that is meaningful for the organism, by hardwiring in the perceptual system of the organism the physical and geometrical regularities in the environment that are useful for its survival. In this sense, it is semantics that drives the form of the representation rather than the other way around as in digital representations.

In this section, we discussed the whole/part compositionality of iconic representations. The reason that iconic representations do not have propositional contents is that they, unlike propositional or symbolic in general, representations do not compose discursively (Camps 2018; Crane 2003; Fodor 2007; Haugeland 1987; Heck 2007; Kulvicki 2006, 2014, 2015). Quilty-Dunn (2020, 4) puts it a similar way: 'iconic representations acquire accuracy conditions from the way features are holistically bound in each part, together with the spatial arrangement of those parts.' This means that the semantic interpretation of the representation of a visual scene is exhaustively determined by its whole/part relations together with the semantic interpretations of its primitives, that is, the objects and properties to which the analogues of singular terms in perceptual contents (the parts of perceptual content that stand for the linguistic 'That X', which refer directly to the objects individuated in the visual scene), and the attributives in the representation refer. It follows that perceptual, iconic representations are semantically structured. They are also syntactically structured, although not in a discursive way; they have the syntactic structure of the basic relation R cast not in a discursive form but in a whole/part form that constraints the sorts of relations that hold between the syntactic parts of the representations through the Mathematical rules of Mereotopology. It is this

structure that enables the natural mapping between the components of the visual representation and the entities and relations in the visual scene.

#### **4. Whole/Part Compositionality in Vision, Mereotopology, and Operational Constraints in Vision**

I return to elaborate on the whole/part compositionality of visual iconic representations, starting from an example of what I mean by part/whole compositionality, as it reveals itself in the phenomenology of visual perception. 2-D surfaces are composed of edges that, in turn, are composed of line-segments. When one perceives in realistic conditions a line segment, one sees it as part of an edge, which is also a part of a surface, despite the fact that the perceptual system constructs first line segments, then edges, and then 2-D surfaces. In other words, segments of an edge are perceived by perceiving the edge and, thus, are perceived as edge parts and not as independent perceptual units. The same holds for properties such as size and brightness. Burns (1987) shows that size and brightness are not encoded at first as independent features of objects but are perceived in terms of holistic objects.

This bears directly on why iconic symbolic representations having whole/parts compositionality do not satisfy the principle of semantic constituency that characterizes, or even defines, symbolic representations. According to this principle, if a representational vehicle  $s$  is a syntactic part of a representational vehicle  $p$ , the meaning of  $s$  is a semantic part of the meaning of  $p$ . Even though syntactically line segments are parts of edges, which in turn are parts of 2D surfaces, because this is how the neural system constructs the representations, at the semantic, phenomenological level, line segments are not perceived as semantic parts of edges in the sense that to perceive phenomenologically an edge one need not phenomenologically perceive a line segment; in fact, one does not. To put it metaphorically in conceptual terms, one understands an edge without understanding line segments, even though the latter are syntactic constituents of the former. As a matter of course, iconic representations satisfy semantic constituency at some level. If iconic representations bind holistically parts together, some of these parts, say the objects in a visual scene, are individuated as distinct patterns in the representation, and the meaning of the whole representation depends on the meaning of these parts.

The difference of semantic constituency to the extent that it applies to iconic representations from the same principle as it applies to symbolic representations lies in the fact that in iconic representations the meaning of the whole is not determined by combining parts at the syntactic level following the rules of a formal syntax, as it happens in symbolic representations. In symbolic representations, one composes syntactically well-formed expressions following a set of syntactic rules and it is the syntactic composition along the semantics of the atomic parts that determines the meaning of the whole. In iconic representations, in contrast, one does not rely on some formal rules to compile parts to form a whole. Visual iconic representations, for example, are constructed by the process of the visual system and the visual processes of this system that form parts and combine them to form the whole (that is, the percept) are not guided by, and do not underwrite a, set of formal rules but are driven by the stimulus, and are guided by a set of constraints that express the physical and geometrical regularities of the environment in which the perceiving organism lives. In this sense, it is the semantics and pragmatics of the system formed by the representation and the represented entity that guide the formation of the whole, rather than syntax.

A characteristic of many iconic representations is the use of space or some geometrical structure in representing the representatum; this spatial or geometrical structure maps onto the spatial structure of the representatum. Visual representations, a clear-cut case of iconic representations, have a two-dimensional or three-dimensional form. This is the spatial form that grounds the iconic structure of the perceptual representation; spatial representations preserve the spatial relationships of the represented entities owing to the mapping of the spatial relationships between the elements of the representation onto those in the representatum. Visual perceptual representations, being spatial representations, inherit this property and, thus, preserve the spatial relationships of the entities (be it

objects or properties) in the visual scene. This is, of course, one aspect of the iconicity of visual representations, the other one being that some of the elements of the representational content of the visual representation map into the visible features of the objects in the represented visual scene.

The whole/part compositionality constrains the way the parts of an iconic representation are put together to form the whole: 'iconic representations acquire accuracy conditions from the way features are holistically bound in each part, together with the spatial arrangement of those parts.' (Quilty-Dunn, 2020). I said above that line segments are put together to form edges, which in turn, put together in appropriate ways, form 2-D surfaces. Moreover, one does not compile a surface from edges by following some logical, or syntactic in general, rules of construction of complex representations. Owing to the whole/part compositionality of the representation, the parts are perceived as parts of the whole rather than as distinct entities; one does not perceive line segments as distinct entities but as parts of edges. In discursive representations such as ' $(\Phi a)$ ', in contrast, ' $\Phi$ ' retains its autonomy and can be seen independently of the whole ( $\Phi a$ ); this is the result of the compositionality of discursive representations.

The part/whole compositionality seems to entail that since line segments are parts of edges and edges are parts of 2-D surfaces, the line segments are parts of the 2-D surface, a conclusion that seems warranted. However, even though the handles of a door are parts of a door, and the door is a part of a house, the handles are not properly speaking parts of the house? How is the latter case different from the former? The answer to questions like these should be searched in the mathematical models of Mereology and Topology, the two disciplines that examine part/whole compositionality.

All models in mereology start by defining at a first pass the 'is part of' relation by means of three axiom, namely, (P1)  $Pxx$ ; (P2)  $Pxy \wedge Pyx \rightarrow x=y$ ; (P3)  $Pxy \wedge Pyz \rightarrow Pxz$ , where ' $P$ ' is a two-place predicate to be interpreted as the parthood relation. In other words, the parthood relation is reflexive, antisymmetric, and transitive.

I said in 'a first pass' because this definition has a serious shortcoming. If parthood is transitive, since the handle is a part of door and the door is a part of a house, it follows that the handle is a part of the house too. To avoid this conclusion, the general intended interpretation of " $P$ " by the three axioms should be narrowed by introducing some additional conditions restricting its application. Such a condition involves functionality, which requires that  $x$  be a part of  $y$  if  $x$  makes a direct contribution to the functioning of the whole of which  $x$  is a part. In this case, the handle is a functional part of a door and, thus, it is a proper part of the door, and although the door is a functional part of the house, the handle is not a functional part of the house and, thus, the handle is not a proper part of the house. Mathematically put, where ' $\varphi$ ' is any formula in the language, the implication: (1)  $(Pxy \wedge \varphi[x,y]) \wedge (Pyz \wedge \varphi[y,z]) \rightarrow (Pxz \wedge \varphi[x,z])$  may well fail to be a theorem of the mereological theory if  $x$  is not functionally related to  $z$ . Note that the functional restriction has its own problems. A spot on the door that is painted differently from the rest of the door is a part of the door, but it adds directly nothing to the functioning of the whole entity of which it is a part, that is, the door. Be that as it may, this discussion clearly bears on the problem of which parts of an iconic representations can be construed as parts of it properly speaking. I have argued that the part consisting of a part of the back of an object conjoined to a part of the background has no functionality in perceptual processing given the nature of our perceptual system; hence it is not a proper representationally speaking part of the representation, according to the restriction introduced in Mereology.

To address the aforementioned, and other problems, the meaning of " $P$ " is further clarified by a set of additional theorems that differ from one Mereology theory to another. Here I consider only those extensions that will be used in defining the sum or fusion of two parts to form a whole. These are the following:

(i)  $PPxy = \text{df } Pxy \cap \neg Pyx$  (Proper Part), that is,  $x$  is a proper part of  $y$  iff  $x$  is a part of  $y$  and  $y$  is not a part of  $x$ ;

(ii)  $Oxy = \text{df } \exists z (Pzx \cap Pzy)$  (Overlap), that is,  $x$  and  $y$  overlap iff there exists a  $z$  such that  $z$  is a part of  $x$  and  $z$  is a part of  $y$ ; and

(iii)  $Uxy =_{df} \exists z (P_{xz} \cap P_{yz})$  (Underlap), that is,  $x$  and  $y$  undelap iff there exists a  $z$  such that  $x$  is a part of  $z$  and  $y$  is a part of  $z$

To define the sum of two parts the theory is extended so as to constrain further the meaning of parthood. The first extension is:

(P4)  $\neg Pxy \rightarrow \exists z (P_{zx} \cap \neg Ozy)$ , that is, if an individual has a proper part, it has more than one.

(P4) entails the following theorems

(P4a)  $\vdash PPxy \rightarrow \exists z (PPzy \cap \neg Ozx)$ , that is, if  $x$  is a proper part of  $y$ , then there exists a  $z$  such that  $z$  is a proper part of  $y$  and it is not the case that  $z$  and  $x$  overlap;

(P4b)  $\text{EM} \vdash \exists z PPzx \cap \forall z (PPzx \rightarrow PPzy) \rightarrow Pxy$ ; and

(P4c)  $\text{EM} \vdash \exists z PPzx \cap \forall z (PPzx \leftrightarrow PPzy) \rightarrow x=y$ , which means that no two distinct objects can share the same proper parts.

The second extensions is:

(P5)  $Uxy \rightarrow \exists z \forall w (Owz \leftrightarrow (Owx \vee Owy))$ , that is, that if  $x$  and  $y$  undelap, then there exists a  $z$  such that for every  $w$ ,  $w$  and  $z$  overlap iff either  $w$  and  $x$  overlap or  $w$  and  $y$  overlap.

In a Mereological theory that holds (P1) to (P5), the following “sum” definition can be supported:

(Sum Definition)  $x+y =_{df} \exists z \forall w (Owz \leftrightarrow (Owx \vee Owy))$ , where ‘ $i$ ’ is a description operator in the language. This says that the sum of two parts  $x$  and  $y$  is defined as follows: for every  $w$ ,  $w$  and  $x$  overlap iff either  $w$  and  $x$  overlap or  $w$  and  $y$  overlap. Adopting the ‘+’ sum operator allows us to recast (P5) as follows:

(P5')  $Uxy \rightarrow \exists z (z=x+y)$ , that is, that if  $x$  and  $y$  undelap, then there exists a  $z$  such that  $z$  is the sum of  $x$  and  $y$ .

The sum definition determines the sum of two parts that are combined or fused to form a whole. This definition can be generalized to yield the notion of Unrestricted sum:  $\exists w \varphi_w \rightarrow \exists z \text{Siz} \varphi_w$ , where  $\varphi$  is any formula in a language (which picks out the components or parts that form the sum, say the grains of sand in a pile of sand). This definition stipulates that if there are parts  $\varphi_i$ , then there is an object  $z$  that is the sum of  $\varphi_i$ .

The most pressing problem with Mereological theories, a problem that pertains directly to the compositionality of iconic representations, is that the way the “sum” operation is defined, Mereology by itself cannot capture some of the basic properties we attribute to wholes (for example, a whole is a one-piece, self-connected entity, such as an object in a visual scene, as opposed to a scattered entity made up of several disconnected parts, such as parts of different objects). Parthood is a relational concept, whereas wholeness a global property (Varzi 1996, 269), and, thus, the former, and the ensuing definition of “sum”, cannot capture completely the meaning of “wholeness”. This shows immediately if one considers that in Mereology for every whole there is a set of parts, and that for every specifiable set of parts (for example, arbitrary objects) there is in principle a complete whole, i.e., its mereological sum, or fusion. As the “Sum definition” shows, the only restrictions concern the overlapping relations between the combined parts, but this is hardly enough to constrain the definition so that only combinations of parts that result in a whole *qua* object that satisfies our basic understanding of what should count as an object. Thus, within Mereology itself, there is no way to draw a distinction between “good” and “bad” wholes, and, thus, there is no way to distinguish between an integral whole and a scattered sum of disparate parts (yes, this is the *Gavagai* problem).

The problem is that it is not possible, on pure Mereological grounds, to determine the appropriate restrictions that would permit the fusion of parts in ways that allow only the formation of integral or natural wholes (such as objects or whole visual scenes), and would exclude the formation of sums of disparate parts or concrete heap-like composites (such as a pile of bricks, or sums of disparate parts of objects or background). Mereology does not say what constitutes a natural whole, since the existence of a sum  $x+y$  is conditional to the existence of an object  $z$  containing both  $x$  and  $y$ , in the sense that if  $x$  and  $y$  undelap, there exists a  $z$  such that  $z$  is the sum of  $x$  and  $y$ . This allows that parts of different objects be combined to form a whole, which thus consists in a conglomeration of disparate parts, and which is, thus scattered all over the place. Consequently, any attempt to account for basic spatio-temporal relations, such as the relationship between an object and

its surface, or the relation of something being inside or around something else., which are some among the relations that any theory concerned with spatio-temporal entities should supply, cannot be defined directly in terms of mereological primitives only. This is not the only misgiving with the unrestricted form of fusion, as there are arguments that it

does not sit well with certain fundamental intuitions about persistence through time . . . that it is incompatible with certain plausible theories of space . . . or that it leads to paradoxes similar to the ones afflicting naïve set theory. (Varzi 2016, 38)

Many theoreticians accept this limitation of summation but argue that there is no structured way to restrict the notion of fusion or sum and that the standard definition of fusion in mereology is the only plausible option (Varzi 2016, section 4.5).

It is possible, of course, to restrict the definition of “sum” so as to output only sums that satisfy certain conditions, such as a sum of concrete parts must be a non-heap like entity, for example. In this case, if  $\varphi$  is any formula in a language (which picks out the components or parts of that form the sum, say the grains of sand in a pile of sand), and  $\psi$  is a condition that the sum must satisfy (for example, that the sum of some material parts must be a natural whole), the set of all  $\varphi$ -ers has a sum  $\varphi_i$  if and only if every  $\varphi$  is  $\psi$ , that is, it is part of a natural whole. This yields the following definition of restricted sum:  $(\exists w \varphi_w \wedge \forall w (\varphi_w \rightarrow \psi_w)) \rightarrow \exists z \text{Siz} \varphi_w$ .

Accordingly, many theoreticians propose that mereology should be supplemented with some topological theory (Pianesi & Varzi 1996; Norton 2011). Pianesi & Varzi (1996) and Fletcher & Lackey (2022, 15) call the resulting amalgam of Mereology and Topology “Mereotopology”. A topological theory provides a primary predicate that is essential in solving the abovementioned problems, namely the relationship of “connectedness” by introducing the *Connection* predicate ‘C’. According to C, in order for two parts to form a sum it is necessary that they be connected or joined to each other, a demand that introduces the notions of contiguity and of interval relations that promptly show in discussions of iconic representations, as indispensable ingredients for forming natural wholes. ‘C’ is defined as follows: (C1) Cxx; (C2) Cxy  $\rightarrow$  Cyx; (C3) Pxy  $\rightarrow$   $\forall z (Czx \rightarrow Czy)$ . In other words, x is connected to itself; if x is connected to y, y is connected to x; and if x is a part of y, then for every z, if z is connected to x then z is connected to y. This last axiom ensures that putting together disparate parts does not provide an acceptable sum.

There are three ways to accommodate the fusion of Topology with Mereology. The first is to accept that the two theories together can provide an adequate framework to explain the part/whole compositionality, the second is an attempt to fuse Mereology into Topology by defining the *Parthood* relation P of Mereology in terms of the connection predicate C of topology, and the third is an attempt to subsume Topology under Mereology by defining the connection predicate C in terms of P and the vocabulary of Mereology. The details need not concern us, because what is important is that to determine the conditions under which parts can be summed or fused to form natural, that is, integral wholes, and exclude the formation of sums of disparate parts require both notions of ‘parthood’ and connectedness.

My analysis of the whole/parts compositionality befitting iconic representations reflects Werning’s (2012) work. Werning employs neurobiological findings concerning topologically structured cortical feature maps and the mechanism of object-related binding by neuronal synchronization to argue that iconic representations do compose, but their compositionality does not follow the principle of semantic constituency. The use of neuronal synchronization mechanisms underlying object-related binding can be used to complement my discussion of the neuronal implementation of the principles of Mereotopology underlying the formation of well-formed whole objects out of parts. Synchronization is the preferable mechanism nowadays invoked to explain how feature binding takes place to produce the representation of whole objects, and is used by a class of models that purport to explain how different attributes that are registered and processed in different visual areas can be bound together to form a visual object. The main characteristic of these models is that oscillators with neighbouring receptive fields and similar feature selectivities tend to synchronize . . . whereas oscillators with neighbouring receptive fields and different feature

selectivities tend to desynchronize. As a consequence, oscillators selective for proximal stimulus elements with like properties tend to form a synchronous oscillation when stimulated simultaneously. This oscillation can be regarded as one object representation. In contrast, inputs that contain proximal elements with unlike properties tend to cause anti-synchronous oscillations, that is different object representations. This result is in line with the findings of object-related neural synchronization. (Werning 2012, 642)

Synchronization, thus, could be the mechanism, or one among others, underlying the compositionality of parts to form wholes in vision. As the quotation makes clear, neighborhood relations are crucial in guiding synchronization, which brings in the role of Mereotopology in explaining the bindings of parts of objects to produce natural objects. Werning shows that the network of the top-down and bottom-up signals as implemented by neuronal synchronization can be modelled by means of neural networks that, famously, do not employ symbolic representations. This means that object-binding may be a case of composing non-symbolic, iconic representations following the principles of Mereotopology.

Before I turn to examine in some detail the way neural synchronization is involved in perceptual groupings, let me say that the discussion of iconic compositionality in terms of Mereotopology is relevant to various problems plugging the application of the *Picture Principle* in visual representations (see Burge (2022) for a thorough discussion). One of the problems of the principle is that it entails that any part of a picture is a representation of a part of what the picture represents. This is a problem for applying the principle to visual perception, because if one combines the back part of an object in a visual image and a part of the immediate background in the image and forms a whole, this whole is irrelevant in terms of what is computationally relevant in perception, and, thus, it is highly unlikely that this complex part of the image represents anything in perception. In terms of Mereology and Topology, this combination does not result in a natural, integral whole. In this sense it is not true that any part of the representation represents a part of the domain that the representation represents; only parts that are admissible as components of perceptual processes are admitted. Which parts are these is an empirical issue, which means that it is the perceptual system itself that solves the problem of which combinations of parts of an image are admissible as natural wholes.

We saw before that oscillating neurons with neighbouring receptive fields and similar feature selectivities tend to synchronize. It is thus plausible to suggest that synchronization is the mechanism underlying the neural implementation of the constraints of “local proximity” and “feature similarity”. Since such neurons are usually grouped together in the brain, Local Field Potentials (LFP’s) are very useful to study the activity of these neurons and their synchronization, because the LFP’s are transient electrical signals generated in groups of neurons by the summed and synchronous electrical activity of individual neurons. They express the aggregate activity of small populations of neighboring neurons represented by their extracellular potentials. Unlike action potentials that are generated by individual neurons, LFPs measure synaptic potentials pooled across groups of neurons near the recording electrode.

Studies (Baldauf & Desimone 2014; Bastos et al. 2015; Fries et al. 2001, Gregoriou et al. 2009; 2015) illuminate the role of LFPs in tasks involving top-down attention and VWM tasks, since these two play a pivotal role in grouping behaviorally relevant stimuli. These studies suggest that attention increases gamma frequency synchronization, increases low-frequency alpha-band synchronization for distractors, reduces low-frequency alpha-band synchronization of V4 neurons representing behaviorally relevant stimuli, increase theta-band frequency, and increases low-frequency beta synchronization for attended stimuli. As we shall shortly see, top-down visual attention originates in the Prefrontal Cortex (PFC), which seems to modulate through low-frequency waves the activity in visual areas from V1 to V4, a modulation that results in increased synchronization at the gamma-band high-frequencies between either Front Eye Fields (FEF) in the case of spatial attention, or Inferior Frontal Junction (IFJ) in the case of object/feature-based attention, and the visual areas.

Gregoriou et al. (2009; 2015) found that the distributions of the latencies of attentional effects in LFP gamma power in both FEF and V4 were significantly later than the distribution of the latencies

for attentional effects on the firing rates in FEF, and significantly earlier than the distribution of latencies for attentional effects on V4 firing rates. These results indicate that significant attentional effects on LFP gamma power in either FEF or V4 occur later than the earliest attentional effects on firing rates in the FEF. Thus, rather than being caused by enhanced gamma oscillations, increases in firing rates in FEF with attention may initiate the coupled oscillations within and across areas. In contrast, firing rate changes in area V4 occur later and might result at least in part from enhanced gamma oscillations. Fries et al. (2001) found that the coupled oscillations remain even during the delay period where the firing activity in visual areas is subthreshold. More generally, increases in gamma synchrony are found among cells that decrease, or show no change in, their firing activity (Brunet et al. 2014).

During the delay period of VWM tasks, cells in PFC show increased activation rates, whereas it is likely that the activations in visual areas are subthreshold. According to the evidence discussed in the previous paragraphs, the increased firing rates activity in PFC, and the increased activity in visual areas, especially in FEF, due to top-down attention, induce coupled oscillations at the gamma band frequency both within PFC and across visual areas, starting from V4 where the attention effects are more pronounced and extending to other mid-level and high-level visual areas. The attentional effects on mid-level and high-level visual areas during the delay period, thus, manifest themselves in the LFP gamma power and not in the firing rates in the neurons in mid- or high-level visual areas. It should be noted that reports concerning the impact of attention on neuronal synchronization in early visual areas are conflicting (Gregoriou et al. 2015). Chalk et al. (2010) report that attention reduced gamma-synchronization in V1, probably due to a decrease in the inhibitory drive that controls surround suppression.

We examined evidence showing the role of attention in increasing gamma-band high-frequencies synchronization both between visual areas, and between visual areas and higher cortical areas. Is this an attentional effect, or is this the way attention affects perceptual processing? In other words, how exactly does top-down attention affect the oscillations in the inter-communicating areas? The answer is that top-down attentional effects on visual areas are transmitted through low frequencies in alpha- and beta-bands and modulate the gamma-band activity in the modulated visual areas. Several studies in monkeys and humans show that spatial attention reduces local low-frequency alpha-band synchronization in visual areas and V4 (Gregoriou 2015; Fries et al. 2001), in contradistinction to the gamma band activity that is increased by attended stimuli in V4. Specifically, distractors increase alpha-band activity, whereas attended stimuli reduce alpha-band activity. Beta-band activity, on the other hand increases for attended stimuli. As Anderson et al. (2011) show, FEF that plays a predominant role in directing top-down spatial attention, excites inhibitory neurons in target areas on the visual cortex during attentional modulation. Similarly, Gregoriou et al. (2015) suggest that alpha band waves may play a role in suppressing irrelevant stimuli, enhancing, thus, the activations of neuronal assemblies representing the attended stimuli.

Research (Bastos et al. 2015; Michalareas et al. 2016; Fries et al. 2001; Fries 2015) suggest that gamma band waves subserve feedforward signalling, whereas the alpha-beta band waves subserve top-down, feedback flow of information. When top-down attention affects visual processing, the predominantly bottom-up directed gamma-band (high frequency 30-90 Hz) influences are controlled by predominantly top-down directed alpha-beta band (8-20 Hz) influences. Plomp et al. (2014) and van Kerkoerle et al. (2014) show that stimulation of V1 induces enhanced gamma-band activity in V4 (V1 to V4 feedforward projections), while stimulation of V4 under visual stimulation with a background stimulus induces enhanced alpha-beta activity in V1 (V4 to V1 feedback projections), which likely suppresses the background stimulus. In general, attentional top-down influences carried by low-band frequency waves are thought to modify gamma-band synchronization in the lower areas that receive the attentional feedback enhancing the feedforward signals emanating from these areas. Top-down signals increase both the synchronization strength, as measured by LFPs power, and the synchronization frequency of the gamma-band synchronization.

Fries et al. (2001) found that synchronization or coherence was modulated by spatial attention very early in the response to the stimulus. The Spiked-Triggered Averages (STAs) for the 100-ms. period after response onset (starting 50 ms. after stimulus onset because it takes about 50 ms. for the brain to start responding to the incoming stimuli) contained large low-frequency modulations with superimposed gamma-frequency modulations. The low-frequency alpha-band (10 Hz) synchronization was reduced by attention, whereas, there was a smaller gamma-frequency peak at around 65 Hz that was enhanced by attention. Both the visual evoked potential (VEP) and the spike histogram contained strong stimulus-locked gamma-frequency oscillations in the first 100 ms. of the response. Since gamma-band synchronization is related to bottom-up processing, this finding shows the bottom-up, stimulus locked activity in visual areas. However, oscillatory synchronization during the later sustained visual response was not stimulus locked, which shows that effects of spatial attention that modulates perceptual processing. This is shown by the fact that the low-frequency synchronization in the alpha-band was reduced by attention (alpha band synchronization is reduced for attended stimuli and increased for distractors). Since attention strengthens the representation of the attended stimulus in V4, it facilitates the bottom-up signals from V4 to IT and other cortical areas, which explains the increase in gamma-band frequency. Thus, attention increased gamma high-frequency and reduced low-frequency synchronization of V4 neurons representing the behaviorally relevant stimulus. This was observed even during the delay period and in the first few hundred milliseconds after response onset, when firing rates were not affected, because spatial attention fixated at a certain location modulates the preparatory activity of the neurons whose receptive field falls within the attended location before the presentation of the stimulus and this effect is carried on after stimulus onset before the attentional cue is being presented.

In contradistinction to alpha-band waves that suppress irrelevant stimuli, since distracting stimuli enhance alpha band oscillatory activity, beta-band wave activity may enhance the activation of attended stimuli by inducing stronger synchrony in lower frequency bands, which enhances the top-down signals. Bastos et al. (2015) argue that top-down signals that facilitate processing of attended stimuli are carried by beta-band (14-18 Hz) synchronization. As Bastos et al. (2015) suggest, cognitive tasks enhance top-down beta-band influences. Moreover, when attention selects a stimulus and enhanced top-down signals (carried by beta-band waves) reach the representation of the attended stimulus in visual cortical areas, this may lead to enhanced bottom-up signalling of that stimulus carried by gamma-band waves. Bosman et al. (2012) show that bottom-up causal influences from V1 to V4 are enhanced when they carry information about the attended stimulus, in accordance with Fries et al. (2001) finding that neurons activated by attended stimuli show increased gamma-frequency.

I claimed above that the top-down flow of information is subserved by alpha and beta-band synchronizations and the bottom-up flow of information by gamma (and theta)-band synchronizations. We also saw that there is substantial evidence that gamma-band causal influences between FEF and V4 predominate in the FEF-to-V4 direction after an attentional cue (post-cue condition), but subsequently predominate in the V4-to-FEF direction. In view of the fact that FEF are anatomically higher than V4 and, thus, the connections from FEF to V4 seem to be feedback/top-down projections, one might think that gamma-band synchronization subserves feedback flow of information, which contradicts the evidence that feedback is carried by low frequency waves. However, there is no discrepancy since FEF may be anatomically higher in the hierarchy than V4 but functional things are different.

FEF is situated in the PFC at a site heavily interconnected with the parietal cortex and is a part of the dorsal visual system. The mean activation latency of FEF neurons is 70 ms poststimulus. Signals arrive at FEF with a slight (if at all) time delay time with respect to the signals arriving at V1 (50-80 ms.) and V2 (85 ms.) and much earlier than they arrive at V4 despite the fact that FEF is anatomically higher than V4. Thus, functionally, FEF is at a lower level than V4 and, therefore, the projections from FEF to V4 are feedforward. This describes prestimulus conditions since it indicates the potential functional relations of FEF and V4. FEF contains visual and movement neurons (O'Shea et al. 2004).

Bastos et al., (2015) show that in the post-cue condition, that is, before spatial attention triggered by the cue intervenes, 8L area of FEF is lower than V4, which means that the projections from area 8L of FEF to V4 are feedforward. Thus, Gregoriou's et al. (2009) finding that gamma-band causal influences between macaque FEF and V4 predominate in the FEF-to-V4 direction after an attentional cue reinforces rather than contradicts the view that feedforward influence are subserved by gamma-band synchronizations. Notice that this description purports to explain the increase in gamma synchronization between FEF and V4 and the fact that the causal influence is from FEF to V4, in other words, that FEF projects feedforward signals to V4. It does not account for the top-down attentional effect of FEF, and especially of area 8M, on V4, because this is carried by low-frequency waves, as top-down signal propagation predominantly is, from FEF to V4.

Recall that Fries et al. (2001) found that spatial attention modulates synchronization in V4 at 150 ms. after stimulus onset, when large low-frequency modulations were superimposed by gamma-frequency synchronizations. The low frequency waves on V4 carry the top-down attentional modulation from area 8M of FEF to V4. Bastos et al. (2015) show that area 8M is slightly higher than V4. The gamma-band synchronization in the FEF to V4 direction before the cue probably shows the effect of the processing of the visual stimulus in area 8L of FEF in which the visual neurons of FEF are clustered and which process the stimulus and are independent of attention. Moreover, once the representation in V4 of an item at the attended location has been boosted by the top-down attentional effects carried by low-frequency waves, the feedforward projections from V4 to FEF dominate and enhanced bottom-up signalling of that stimulus carried by gamma-band waves prevails. This explains Gregoriou et al. (2009) finding that that gamma-band causal influences between macaque FEF and V4 predominate in the V4-to-FEF direction later in the post-cue period.

One might be puzzled by the fact that the same areas (for example, V4 and FEF) can have both feedforward and feedback projections between them, which leads to the functional hierarchy exhibiting dynamic changes. Fries (2015) argues that when two neuronal areas are bidirectionally connected, unidirectional entrainment (that is the causal influence of one on the other) occurs separately in both directions of the bidirectional link. Anatomical data show that for each direction of communication, the linked brain areas have specialized neuronal groups for sending and receiving signals; that is, a specific brain area has neurons receiving inputs and different neurons sending outputs (Fries 2015; Markov et al. 2014).

Jensen et al. (2014) examine the role of alpha oscillations in relations to gamma oscillation in attentional effects on the visual areas. Irrespective of the exact nature of the signals that MVPA decodes, it is clear that VWM involves visual areas. When the test display appears, the perceptual areas also receive bottom-up, stimulus driven activation and neuronal assemblies in these areas encode the perceptual information in the test display. This information is compared to the perceptual information concerning the sample item that is stored in the connection weights of neuronal assemblies in visual areas that are reactivated by both the bottom-up signal from the test item owing to the distributed nature of representations, and by the top-down signals from PFC.

Work by Nakatani and Leeuwen (2006) elucidates the role of attention in perceiving ambiguous figures relating attentional activity to the synchronized activity in the right parietal areas that are responsible for perceptual awareness, and in the right frontal areas that correlate with perceptual flexibility and, hence, to perceptual switching between the two possible percepts of the ambiguous or bistable figure. Note that the same areas are also involved in top-down selective attention (Corbetta and Shulman 2002). Nakatani and Leeuwen's (2006) research shows two cycles of synchrony in the gamma band; the first occurs 800–600 ms., and the second 400–200ms before button pressing. The first period of synchronicity coincides with a drastic suppression of eye blinks that is related to attentional demands, as these demands make viewers to focus closely postponing saccades (Ito et al. 2003). The second period of synchronicity in the observed activity patterns in PFC coincides with the maximum saccade frequency that reaches its peak at about 250ms before the switch response. Since saccade frequency is associated with shifts of attention (Leopold and Logothetis 1999), the second period of synchronicity probably reflects the final focus of spatial attention after a series of attentional

shifts, which, by determining the critical points on the image also determines which interpretation of the ambiguous figure will be perceived.

Nakatani and Leeuwen (2006) also explored the role of the activity in frontal and occipital cortex during switching episodes. They found that the theta activity in the frontal cortex is a general characteristic of the processing activity of viewers that perform frequent switches, but is not specifically related to perceptual switching. Increased theta band activity in the frontal cortex is related to the concentration of attention to a task and to the inhibition of eye blink (Yamada 1998), as is the activity in the first period of synchrony in the gamma band. The alpha band activity observed in the occipital cortex is related to frequent perceptual switches. Increased alpha activity in the occipital cortex is related to attention to the stimulus by enhancing the efficiency of information processing (Yamagishi et al. 2003). Thus, the frontal and occipital cortex activity during perceptual switches signifies the crucial role of attentional modulation of the perception of ambiguous figures and its effects on the rate of perceptual switches.

The previous studies concern the role of synchronization in attention tasks. Attention is closely related to tasks involving VWM as well and, so, it would be interesting to see what role synchronization plays in VWM tasks. Examining the role of synchronic activity among brain regions in memory tasks, Roux & Uhlhaas (2014) proposed that gamma-band oscillations are specifically involved in the active maintenance of VWM information. Theta-band oscillations are specifically involved in the temporal organization of NWM items, a view that generalizes Nakatani and Leeuwen's (2006) proposal. Finally, alpha-band oscillations are involved in the inhibition of task-irrelevant information, which results in enhancing the efficiency of information processing, as Yamagishi et al. (2003) proposed. However, other factors may contribute to the enhancement of the efficiency of information processing in occipital area, and the increased alpha activity may reflect these factors as well. Such factors may a direct enhancement of task-relevant information (Miller & Cohen 2001), or the sharpening of the representations of different object categories in the extrastriate cortex by an increase of the distinctiveness of their distributed neural representations (Fuster et al. 1985).

These results are based on studies demonstrating amplitude modulation of neural oscillations presumably emanating from particular brain regions involved in WM. During a delayed match-to-sample task while recording human EEG, Tallon-Baudry et al. (1999) observed that occipital gamma and frontal beta oscillations were sustained across the retention interval. Moreover, as the delay interval increased, these oscillations decreased in parallel with decreased performance on the task. Anderson et al. (2014) showed that the spatial distribution of power in the alpha frequency band (8–12 Hz) tracked both the content and the quality of the representations stored in visual working memory. Recall that in memory tasks there is increased power in the alpha frequency band in the occipital cortex related to the enhancement of the efficiency of information processing by attention to the stimulus (Yamagishi et al. 2003). These empirical findings together support both the view that neural oscillations are critical for VWM maintenance processes, and the view that in VWM tasks posterior visual processing areas play a critical role in sustaining the representations held in VWM. Finally, Lee et al. (2005) found evidence of enhanced local field potentials (4–10Hz) in area V4 of the monkey during a visual working memory task.

Long-range synchronization of the oscillations between brain regions likely also plays an important role in VWM function (Crespo-Garcia et al. 2013) as well. In a human MEG study, synchronized oscillations in the alpha, beta, and gamma bands were observed between frontoparietal and visual areas during the retention interval of a delayed match-to-sample visual working memory task. These synchronized oscillations were sustained and stable throughout the delay period of the task, were memory load dependent, and were correlated with an individual's VWM capacity (Palva et al. 2010).

These studies bring to the fore the crucial role of synchronous oscillations in alpha- beta- theta- and gamma-frequency-bands in top-down attention and in VWM tasks. Top-down attention effects are carried by low frequency oscillations that synchronize the LFPs oscillations between the affecting

and affected brain areas, while bottom-up signals are carried by high-frequency gamma-band oscillations.

The foregoing discussion shows the close collaboration between cognitive centers and visual areas in the brain in VWM, since the higher-level cognitive centers guide attention through which they sustain the perceptual representations in visual areas during memory tasks. This suggests that the perceptual information used in memory tasks is most likely represented in visual areas and, in this sense, memory recruits the representations in these areas to achieve its goals. This is the basic premise of sensorimotor recruitment models of VWM, a class of models that hold that the systems and representations engaged to perceive information can also contribute to the short-term retention of that information in VWM (D'Esposito & Postle 2015).

We can bring in, now, the upshot of the discussion about Mereology and Topology, which is that to solve the problem of how to constrain the fusion of parts so that only natural objects be accepted as proper wholes, one must consider both "parthood" and "connectedness". Since the visual system has solved this problem, it is plausible to assume that its computations directly implement, in one form or another, the Mereotopological principles of, say, "parthood" and "connectedness" and combine them in such a way that only natural wholes are represented in visual perception under normal conditions. The reader has perhaps made this connection as a result of our discussion of the neural mechanisms of synchronicity that underlie the compositionality of visual representations. If iconic compositionality is realized by neural mechanisms and the ways parts are composed are hardwired in the visual system, and if these compositions are expressed by the principles of Mereotopology, it follows that the principles of Mereotopology express the functioning of the relevant neural mechanisms.

At the same time, the operation of the visual system, as we have seen, is characterized by a set of operational constraints. Thus, these constraints must be realized by the neural circuits in our visual system. This is what Han et al. (2002) studies show. Han et al. (2002) studied the neural mechanisms underpinning the operational constraints of "local proximity" and "similarity". The findings of their study suggest that proximity grouping resulted in short latency modulations of medial occipital activity that was followed by longer latency modulations in the occipito-parietal cortex. Proximity grouping that relies on "local proximity" induced similar medial occipital modulations at 110 ms., which suggests that it depends mainly on representations of spatial relationships between local elements, and is independent of visual features. Grouping by color similarity that relies on feature similarity produced only long-latency occipito-temporal modulations.

In view of these considerations, it follows that the operational constraints in vision should reflect the principles of Mereotopology. I shall argue now that they do so, starting with a brief account of the operational constraints in vision. As I have repeatedly said (Author), perceptual computations are constrained by a set of what I have called operational constraints. Burge (2010) calls them "formation principles", Echeverri (2017) calls them "object constraints", and some among them are also known as Spelke (1990) principles of "object perception". These principles also figure in Haugeland's account of perception, where he claims (Haugeland 1998, 261) that non-concept possessing creatures and we share various innate "object-constancy" and "object-tracking" mechanisms that automatically 'lock onto' medium sized lumps of matter. These mechanisms implement the operational principles. The operational constraints reflect general or higher-order physical regularities that govern the behavior of objects in our world and the geometry of our environment, and which have been ingrained in the perceptual system through causal interaction with the environment over the evolution of our species. This means that they reflect generalities in the world given our physical constitution and needs for survival.

Empirical studies by Spelke (1990), Spelke et al. (1995), and Karmiloff-Smith (1992) strongly support the assumption that the infant, from the beginning of life, is constrained by a number of domain-specific principles about material objects and some of their properties. These constraints involve 'attention biases toward particular inputs and a certain number of principled predispositions constraining the computation of those inputs.' (Karmiloff-Smith 1992, 15) Among these

predispositions are the conception of object persistence, and four basic principles: boundness, cohesion, rigidity, and no action at a distance.

The operational constraints function at almost all levels of visual processing, are hardwired in the brain, and do not entail that perception is cognitively penetrated since they do not constitute cognitive states that affect perceptual processing (Raftopoulos 2009; 2019, chapter 3). One of these principles is *cohesion* (Bloom, 2000). 'Objects are connected and bounded bodies that maintain both their connectedness and their boundaries as they move freely' (Spelke et al. 1995, 45). That is, the *cohesion principle* dictates that two surface points lie on the same object only if the points are linked by a path of connected surface points. This entails that if some relative motion alters the adjacency relations among points at their borders, the surfaces lie on distinct objects, and that all points on an object move on connected paths over space and time. When surface points appear at different places and times such that no connected path could unite their appearances, the surface points do not lie on the same object. This constraint reflects the principle of "connectedness" of Topology that, as we have seen, supplants the rules of Mereology to restrict the definition of "sum" so that only compiling parts that yield natural whole objects is an acceptable notion of "sum". This is so because if two surface points lie on the same object only if the points are linked by a path of connected surface points, then the parts on which the two points lie are connected. The importance of the cohesion principle is manifested by the finding that some violations of cohesion seem to destroy infants' representations of enduring object (Chiang & Wynn, 2000; Huntley-Fenner et al., 2002). Mitroff et al., (2004a, b) and vanMarle & Scholl, (2003) show that even adults' visual processing is critically affected by cohesion violations.

Another principle, closely related to "cohesion" is the *principle of solidity*; 'objects move only on unobstructed paths: no parts of two distinct objects coincide in space and time' (Spelke, et al., 1992, 606). This is clear expression of Theorem (P4c) of Mereology presented above that states that no two distinct objects can share parts, and shows how the perceptual system has hardwired basic regularities of the environment, in this case, a basic Mereological property of solid objects.

Another constraint is the *boundness principle*, according to which two surface points lie on distinct objects only if no path of connected surface points links them. This principle determines the set of those points that define an object boundary, and entails that two distinct objects cannot interpenetrate, because two distinct bodies cannot occupy the same place at the same time. Finally, the *rigidity* and *no action at a distance* principles specify that bodies move rigidly (unless the other mechanisms show that a seemingly unique body is, in fact, a set of two distinct bodies) and that they move independently of one another (unless the mechanisms show that two seemingly separate objects are in fact connected). These constraints guide the perception of the motions of objects, of the layout of adjacent objects, of object boundaries, and of object segmentation by both adults and infants, and play a crucial role in the segmentation processes that take place in the visual system upon viewing a scene. These constraints relate both to "connectedness" and to theorem P4a of Mereology. The former because if there is no path of connected surface points lining the two surface points, they belong to unconnected parts that form distinct objects since they cannot be combined to form a natural whole. Theorem P4a applies because it rules out the possibility that two object parts that combine to form a whole object can have any common parts (they cannot overlap), because if that were the case, two distinct object wholes formed of parts that can overlap could penetrate one the other over the points of overlapping.

There are more constraints at work in perception than those mentioned above. The formation of the *full primal sketch* in Marr's (1982) theory, for instance, which involves the grouping of the edge fragments formed in the *raw primal sketch*, relies on the principles of "local proximity" (adjacent elements are combined), which shows the Topological principle of "connectivity" at work, and of "similarity" (elements with similar features are combined). All these principles are parts of "perceptual grouping", which refers to the function of the human visual system to organize discrete entities in the visual field into chunks or perceptual objects. The principle of local proximity states that spatially close objects or object parts tend to be grouped together, constraining thus which parts

can be combined to form natural wholes. The principle of similarity claims that elements with similar features in the field tend to be grouped together. Grouping processes have been assumed to take place at an early stage in the visual processing stream. Perceptual grouping also relies on the more general principle of “closure” (two edge segments could be joined even though their contrasts differ because of illumination effects) (Bruce and Green 1993, 131–132). Other assumptions that are brought to bear upon the early visual processing to solve the problem of the underdetermination of perception by the retinal image are those of “continuity” (the shapes of natural objects tend to vary smoothly and usually do not have abrupt discontinuities), “proximity” (since matter is cohesive, adjacent regions usually belong together and remain so even when the object moves), and “similarity” (since the same kind of surface absorbs and reflects light in the same way, the different subregions of an object are likely to look similar).

The formation of the 21/2D sketch is similarly underdetermined, in that there is a great deal of ambiguity in matching features between the two images form in the retinas of the two eyes, since there are usually more than one possible matches. Stereopsis requires a unique matching, which means that the matching processing must be constrained. The formation of the 21/2D sketch, therefore, relies upon a different set of operational constraints that guide stereopsis. ‘A given point on a physical surface has a unique position in space at some time’ (Marr 1982, 112), and matter is cohesive and surfaces are generally smooth. These operational constraints give rise to the general constraints of “compatibility” (a pair of image elements are matched together if they are physically similar, since they originate from the same point of the surface of an object), “uniqueness” (an item from one image matches with only one item from the other image), and “continuity” (disparities must vary smoothly).

I think that it is not hard to see, and I have given some examples, that almost all of these operational constraints are directly or indirectly related to the connectedness and parthood Mereotopological relations among elements in perceptual computations and that their role is exactly to ensure the formation, and thereby visual representation, of natural wholes. All these suggest that the *Picture Principle* as it applies to perception is severely constrained and that it is not the case that any way you cut a visual representation of a visual scene you get a representation of some part of this scene.

We know that adjacent regions in a visual scene are registered by adjacent regions in the retina, and, through retinotopic projections, are represented by adjacent regions in most of the brain. Thus, neighborhood relations, which is a topological notion strongly associated with ‘connectedness,’ are retained in the brain. Moreover, parthood relations in a visual scene are also retained in the brain because a neural representation of, say, a whole object or scene consists, in the sense of containment, of the neural representations of its parts, the same way a picture of a scene contains as parts pictures of the constituents of the scene. Thus, mereological relations are also retained in the brain. It follows that even though mental states are not spatially arrayed in space as pictures are, mereotopology is still applicable and can be put into use to show how the brain distinguishes natural wholes from mere collections of parts (recall that mereology by itself cannot solve this problem but requires the topological notion of ‘connectedness’). This means that the mathematical tools of mereotopology can be used not only to describe the compositionality of visual perceptual contents, but also the compositionality of their neural vehicles.

## 5. Discussion: How the Environments Affects Minds

If the theses defended above are correct, it follows that the structure of the information-carrying vehicles and events in the environment structures the make-up of the representational perceptual vehicles of organisms (by hardwiring, as it were, in them, the basic principles that govern the ways objects in the environment compose and interact) in order to enable them to extract successfully the information in the environment and organize the information flow. The reason for this is that way objects in the environment compose must be reflected in the ways the respective perceptual representational vehicles compose for the perceptual system to be able to extract successfully the

information carried by the environmental structures. This make-up, in turn, determines the way the organism manipulates the incoming information and, hence, interacts with the environment.

To give an example of how this works, I argued in the preceding paragraph that our visual perceptual system has been structured in part by pressures put upon it by the organization and relations of the objects in the environment, by means of hardwiring the basic principles, expressed mathematically by Mereotopology, that govern the way objects and events in the environment compose to form more complex objects; in this sense, these principles are internalized. The view that the structure of the information-bearing vehicles (that is, their interrelations and organization) of exogenous information can be internalized into cognitive routines has also been defended by Sterelny (2012). Not only can the structure of information-bearing vehicles of exogenous information be internalized, but, also, the vehicles themselves can be internalized in iconic representations. The reason is that in iconic representations, the physical transformations of the information-bearing vehicles in the environment must be reflected and matched by the physical transformations of the representational vehicles, whether it be external representations or internal, mental representations.

This conclusion rests on the assumption that material objects and physical events, as well as social constructions, constitute informational domains and, thus, can be construed as information-carrying vehicles; the information they carry can be construed as exogenous information since it is deposited in material structures that lie outside the organisms that exploit this information. These informational domains have a structure expressed by the organized relations among the information-carrying vehicles. '[D]evelopmental environments contain heterogeneous vehicles of informational content (external artefacts, linguistic labels, the behavior of other agents) organized in systems.' (Buskell 2021, 8540). Moreover, the structural relationships between these vehicles can make a difference to the development of cognitive skills, even if they are only present for short periods of time. (Buskell 2021, 8540) Buskell's point is important because it shows how cognitive capacities can be created by pressures to agents to exploit effectively environmental information, irrespective of whether these cognitive novelties are genetic, epigenetic, or the result of the restructuring of the brain. These pressures cause the brain to reorganize its structure, or get new structures, which, in turn, render cognitive innovations possible. It is interesting to note that, if this view is correct, the external world imposes, as it were, its structure on the mind so that the transformations in the worldly states of affairs map in a natural way and not by some convention onto the transformations in the states of the mind. As the reader recalls, this is a characteristic trait of iconic representations, which may suggest that at the level of the interactions of the mind and the environment at the level of actions, the mind represents iconically the world, transferring the structure and processes of the world and of the actions on the world into the mind.

All these are important because arguments that support the view that our minds impose structure onto the world, which is correct to some extent, tend to overlook the fact that before that the world has imprinted in many respects its own structure in our minds, rendering them more adept at perceiving and understanding it.

## References

1. Anderson, J. C., Kennedy, H., & Martin, K. A. (2014). Pathways of attention. Synaptic relationships of frontal eye fields to V4, lateral intraparietal cortex, and area 46 in macaque monkey. *The Journal of Neuroscience*, 31(30), 10872-10881.
2. Ayers, M. (2019). *Knowing and Seeing: Groundwork for a New Empiricism*. Oxford: Oxford University Press.
3. Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, 344, 424-427.
4. Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., Weerd, P. D., Kennedy, H., & Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, 85, 390-401.
5. Beck, J. (2012). The generality constraint and the structure of thought. *Mind*, 121(483), 563-601.
6. Beck, J. (2018). Analog mental representation. *WIREs Cognitive Science*, DOI: 10.1002/wcs. 1479.
7. Beck, J. (2019). Perception is analog: the argument from Weber's law. *Journal of Philosophy*, 116(6), 319-349.

8. Bloom, P. (2000). *How Children learn the Meanings of Words*. Cambridge, MA: MIT Press.
9. Bruce, V. & Green P. R. (1990). *Visual Perception*, 2nd ed. Sussex; Lawrence Erlbaum Associates.
10. Brunet, N. M., Bosman, C. A., Vinck, M., Roberts, M., Oostenveld, R., Desimone, R., De Weerd, P. & Fries, P. (2014). Stimulus repetition modulates gamma-based synchronization in primate visual cortex. *Proceedings of the National Academy of Science, USA.*, 111, 3626-3631.
11. Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., & Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron*, 6, 75(5), 875-88.
12. Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.
13. Burge, T. (2014a). Reply to Block: adaptation and the upper border of perception. *Philosophy and Phenomenological Research*, 89(3), 573-583.
14. Burge, T. (2018). Iconic representation: Maps, pictures, and perception. In W. Shyam & F. A. Dorio (Eds.), *The Map and the Territory* (pp. 79-100). Dordrecht: Springer.
15. Burge, T. (2022). *Perception: First Form of Mind*. Oxford: Oxford University Press.
16. Burns, B. (1987). Is stimulus structure in the mind's eye? An examination of dimensional structure in iconic memory. *The Quarterly Journal of Experimental Psychology*, 39(A), 385-408.
17. Burnston, D. (2017). Cognitive penetration and the cognition-perception interface. *Synthese*, 194, 3645-3668.
18. Buskell, A. (2021). Cognitive Novelty, Informational Form, and Structural-Causal Explanations. *Synthese*, 198, 8533-8533. <https://doi.org/10.1007/s11229-020-02585-4>.
19. Camp, E. (2018). Why maps are not propositional. In A. Grzankowski, & M. Montague (Eds.), *Non-propositional Intentionality* (pp. 19-45). Oxford: Oxford University Press.
20. Carey, S. (2009). *The Origins of Concepts*. Oxford, N.Y.: Oxford University Press.
21. Chalk, M., Herrero, J. L., Gieselmann, M. A., Delicato, L. S., Gothardt, S. & Thiele, A. (2010). Attention reduces stimulus driven gamma frequency oscillations and spike field coherence in V1. *Neuron*, 66, 114-125.
22. Chiang, W.-C., & Wynn, K. (2000). Infants' representation and tracking of multiple objects. *Cognition*, 77, 169-195.
23. Chomsky, N. (1965). *Aspects of a Theory of Syntax*. Cambridge, MA: The MIT Press.
24. Crane, T. (1992). The Nonconceptual content of experience. In T. Crane, ed., *The Contents of Experience: Essays on Perception* (pp. 136-157). Cambridge: Cambridge University Press.
25. Crane, T (2003). *The Mechanical Mind*, 2nd edition. London: Routledge.
26. Crespo-Garcia, M., Pinal, D., Cantero, J. L., Diaz, F., Zurrón, M., & Atienza, M. (2013). Working memory processes are mediated by local and long-range synchronization of alpha oscillations. *Journal of Cognitive Neuroscience*, 25, 1343-57.
27. Cummins, R., & Roth, M. (2012). Meaning and content in Cognitive Science. In R. Schantz (eds.), *Prospects for Meaning* (pp. 365-382). Cottingen: De Gruyter.
28. Deutscher, G. (2005). *The Unfolding of Language: An Evolutionary Tour of Mankind's Greatest Invention*. New York, N.Y: Picador.
29. Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: The MIT Press.
30. Dretske, F (2000). *Perception, Knowledge, and Belief: Selected Essays*. Cambridge: Cambridge University Press.
31. Echeverri, S. (2017). Visual reference and iconic content. *Philosophy of Science*, 84, 761-781.
32. Everett, D. L. (2017). *How Language Began*. New York, N.Y: Liverlight.
33. Fletcher, S. C. & Lackey, N. (2022). The introduction of topology into analytic philosophy: two movements and a coda. *Synthese*, 200, 197, [doi.org/10.1007/s11229-022-03689-9](https://doi.org/10.1007/s11229-022-03689-9).
34. Fries, P. (2015), Rhythms for cognition: communication through coherence. *Neuron*, 88(1), 220-235.
35. Fodor, J. (2007). The revenge of the given. In B. P. McLaughlin and J. Cohen (eds.), *Contemporary Debates in the Philosophy of Mind* (pp. 105-117). Malden, MA: Blackwell.
36. Frege, G. (1962), (1892). Begriff und gegenstand. In G. Patzig, ed., *Funktion, Begriff, Bedeutung*. Vandenhoeck and Ruprecht.
37. Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Research*, 330, 299-307.

38. Gazzaley, A., Rissman J., & D'Esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive and Affective Behavioral Neuroscience*, 4, 580–99.
39. Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Science*, 16, 129-135.
40. Giardino, V., & Greenberg, G. (2015). Introduction: varieties of iconicity. *Review of Philosophy and Psychology*, 6, 1-25.
41. Gombrich, H. E. (1951). *The Story of Art*. Oxford: Oxford University Press.
42. Goodman, N. (1972). Seven strictures on similarity. In *Problems and Projects* (pp. 437-446). New York, N.Y.: Bobbs-Merrell.
43. Goodman, N. (1976). *Languages of Art*, 2nd edition. Indianapolis: Hackett.
44. Goodman, N. (1978). *Ways of Worldmaking*. Indianapolis, IN: Hackett Publishing Company.
45. Goodman, N. (1984). *Of Mind and Other Matters*. Cambridge, MA: Harvard University Press.
46. Greenberg, G. (2013). Beyond Hyman. *Philosophical Review*, 122(2), 215-287.
47. Gregoriou, G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, 324, 1207-1210.
48. Gregoriou, G., Paneri, S., & Sapounzis, P. (2015). Oscillatory synchrony as a mechanism of attentional processing. *Brain Research*, 1626, 165-182.
49. Han, S., Ding, Y., & Song, Y. (2002). Neural mechanisms of perceptual grouping in humans as revealed by high density event related potentials. *Neuroscience Letters*, 319, 29-32.
50. Haugeland, J. (1981). Analog and analog. *Philosophical Topics*, 12(1), 213-225.
51. Haugeland, J. (1987). An overview of the frame problem. In Z. Pylyshyn (ed.,) *The Robot's Dilemma: The Frame Problem and Artificial Intelligence*. Norwood, N.J.: Ablex Publishing Company.
52. Heck, R. G. Jr. (2007). Are there different kinds of content? In B. P. Mclaughlin & J. Cohen (Eds.,) *Contemporary Debates in the Philosophy of Mind* (pp. 117-139). Malden, MA: Blackwell.
53. Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of distinct entities. *Cognition*, 85(3), 203–221.
54. Hopkins, R. (1995). Explaining depiction. *The Philosophical Review*, 104(3), 425-455.
55. Hopkins, R. (1998). *Picture, Image, and Experience*. Cambridge: Cambridge University Press.
56. Hyman, J. (2006). *The Objective Eye*. Chicago, ILL: Chicago University Press.
57. Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
58. Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neuroscience*, 37(7), 357-369.
59. Kerkoerle, van T., Self, M. W., Dagnino, B., Gariel-Mathis, M-A., Poorta, J., van der Togta, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *PNAS*, 111(40), 14332-1434.
60. Kosslyn, S.M. (1994). *Image and Brain*. Cambridge, MA: The MIT Press.
61. Kulvicki, J. V. (2006). *On Images: Their Structure and Content*. Oxford: Clarendon Press.
62. Kulvicki, J. V. (2014). *Images*. New York, N.Y: Routledge.
63. Kulvicki, J. (2015). Analog representation and the parts principle. *Review of Philosophy and Psychology*, 6, 165-180.
64. Lyons, J. C. (2022). Three grades of iconicity in perception. *Asian Journal of Philosophy*, 1(50), <https://doi.org/10.1007/s44204-022-00039-x>.
65. Markov, N. T., Vezoli, J., Chameau, P. Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Giroud, P., & Uhlmann, S. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *The Journal of Comparative Neurology*, 522, 225-259.
66. Marr, D. (1982). *Vision*. New York, N.Y.: Freeman.
67. Michalareas, G., Vezoli, J., Pelt, S. V., Schoffelen, J-M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, 89(2), 384-397.

68. Miller, E. K., & Cohen, J. D., (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
69. Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004a). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perceptual Psychophysics*, 66(8), 1268–1281.
70. Mitroff, S. R., Scholl, B. J., Wynn, K. (2004b). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science*, 15, 420–425.
71. Nakatani, H., & van Leeuwen, C. (2006). Transient synchrony of distant brain areas and perceptual switching in ambiguous figures. *Biological Cybernetics*, 94, 445–457.
72. Newall, M. (2011). *What is a Picture?* Hampshire: Palgrave Macmillan.
73. Norton, J. D. (2011). Challenges to Bayesian confirmation theory. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Philosophy of Statistics* (pp. 391–439). Handbook of the Philosophy of Science 7. Amsterdam: North-Holland.
74. Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (eds.) (pp. 259–303), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.
75. Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Science USA*, 107, 1580–85.
76. Peacocke, Cr. (1986). Analogue content. *Aristotelian Society Supplementary Volume*, 60(1), 1-18.
77. Peacocke, C. (2001). Does perception have a nonconceptual content? *Journal of Philosophy*, 98(5), 239–269.
78. Peacocke, Cr. (2019). *The Primacy of Metaphysics*. Oxford: Oxford University Press.
79. Peirce, C. S. (1977, 1903). Logic as semiotics: the theory of signs. In J. Buchler, (ed.), *The Philosophical Writings of Peirce*, pp. 98-119 (1955). New York, N.Y: Dover.
80. Pianesi, F., & Varzi, A. C. (1996). Events, topology and temporal relations. *The Monist*, 79(1), 89–116.
81. Planer, R. J. (2021). What is symbolic cognition. *Topoi*, 40, 233-244.
82. Plomp, G., Quairiaux, C., Kiss J. Z., Astilfi, L., & Michel, C. M. (2014). Dynamic connectivity among cortical layers in local and large-scale sensory processing. *European Journal of Neuroscience*, 40, 3215-3223.
83. Quilty-Dunn, J. (2016). Iconicity and the format of perception. *Journal of Consciousness Studies*, 23(3-4), 255-263.
84. Quilty-Dunn, J. (2020). Is iconic memory iconic? Forthcoming in *Philosophy and Phenomenological Research*, 101(3), 660-682.
85. Rescorla, M. (2009). Predication and cartographic representation. *Synthese*, 169, 175-200.
86. Roux, F., & Uhlhaas, P., J. (2014). Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information? *Trends in Cognitive Science*, 18, 16–25.
87. Sainsbury, R. M. (2005). *Reference without Referents*. Oxford: Oxford University Press.
88. Shepard, R. N. (1978). The mental image. *American Psychologist*, 33(2), 125-137.
89. Shin, Sun-Joo (2003). *The Iconic Logic of Peirce’s Graphs*. Cambridge: Cambridge University Press.
90. Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14, 29–56.
91. Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99, 605–632.
92. Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants’ knowledge of object motion and human action. In D. Sperber, & D. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* (pp. 44–78). Oxford: Clarendon Press.
93. Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA: MIT Press.
94. Sterelny, K. (2020). Afterword: tough questions; hard problems; incremental progress. *Topics in Cognitive Science*, 12(3), 766-783.
95. Tallon-Baudry, C., Kreiter, A., & Bertrand, O. (1999). Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Visual Neuroscience*, 16, 449–59.
96. VanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, 14, 498–504.
97. Varzi, A. C. (2016). Mereology. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2016 edition.

98. Werning, M. (2012). Non-Symbolic compositional representation and its neuronal foundation. In M. Werning, W. Hinzen, & E. Machery, (Eds.), *The Oxford Handbook of Compositionality* (pp. 633-724). Oxford: Oxford University Press.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

---

<sup>i</sup> Kulvicki (2014, 57) defines *Outline shape* and *Occlusion shape* as follows: *Outline shape* is the pattern of angles that the rays emanating from an observer (considered as a point source) make as they touch the outlines of the object from that point in space. *Occlusion shape* is the flat pattern that the object forms on a planar surface that is placed perpendicularly to some line of sight cutting the rays that were passing through the space that the plane occupies.