The Representational Formats of Cognition and Visual Perception and their Interface: Part 1

Athanassios Raftopoulos (raftop@ucy.ac.cy)
Department of Psychology, University of Cyprus
P.O.BOX 20537, Nicosia 1678, Cyprus

Abstract

I examine the representational formats of perceptual states and cognitive states related to perception, such as perceptual beliefs stored in long term memory, and argue, first, that despite their important differences they both have an iconic ingredient. Then, I explain how this common iconic component of perceptual and cognitive contents allow cognitive states to modulate perceptual processing focusing on a recent argument made by Burnston (2017) to the effect that owing to their differing representational formats cognition cannot affect directly perception.

Keywords: analog/symbolic representation; representational formats in memory; cognition/perception interaction

Introduction

Cognitive and perceptual states are held by many (Ayers 2019; Burge 2010; Burnston 2017; Carey 2009; Crane 2003; Dretske 1981; Fodor 2007; Goodman 1976; Heck 2007; Haugeland 1987; Jackendoff 1987; Peacocke 1986, 2019) to be cast in different representational formats, namely, digital or symbolic, and analog formats respectively.

Assuming this distinction, if one holds that cognition and perception interact the problem immediately emerges as to how this interaction could take place in view of the differing representational formats. (I do not discuss whether the analog/digital distinction of states is parallel to the analog/digital distinction of contents. I assume this since my arguments can to accommodate both cases.)

Burnston (2017) has argued that cognitive states cannot penetrate perceptual processes in the way this interaction is usually thought, i.e., by the *Internal Effect View* (IEV)

> P is penetrated if, over a specific input, it would perform a certain computation C leading to content R1 in the absence of a cognitive state, S, but performs a different computation C2, yielding content R2, when S is present, where the causal, semantic coherence, and computation conditions are met.

In IEV, cognitive states modulate perception when they modify the computations performed by perceptual processes. In this case, the cognitive states causally affect the perceptual states; the causal condition. Finally, the content of the modified perceptual state should be intelligibly related to categorical facts concerning the content of the penetrating cognitive state; the semantic coherence condition. Consider the putative case of CP of color perception in the Delk & Fillenbaum (1965) experiment. Participants were shown paper-cut objects presented in an orange-red color and when asked to match the background color to that of the presented object tended to make the background a more saturated red for stereotypically red objects (hearts) than for stereotypically non-red objects. Perception should receive as input the orange-red and the participants should match the background orange-red color, R1, but owing to the penetrability of the perception by the belief that hearts have a saturated red color, choose a more saturated red R2. Let us call this, the direct effect of cognition on perception.

Burnston (2017) argues that owing to their different representational format, cognitive states cannot affect perceptual states that have purely iconic contents, in the way described by IEV. Owing to their symbolic format, cognitive states lack the necessary structure that would allow them to modify perceptual computations. Cognitive representations are discrete or atomic, have no referentially relevant internal structure, and their contents do not specify any properties of their referents. This is why they cannot affect perception whose iconic states have a rich internal structure that maps naturally to the representatum.

Burnston thinks that the cognition/perception interaction is described by the *External Effect View* (EEV). Accordingly, cognitive effects do not modify the perceptual computations but, instead, change or bias the distribution of probabilities of all possible perceptual processes that could be applied to a stimulus so that the perceptual processes associated with the concept(s) figuring in the affecting cognitive contents have their probability of being applied to that stimulus increased (indirect cognitive effect).

I have argued (Raftopoulos 2009; 2019) that cognitive states affect perception directly in the way described by IEV. I defend here this view in the face of Burnston's objections in two parts. The first, which concerns the formats of perceptual and cognitive representations is developed here. The second is about the mechanism by means of which cognitive states modulate perceptual processing. I sketch an account of this issue. By perception I mean 'early vision', because it is there that cognition interfaces with purely iconic perception.

The cornerstone of my defense is that cognitive states that are thought to affect directly perceptual processing, namely perceptual beliefs, have an iconic ingredient, in addition to their symbolic/conceptual content. This means that they are hybrid states that have the requisite structure to be mappable in systematic, natural ways to their representatum and this structure allows them to affect directly perceptual processing.

Digital Representations

For Jackendoff (1987, 181-2), symbolic representations, which he calls algebraic representations, represent their objects by means of arbitrary symbols that have no bearing to the physical appearance of the objects. In algebraic representations, spatial relations are formal relations between pairs of symbols and, thus, have a formal character that does not distinguish them from other representations that happen to have the same formal character; the algebraic representation x *is to the left of y* is parallel in formal character to the representation *x possesses y*.

According to Goodman (1976), a representation is symbolic or digital if it contains discrete symbols, signs that refer through a convention ('cat', for example, refers to a certain kind of animals through an agreement of a linguistic community). A symbolic notation is discrete or differentiated, according to Goodman (1976, 148-152), if it is semantically and syntactically disjoint, as opposed to semantically and syntactically dense. A representational system or scheme is differentiated if "for every two characters K and K' and a mark m that does not actually belong both, determination either that m does not belong to K or does not belong to K' is theoretically possible." (Goodman 1976, 135-136)

Thus, the first property of symbolic representations is that they are necessarily discontinuous and non-dense. Secondly, since symbols refer only through some convention, any composition of symbols that is also a (complex) symbol refers through conventions and does not bear any other relation to its referendum. Consider the symbol/concept 'CAT'. 'CAT''s structure is that of a simple concatenation of less complex symbols. 'CAT' refers solely by convention, and, so, no part of 'CAT' refers to cat body parts or to their features and there is no natural correspondence between the constituents of the representing symbolic structure and the body parts and features of a cat. "The word dog, in contrast (the contrast is to iconic representations, my comment) contains no information about ears or any other part of a dog." (Carey 2009, 135).

Third, symbolic representations have canonical decomposition and have syntactic structure because they consist of distinguishable discrete parts. A symbolic representation is compositional because its syntactic structure is determined by the syntax of its parts and the syntactic features that are used in the composition. The complex symbol p&q, whose truth value depends on the truth values of p and q and the way the logical connective '&' functions, is such an example. According to Fodor (2007), symbolic/conceptual representations are discursive and can be recombined the right sort of way. Symbolic representations allow the representation of logical connectives and quantifiers. Thus, states with symbolic contents can recombine in systematic ways to produce new states. This is closely related to the fact that symbolic contents are conceptual, and the recombinability of concepts to form new thoughts is taken for granted.

Fourth, the discreteness of symbolic representations is important for their function as concepts. Concepts are like on-off switches; 'on' means that a token belongs to a certain type, and 'off' means that the token does not belong to that type. For any token K it is theoretically possible to determine whether K belongs to the type L or not. This is possible because symbolic representations are differentiated.

Analog or Iconic Representations

Those who think that perception and cognition have different representational formats hold that while cognitive states have symbolic contents, purely perceptual states have analog or iconic content. The relationship between analog and iconic is complicated and depends largely on how one understands 'analog'. All analog representations are iconic, but whether iconic representations are analog is debatable.

Let us start with the minimal conception of analog as an iconic representation. A representation is iconic if it has an inherent structure that maps naturally onto the structure of the represented entity. This entity can be an object, in which case the iconic representation represents it through some mapping of shapes (similarity between the shape of the representing entity and the shape of the represented entity) or some other characteristic feature. Or, it can be a visual scene in which case the scene is represented through a mapping from the spatial structure of the representational content to the spatial structure of the visual scene, and through mappings from the attributives (the representations of the features of objects in the scene) to the attributes in the visual scene (although it is not necessary that the mapping contains all of the features in the visual scene). As Jackendoff (1987, 181-2) explains, 'In a geometric representation 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.' Thus, the relations between aspects in the representation are not arbitrary but depend on the relations between the aspects of the visual scene onto which they map; an analog representation approximates the representatum (Peacocke (1986).

It is an essential characteristic of the iconic structure of perceptual representations that it does not support logical operations. Logical connectives and quantifiers cannot be among the analog representational content of perception, as they can be part of the content of propositional states. This can be inferred from two facts. First, that there are no logical contradictions in perception (illusions are not logical contradictions), while a proposition whose form is p & -p is a logical contradiction. Second, from the fact that if one tries to take a picture of a situation expressed by a disjunction, say that O1 is either to the left of O2 or to the right of O2, one gets a picture either of O1 being to the left of O2, or a picture of O1 being to the right of O2, depending on the actual spatial configuration. This, however, is not a picture that displays the disjunctive fact described above; one cannot analogically express the fact of the occurrence of a logical connective.

Ayers (2019, 77-78) notes that pictorial 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'.

According to Carey (2009), analog representations are iconic in that their parts represent parts of what the representation as a whole represents, which Fodor (2007, 173) calls the *Picture Principle*, 'if P is a picture of X, then parts of P are pictures of parts of X.' Kosslyn (1994, 5) holds the same view about iconic or depictive representations. Iconic representations have a whole/part structure, as opposed to the compositional structure of symbolic representations.

There are several problems with this proposal. Although it makes sense to talk about spatial parts when the representation itself has spatial properties (as pictures, and maps have), there are iconic representations that do not involve spatial parts properly speaking because they are not arrayed in actual space. This means that the *Picture Principle* cannot be a necessary condition for iconic representations (Beck 2019; Blachowicz 1997; Peacocke 2019). This has repercussions for the imagery debate. The problem that the parts principle, as the *Picture Principle* is also called, poses for mental imagery is that according to Kosslyn (1994) the *Principle* applies to mental images, but mental images are quasi-pictures because they are not arrayed in space, as literal pictures are, and it is not clear in which sense one could talk of parts in non-spatial representations. Kosslyn's (1994, 5) introduces the notion of 'functional space', according to which the space in which mental images appear need not be physical but can be like an array in a computer that specifies spatial relations purely functionally. This means that the physical locations of each element in a spatial array are not themselves arranged into an array and it is only 'in virtue of how this information is read and processed that it comes to function as it were arranged into an array'.

Mental images need to be 'physically instantiated in a way such that they can "act" or "function" like images in an appropriate 'processing" environment.' (Haugeland 1987, 91) A representation may be iconic even if it is not arrayed in space and does not have spatial features, since it suffices that it determine a geometrical structure that maps naturally onto some spatial structure. The view that representations in perception or imagery even though not literally arrayed in space can be iconic representations of spatially arrayed properties comes from our knowledge of the topologies involved in perception and their inter-mappings. Recall that a representation is iconic if it has an inherent structure that maps naturally onto the structure of the represented entity. The iconic nature of perceptual representations is grounded successively in the layout of the retinal cells that maps onto the spatial layout of the environment, and in the orderly retinotopic mapping of the visual world onto the surface of the cortex through the retinotopic mapping of the surface of the cortex onto the retinal cells. The physical layout of the retinal cells and the other receptors higher in the hierarchy of the brain renders registration of information from the retinal image iconic. The iconic registration of the retinal image maps onto representational states in the brain rendering them in turn iconic, and both map onto to visual perceptual representation in experience rendering it iconic as well. All these mappings are grounded in the mapping of the topology of information registration in the retina onto the topology of spatial and featural structures in the environment and this results in perceptual representations that preserve the spatial and featural structure of the represented visual scene.

Perception, thus, is a primary candidate for an iconic scheme that satisfies the *Picture Principle* because perceptual representation is arrayed in space and, thus, is amenable to part/whole structural analysis. Note, however, that there are some limitations to what constitutes a proper part of an image. Not all combinations of features could be considered genuine parts of the image. Consider the back part of an object and a part of the immediate background and combine them. In terms of what is computationally relevant in perception, it is highly unlikely that this complex part of the image is represented by anything in perception. Thus, it is not true that any part of an iconic representation represents a part of the image that the representation represents; only parts that are admissible as components of perceptual processes are admitted.

Some iconic representations are dense, continuous, homogeneous, unit fee, and come in information packages, a set of properties traditionally assigned to analogicity (analogtr). A set is dense if between any two elements in the set there is always a third element; the set of real numbers is dense but the set of natural numbers is not because between two consecutive natural numbers there is not a third number. In the brain, some neurons have continuous activation functions, which means that the set of the activation values of a neuron is dense. Consider a neuronal assembly that represents the color red and has continuous activation values. This entails that 'red' may be represented by a continuous, dense set of activation values, some subsets of which determine the different shades of red (deep red, bright red, etc.,) but others have no phenomenological cash values. Or, consider a mercury thermometer in which the magnitude of mercury represents temperature. Both the representing magnitude and the represented temperature vary continuously and are dense.

Peacocke (1986) argues that analog magnitide representations are unit free (see also Jackendoff (1987)), while digital representations of magnitudes come in some unit. Dretske (1981, 73) argues that all signals are "pregnant with information"; they carry many messages when they carry any. A picture of a red cup not only carries information about its color, but also about its shape, etc.; one cannot see a red cup without seeing it as having some specific shape, etc. By contrast, you can form the belief that a cup is red without forming beliefs about its shape.

Blachowicz (1997) examines the properties that analogtr representations are supposed to have and concludes that many analog representations exhibit all these properties (see also Beck 2019), but, excepting, 'relational identity' they are not necessary to being an analog representation, which means that if a representational scheme satisfies relational identity should be considered to be analog despite the fact that it is not continuous or dense. Reference to a similar condition for analog representations is found in Beck (2019, 331-333), according to whom, a representation is analog if it mirrors (that is, it is isomorphic to, or bears some structure-preserving relation toward) what it represents; similarities among the elements in the represented domain are mirrored by similarities among the elements in the representational scheme.

Maley (2011, 122) offers a covariational account in which a representational medium R of a domain Q is analog just in case there is some property P of R such that the quantity of P determines Q and as Q increases or decreases by an amount d, P increases or decreases as a monotonic function of Q + d or Q-d. This demand is further developed by Kulvicki (2015).

Kulvicki argues that analog representations are those that bear a certain mirroring relationship to the domain they represent, a requirement that may be satisfied by non-dense representational schemes. Analog representations require structure preserving syntactic-semantic links (syntactic refers to the representing medium, while semantic refers to the represented domain) that result in representations with vertically articulated content. A representation has a vertically articulated content when it represents objects as being P but also represents them as being Q, where Q is an abstraction from P. A mercury thermometer is such a representation because it represents temperatures and when it designates a certain temperature T1 through the measurement of some measured height of the mercury, it also represents indefinitely many abstractions from T1, that is, other temperatures that correspond to heights that are very close to the measured type that, as such cannot be discriminated from that height.

In these accounts, the traditional properties assigned to analog representations are dropped and analogicity is defined in terms of an appropriate mapping of the representation onto the represented domain that captures semantical properties and relations in the represented domain. Thus, the defining character of analogicity is the iconic character of the representation. Let us call this the revised view of analogicity (analogr). Discrete representational systems could count as analogr. I will not delve on whether perception, which is iconic, has some of other properties of analogtr because whether cognitive states could affect perceptual processing hinges on the iconic nature of perceptual representations.

Representations in Visual Perception, VSTM and LTM

Research supports the distinction between dense, purely iconic perceptual representations, on the one hand, and the less dense hybrid representations in VSTM (Visual sort term memory) and VLTM (Visual long term memory) (should one espouse the view that these representations are not purely symbolic and may have an iconic component), or the purely symbolic representations used in VSTM and VLTM supporting conceptual thought, on the other hand. Since attention is usually involved in storing information in VSTM and VLTM ('usually' because information can get into LTM in the absence of attention and on certain occasions bypassing working memory), the attentional modulation of the output of early vision restricts not only the number of objects that can be held in memory, but also impoverishes the information about those objects that is stored in memory. It is thought that iconic representations are high-density representations in the order of 100,000 bits of information (Itti & Baldi, 2005), while the representations in VSTM have much lower density, about 30-40 bits of information (Vogel et al. 2001).

Coding of purely iconic representations in early vision is done through basis functions. These representations are modal and represent by means of dense basis functions that work at the early perceptual levels. A color, for instance, is represented by a vector or a pattern of activation values (scalars that represent the relative activity of red, green, and blue) across columns of neurons that distributively represent colors. The basis functions in early vision are dense in the sense that the relevant neuronal activations that realize the representations take continuous values. Thus, the states in early vision are represented by neural firing rates that vary continuously. In this sense, the representations in early vision are analog in the traditional sense, they are homogeneous, and satisfy the *Picture Principle*.

Not all neurons in the brain have continuous activation functions. Some neurons fire a fixed amount once they reach a threshold or do not fire at all. In this case, the properties of objects in the visual scene could be represented either by the total number of neurons firing above a certain threshold (Beck 2019, 333), or by the firing patterns of these neurons. Either way, both the medium and the representational format are discrete. The representations, thus, are not analogtr, but they are still iconic since they preserve the spatial structure of the perceived visual scene and since the elements in the representation map onto elements in the represented scene. In addition, the coding in early vision, even if non-dense has a higher-density than the coding in VSTM or VLTM. Moreover, the iconic, non-continuous representations in early vision are purely iconic and do not contain any symbols. One could say that in this case the representations are near-continuous.

A few words about basis functions are in order. In mathematics, a basis function is an element of a particular basis for a function space. Every continuous function in the function space can be represented as a linear combination of basis functions, just as every vector in a vector space can be represented as a linear combination of basis vectors. So, in the case of color representations, where colors constitute the function space, the activities of neurons that represent red, green, and blue are the basis functions and every color can be represented by a combination of the activation values across neuronal assemblies that distributively represent red, green, and blue, that is, in terms of the relevant basis functions. Note that in this case, the components in the brain that realize the basis functions are modeled as neurons in the relevant neuronal assemblies and, given the fact that representations in the brain are distributed, each neuron can be a component of different basis function (Pouget and Snyder 2000). The idea is that by taking a weighting sum of three basis functions of the visual signals, one obtains the color (hue, saturation, and lightness) at a particular location in a visual scene.

Let me elaborate. Color processing starts with three kinds of cone cells/receptors in the retina that respond to short (S), medium (M), and long (L) wavelengths (the proportions of these three kinds of cells in the retina is L:M:S/ 10:5:1). The three types of cells have Bell-like overlapping response curves, or spectral sensitivities, with peaks at 440, 530 and 560nm respectively (De Valois & De Valois 1993), and are usually called the blue, green, and red cones. Although this is misleading, since the peak of the third type of cones is at 560 nm but the red color spectrum is between 600 and 700nm, I will keep this nomination. These three cone spectral sensitivities can be thought of as the three basis functions out of which all other spectra could be constructed, which entails that each color can be described by three numbers, the values across the three basis functions to which each color corresponds (I will explain next what "corresponds" means). This, in turn, signifies that each color corresponds to a vector in the three-dimensional space defined by the three basis functions.

Let N be the number of all possible spectral signals. These signals define a N-dimensional space and every environmental spectral signal is N-dimensional. As a given signal projects onto the retina it is filtered by the three basis functions (that is, the three types of spectral sensitivities of the retinal receptors), in the sense that the incoming signal activates these cells in proportion to its spectral signature and these cells, in turn, output their signal to LGN and V1 cells. So, all the information in the incoming signal is filtered by the responses of the three types of cells in the retina and propagates further in the visual cortex. For each incoming signal, each type of cell responds with a specific output, a number corresponding to a wavelength in its sensitivity curve, and the three numbers (since there are three cone types/basis functions) encode in the retina the incoming signal. This can be thought of as a projection of the incoming signal, through the three basis functions, onto a three-dimensional subspace. Thus, if we measure the responses of the three different spectral sensitivity curves/basis functions, we are measuring the projection of the N-dimensional input vector onto each one of the dimensions of the three-dimensional space defined by the three basis functions. (This is the sense of "corresponds that I used in the previous paragraph.) The three coordinate values (the triplet of cone responses) defines a specific coordinate in the three-dimensional space, which is the projection in three-dimensional space of the N-dimensional input signal. It follows that any color system can be defined by simply specifying a set of three linearly independent basis functions and record the projection of any color in the system onto the thus-defined three-dimensional space as it is filtered upon its projection onto the retina by the basis functions.[1]

It is important to point out that through the basis functions, N-dimensional environmental inputs are projected in a systematic way onto brain states, in such a way that despite the reduction in the dimensionality of the signal, from N dimensions to the three dimensions of the filtering basis functions, the structure of the represented domain, the structure of the color space for instance, is preserved in the representing states since similarities in the represented domain are preserved in the representing medium; similarities in the represented domain are translated to closeness in the representing medium so that the representations of similar input signals form clusters, which in terms of experiential content correspond to phenomenal color spaces. Since a representation is iconic if it has an inherent structure that maps naturally onto the structure of the represented entity so that similarities among the elements in the represented domain are mirrored by similarities among the elements in the representational scheme, the basis functions provide a means by which the brain can built iconic representations of some environmental structures. Moreover, as we shall see later on, despite the reduction in dimensionality owing to the projection onto the retina, all the information in the input is retained through the Garbor functions performed by V1 cells.

VSTM coding is done by means of basis functions as well, but these functions are sparser since the relevant activations do not take continuous or near continuous, but discrete values. VSTM codes of colors, for example, concern categories like 'red', 'light' etc., but lacking a dense structure they do not encode the fine color information regarding hues, intensities, etc., that is available to low-level color channels. Thus, information stored in VSTM does not allow the fine discriminations made available via low-level color channels and the representations in visual areas differ from the representations stored in VSTM.

It is debatable whether the representations in VLTM function as descriptors that code in a symbolic all or nothing manner, or by means of sparse basis functions of the short used in VSTM. VLTM cannot store information in a richer format than that of VSTM, although it can store more information. Symbols/concepts are stored in LTM. Concepts are in a sense on-off switches, the 'on' meaning that a token belongs to a certain type (concept), and the 'off' meaning that the token does not belong to that type. It is, therefore, possible that representations in LTM function as descriptors.

The role of representations in LTM as concepts is compatible with the possibility that these representations function by means of sparse basis functions. If VLTM store information by means of basis functions, the information stored may be described as symbolic in the sense that the basis functions concern types of visual object-features that form a discrete set of values, as opposed to the continuum of values of the basis function used in iconic information. The representations stored in memory, apparently owing to attentional filtering, do not contain information about, say, determinate hues, but only information about the category of the color (bright red). By being symbolic, the information stored in VLTM representations enables categorization since the representations abstract away much of the detailed iconic information and allow different tokens that differ in various features to be subsumed under the same type.

Could VLTM representations be rich albeit less rich than iconic perceptual representations, and in what sense? Since it is debatable whether VLTM representations function as descriptors that code in an all or nothing manner, or by means of sparse basis functions, if the latter they may encode rich visual information from a visual scene perceived in the past and their role exceeds their function as descriptors. Representations in VLTM could store information iconically. To shed light on this problem, let us examine the research conducted by Hollingworth et al.

Hollingworth and Henderson's (2002) research on the role of LTM in retaining information from objects attended in after attention is withdrawn provides evidence that performance in online change detection and discrimination tasks is mainly supported by the maintenance of visual object representations stored in LTM during the online perceptual processing of the scene, rather than in VSTM. This conclusion is further reinforced by Hollingworth et al. (2001) who found that online change detection performance is strongly influenced by the semantic consistency between that target object and the scene in which it appeared, a variable which is known to influence the representation of an object in LTM but not in VSTM.

This research supports the following. First, when attention is oriented to an object, in addition to low-level perceptual processing, visual processing leads to the construction of representations at higher levels of analysis. These may include a visual description of the attended object abstracted from low-level sensory properties, and conceptual representations of object identity and meaning. Higher-level visual representations can code detailed information about the visual form of an object, specific to the viewpoint at which the object was observed. Second, these abstracted representations are indexed to a position in a map coding the spatial layout of the scene forming an object file that preserves abstracted visual representations rather than sensory information and supports the short-term retention of conceptual codes. Third, processing of abstracted visual and conceptual representations in VSTM and the indexing of these codes to a particular spatial position leads to their consolidation in LTM. LTM codes for objects are likewise indexed to the spatial position in the scene map from which the object information was encoded, forming LTM object files. Fourth, when attention is withdrawn from an object, VSTM representations decay rapidly leaving only the spatially indexed, LTM object files that are relatively stable. The fact that changes to objects on the saccade away from that object are often detected immediately (Hollingworth et al. 2001) suggests that visual object representations can be retained in VSTM at least briefly after attention is withdrawn from an object. Fifth, the retrieval from LTM of higher-level visual codes specific to the viewed orientation of a previously attended object accounts for the ability to detect token and rotation changes and to perform accurately on orientation-discrimination tests.

Hollingworth (2004) provides strong evidence that the online representation of previously attended objects is supported by the same form of representation supporting long-term object memory. These data provide support for the visual memory theory of scene representation, according to which, as the eyes and attention are oriented from object-to-object within a scene, higher-level visual representations of attended objects are activated, are maintained briefly in VSTM, and are consolidated into LTM. VSTM representation is soon replaced as attention moves on to other objects, but higher-level visual representations of previously attended objects accumulate in LTM, forming a robust and relatively detailed representation of the scene.

There is dissociation between VSTM and VLTM and sensory persistence (iconic memory) in terms of format and content (abstracted vs. sensory–pictorial), capacity (limited vs. large capacity), and time course (relatively robust vs. fleeting). On another occasion, Hollingworth remarks that his research provides evidence of representations in LTM with similar format and content as those in online perception. There is also a strong conceptual involvement, including semantic associations, in VLTM. Hollingworth's study demonstrated that visual memory is sensitive to semantic associations, task demands, and viewer strategies.

Hollingworth and colleagues' research is not the only ones to suggest a picture of VSTM according to which VLTM stores an abundance of detail of visual scenes encountered in the past. Even though current models of visual perception posit a hierarchy of processing stages that reach more and more abstract representations in higher-level cortical areas in the brain, many studies in addition to the ones discussed in the previous paragraphs suggest that visual processing stages in the brain do not necessarily discard visual details by showing that participants successfully maintained detailed representations of thousands of images (Brady et al. 2008). It seems, thus, that VLTM representations can contain not only gist information, as is traditionally thought, but also details sufficient to discriminate between exemplars and active visual states. 'Whereas in everyday life we may often fail to encode the details of objects or scenes, our results suggest that under conditions where we attempt to encode such details, we are capable of succeeding.' (Brady et al. 2008, 14328). To be able to maintain such featural details, the representations of objects and their features in VLTM might be stored throughout the entire hierarchy of the visual processing stream, including early visual areas that are reactivated on demand by means of top-down processes (Ahissar and Hochstein 2004; Wheeler at al. 2000).

If VLTM stores information not in an all or nothing descriptive manner but by means of basis functions, it is possible that the representations in VLTM have in iconic format, which is more abstract than the iconic format of visual representations. To be iconic, they have to satisfy only the condition that their structure capture semantical properties and relations in the represented domain. Note, however, that owing to the presence of conceptual elements, the format of VLTM representations cannot be purely iconic. It follows that the VLTM representations have a hybrid format; symbolic, owing to the conceptual ingredient, and iconic, owing to the perceptual ingredient. Since, unlike perception, LTM also stores cognitive information in propositional form, it is likely that the representations in VLTM are cast in a propositional form that contains iconic elements, which means that concepts may take advantage of the iconic information.

Early Vision/Cognition Interaction: A Sketch

The cornerstone of the view that cognitive states cannot affect directly perceptual states is that the symbolic cognitive states do not have the requisite structure to modulate iconic perceptual states. We saw that cognitive states in visual memory may be hybrid states that have both conceptual components and iconic components. Let us return to Delk and Fillenbaum (1965). The participants, recognizing the paper cut surface as a heart, activate in memory their knowledge about hearts, among which the most relevant to the task at hand is the belief that hearts typically have a bright red color. This representation is effectuated by means of a sparse function. Assume that in VLTM the typical red color of a heart is represented by a triplet of values coding for 'deep', 'bright', and 'red', <a, b, c>, which is a hybrid symbolic and iconic representation. This symbolic/conceptual representation maps onto the relevant perceptual space not simply in the sense that the color is in a certain color-range, but also in the sense that owing to its iconic elements this representation maps, partly, onto a region in the phenomenal similarity color space.

Representations in memory have an iconic component and partition the phenomenal color space in a certain way and the components of the sparse function that represents the belief map naturally onto this space. The symbolic-phenomenal space mapping is natural since it is the same type of functions that underlies both iconic representations of colors in pure perception and hybrid representations of colors in VSTM, and since the concepts in the VSTM are formed by processes that are partly guided by the stimulus. It is the nature of the stimulus that elicits the relevant concept and is not arbitrary which concept will be activated.

The iconic component of memory color representations endows them with structure that maps onto the phenomenal color space through their mutual mappings onto environmental colors. This structure allows, in turn, memory representations to modulate perceptual processing of colors, through a mechanism currently investigated.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science, 8*, 457–464.

Ayers, M. (2019). *Knowing and Seeing: Groundwork for a New Empiricism*. Oxford: Oxford University Press.

Beck, J. (2019). Perception is analog: the argument from Weber's law. *Journal of Philosophy*, 0022362X/19/1606/319–349.

Blachowicz, J. (1997). Analog representation beyond mental imagery. *Journal of Philosophy, 94*(2), 55-84.

Brady, T., F & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision, 15*(15), 1-24.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Burnston, D. (2017). Cognitive penetration and the cognition-perception interface. *Synthese, 194*, 3645-3668.

Carey, S. (2009). *The Origins of Concepts*. Oxford, N.Y.: Oxford University Press.

Crane, T (2003). *The Mechanical Mind*. London: Routledge.

Delk, J. L. & Fillenbaum, S. (1965). Differences in perceived colour as a function of characteristic color. *The American Journal of Psychology*, *78*(2): 290-93.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: The MIT Press.

Fodor, J. (2007). The revenge of the given. In B. P. McLaughlin and J. Cohen (Eds.), *Contemporary Debates in the Philosophy of Mind*. Malden, MA: Blackwell.

Goodman, N. (1976). *Languages of Art*, 2nd edition. Indianapolis: Hackett.

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.

Heck, R. G. (2007). Are there different kinds of content? In B. P. Mclaughlin & J. Cohen (Eds.) *Contemporary Debates in the Philosophy of Mind*. Maden, MA: Blackwell.

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm. *Memory & Cognition, 29*, 296–304.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113-136.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 519–537.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. In *Proceedings of the IEEE Conference*, 1, (pp. 631–637).

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

Kulvicki, J. (2015). Analog representation and the parts principle. *Review of Philosophy and Psychology, 6*, 165-180.

Maley C. (2011). Analog and digital, continuous and discrete. *Philosophical Studies, 155*, 117-131.

Peacocke, Cr. (1986). Analogue content. *Aristotelian Society Supplementary Volume, 60*(1), 1-18.

Peacocke, Cr. (2019). *The Primacy of Metaphysics*. Oxford: Oxford University Press.

Raftopoulos, A. (2009). *Perception and Cognition: How do Psychology and Neural Sciences inform Philosophy*. Cambridge, MA: The MIT University Press.

Raftopoulos, A. (2019). *Cognitive Penetrability and The Epistemic Role of Perception*. Palgrave, McMillan.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human perception and Performance, 27,* 92-114.

Wheeler, M. E., Petersen S., E., & Buckner R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of National Academy of Science USA, 97*, 11125–11129.