PERCEPTUAL PLURALISM

Jake Quilty-Dunn
Faculty of Philosophy, University of Oxford

Forthcoming in *Noûs* – penultimate draft

Abstract: Perceptual systems respond to proximal stimuli by forming mental representations of distal stimuli. A central goal for the philosophy of perception is to characterize the representations delivered by perceptual systems. It may be that all perceptual representations are in some way proprietarily perceptual and differ from the representational format of thought (Dretske 1981; Carey 2009; Burge 2010; Block ms.). Or it may instead be that perception and cognition always trade in the same code (Prinz 2002; Pylyshyn 2003). This paper rejects both approaches in favor of *perceptual pluralism*, the thesis that perception delivers a multiplicity of representational formats, some proprietary and some shared with cognition. The argument for perceptual pluralism marshals a wide array of empirical evidence in favor of iconic (i.e., image-like, analog) representations in perception as well as discursive (i.e., language-like, digital) perceptual object representations.

§1. Introduction

Most philosophical writing on the nature of perceptual representation focuses on content. For example, philosophers have debated whether perception has conceptual content (Heck 2000; Byrne 2005), high-level content (Bayne 2009; Brogaard 2013), and rich conscious content (Cohen & Dennett 2011; Block 2014a), just to name a few disputes.[1]

This preoccupation with content, though fruitful in many respects, is peculiar. Elsewhere in philosophy of mind and cognitive science, debates center around the *vehicles* of

[1] There is also fierce debate about whether perception has content at all (Brewer 2006; Schellenberg 2011). I will not argue for the assumption that perception is representational except indirectly, by demonstrating the explanatory benefits of positing particular types of representational structure.

content—mental representations themselves, and how they, rather than their contents, are structured. For example, debates about whether there is a language of thought (Fodor 1975; Dennett 1978), whether some thoughts are map-like (Camp 2007; Rescorla 2009), whether concepts are sensory-based (Prinz 2002; Carey 2009), whether representations are stored in the language faculty (Chomsky 1980; Devitt 2006), whether implicit attitudes are associative (Gawronski & Bodenhausen 2006; Mandelbaum 2016), whether reasoning is formal or model-based (Braine & O'Brien 1998; Johnson-Laird 2006), the structure and development of mathematical cognition (Carey 2009; Dehaene 2011), and the nature of mental imagery (Pylyshyn 2002; Kosslyn et al. 2006) all primarily concern representational vehicles and their structures.

There is good reason for this focus on vehicles throughout cognitive science. Mental representations are the elements of mental computational processes. Contents may be such exotic entities as sets of possible worlds or Fregean senses, while the vehicular properties of mental representations play a direct causal/computational role in the mind.[2] A full understanding of how mental representations function in the mind and produce behavior thus requires a grasp on their representational structures as well as their contents.

In addition to aiding our understanding of the functional role of perception, a theory of how perceptual representations are structured may furnish us with an account of how perceiving differs from thinking. In recent years, a movement has been building that seeks to distinguish perception from cognition by appeal to the structure of perceptual representation. If perceptual representations are structured differently than cognitive representations, then we may be able to draw the border between perception and cognition by appeal to differences in representational structure. This *representational strategy* for drawing the perception–cognition border may help explain phenomenological, epistemological, and functional differences between perceiving and thinking. For example, the apparent richness and fine-grainedness of perceptual phenomenology may reside in the unconceptualized structure of perceptual states (Evans 1982; Tye 2006; Block 2014a; cf. Cohen & Dennett 2011). The view that perceptual representations have a distinct epistemic role from belief (Pryor 2000; cf. Sellars 1956; Siegel 2017) could be explicated in part by their having a structure unlike that of belief (Hopp 2011). And differences in format suggest differences in computational role (e.g., an unconceptualized

[2] Endorsing the truth of this claim does not require denying that mental computations may also be in some sense *sensitive* to the contents of computed representations (Rescorla 2014), nor that some semantic properties may be directly available to computational processes.

perceptual representation might not be able to figure in deductive inference), which may provide a partial characterization of the different functional roles of perception and thought.

Some theorists instead pursue an *architectural strategy*, which aims to ground the perception–cognition border in non-representational aspects of mental architecture, such as the encapsulation of perception from cognition (Firestone & Scholl 2016; Byrne & Siegel 2017; Mandelbaum forthcoming; Quilty-Dunn 2017).[3] While the representational strategy seeks a representation-based distinction between perceiving and thinking, the architectural strategy seeks a distinction based in different types of processes. The architectural strategy therefore allows in principle that perception delivers the same sort of representations used in thought (Pylyshyn 2003). Opponents of the architectural strategy appeal to top–down influences of cognition on perception to undermine the notion of an architectural border between the two systems (Prinz 2006; Clark 2013; Lupyan 2015).

Fred Dretske (1981, 135ff; 2000, 150), Susan Carey (2009, 8), Tyler Burge (2014a, 488; 2014b, 574), and Ned Block (2014b, 560; ms.) all pursue a version of the representational strategy according to which perceptual representations are individuated at least partly in virtue of their *format*. More specifically, they claim that the format of perceptual representation is iconic (or image-like, or analog) while the format of thought is discursive (or language-like, or digital). This paper constitutes a critical evaluation of this thesis. While there is evidence for iconic representations in perception, I will argue for *perceptual pluralism*, the thesis that perception delivers both iconic *and* discursive representations. Though I will not defend the architectural strategy directly, perceptual pluralism undermines the representational strategy pursued by Dretske, Carey, Burge, and Block. The failure of this prominent representation-based approach suggests theorists interested in the perception–cognition border should look to other strategies, including architectural ones.[4]

[3] The architectural strategy can invoke other non-representational factors such as stimulus dependence (Camp 2009; Beck forthcoming), the use of special algorithms, or even simply the use of certain brain areas. Encapsulation has often taken pride of place in architectural approaches to the perception–cognition border, perhaps because of its historical connection to the perennially popular cognitivist approach to the mind outlined by Fodor and Pylyshyn (1988; see also Fodor 1981; Pylyshyn 1984; 1999).

[4] Perceptual pluralism is compatible with non-architectural approaches (whatever they might look like), and even with eliminativist approaches to the perception–cognition border (Shea 2014; Lupyan 2015). However, the arguments in this paper presuppose a border and use experimental evidence to distinguish perceptual processes from cognitive ones.

The arguments below assume some core claims of representational strategists. First, it's assumed that at least some iconic formats are in some substantive way proprietary to perception. That is not meant to preclude the use of iconic representations outside of perception; iconic representations can be used offline in mental imagery, for instance (Block ms.). But I will assume that (e.g.) some icons have a specifically *visual* format, and that this format is distinctly visual even when deployed in non-perceptual contexts such as visual imagery, as is arguably suggested by modality-specific functional interactions between imagery and perception (Kosslyn & Pomerantz 1977; Pearson et al. 2015).[5]

Second, it's assumed in what follows that showing the presence of discursive format in perception is sufficient to establish the form of perceptual pluralism rejected by the representational strategists mentioned above. It is in principle possible that there might be discursive formats that differ from the discursive format of thought. However, none of the theorists who pursue the representational strategy outline what this sort of nonconceptual discursive format might be, and it is yet less clear how such a format might be invoked to explain the evidence detailed below. Moreover, the thesis that perception delivers representations couched in the same discursive format as cognitive representations offers to explain how some perceptual representations feed so quickly and effortlessly into the updating of beliefs and the rational planning of action; the commonality of format would allow cognition to act immediately on the outputs of perception without any intermediating translation mechanism. There is thus some independent reason to think that, if there are discursive representations in perception, then they have the same format as discursive representations in cognition. Indeed, it is a strength of architectural strategies that they allow for a commonality of format between perception and cognition while also providing a principled distinction between the two systems (Mandelbaum forthcoming). I will therefore assume that demonstrating the presence of both iconic and discursive formats in perception provides abductive evidence in favor of perceptual pluralism and against the representational strategy.

[5] It's compatible with this claim, and indeed with the claims of representational strategists more generally, that there are also iconic formats that are in no sense proprietary to perception. There may for instance be cognitive maps that are iconic without being tied to a particular modality. What exactly makes an iconic format proprietarily perceptual is a question of significant interest, but I cannot provide an extensive discussion here. One simple distinction may be that amodal icons like cognitive maps are simply never deployed in perceptual systems—though for theorists who hold that iconicity is the mark of perceptual systems, this claim would have to be qualified to avoid circularity.

§2. Representational Format

Formats are general types of representational structures. The sentence 'This is a tiger' and a picture of a tiger differ in representational format. The phrases 'brown cow' and 'large yawning tiger' have distinct structures but do not differ in format, since they are both instances of English phrases—the difference in representational structure does not mark a difference in format. Different formats are akin to different languages, such that representations that are couched in different formats typically cannot compose together (though new hybrid formats, like hybrid languages, can arise through convention or stipulation in some meta-language). Examples of distinct representational formats include maps, graphs, diagrams, sentences, photographs, hieroglyphs, and blueprints. Some of these formats share relevant structural features, but they all differ enough in how they exploit representational structures and compose representational parts that they each constitute a distinct format.

Perhaps the most influential and fundamental distinction between formats in cognitive science is the distinction between iconic and discursive representations. According to the mainstream cognitivist viewpoint in the past half century, beliefs and other propositional attitudes are discursive (Fodor 1975; 1987; Chomsky 1980; Pylyshyn 1984; Carey 2009; Burge 2010). The systematic and productive ways that concepts seem to recombine into thoughts (Fodor & Pylyshyn 1988), the formal/logical character of deductive inference (Quilty-Dunn & Mandelbaum 2018b), and the word-sized associative links involved in semantic priming (Quilty-Dunn & Mandelbaum 2018a) all call for a theory of propositional thought as literally structured like a sentence or proposition and breaking down into smaller parts that are syntactically concatenated with one another. Though the discursivity of thought remains controversial, I will assume that at least some thoughts are discursive.

Many who believe that thought is typically discursive also believe that there are mental icons, endorsing a pluralist view about mental representation generally (Fodor 1975; Kosslyn 1980; 1994; Dretske 1981; Kosslyn et al. 2006; Camp 2007; 2009; Carey 2009; Rescorla 2009; Burge 2010; Kulvicki 2015; Block ms.). Before surveying the empirical evidence in favor of iconic representations in perceptual systems, it will be useful to characterize the distinction between iconic and discursive formats.

There are two key differences between iconic and discursive representations. The first is that parts of icons correspond to parts of what they represent, while this need not be true of discursive representations. Typically, distance relations between parts of icons correspond to distance relations in what is represented (Kosslyn et al. 2006).

The second difference is that icons represent holistically, i.e., parts of icons encode various properties at once. Icons do not represent by means of a canonical decomposition into constituents that stand separately for distinct individuals and distinct features, while discursive representations may and ordinarily do have such a canonical decomposition.

These principles are not meant to constitute a definition of iconicity or discursivity. My aim is merely to point out some intuitive properties of paradigm cases of iconic representation and show how invoking these properties can explain aspects of perception and imagery. Iconicity as it appears in cognitive science is a natural psychological kind, and thus cannot be defined *a priori*. A full characterization of iconicity will instead emerge from empirically grounded theorizing about perception and imagery.

Consider the difference between a picture of Bob Dylan wearing a checkered shirt (Fig. 1) and the sentence

(1) Bob Dylan is wearing a checkered shirt.

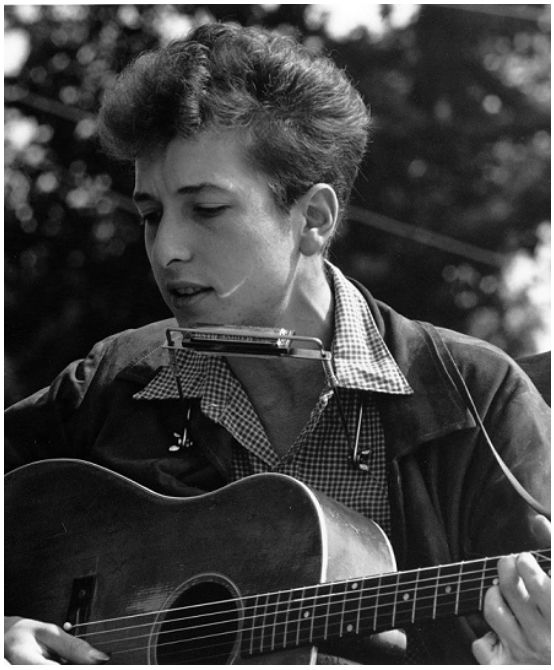


Figure 1—Dylan wearing a checkered shirt

Parts of this picture correspond to parts of Dylan (or his guitar, the background, etc.). The part that corresponds to his shirt also consists of parts that correspond to parts of his shirt. One can point simultaneously to two parts of the picture that correspond to parts of the left side of his collar; one can then move one finger over to a part that corresponds to a part of the

right side of his collar. The increase in distance between ostended parts of the picture corresponds to an increase in distance between parts of the shirt.[6]

The picture also encodes properties in a holistic fashion. The part of the picture that represents the left side of the collar of Dylan's shirt also represents its shape, its visual texture, and its (achromatic) color properties. There is no "canonical decomposition" (Fodor 2007, 108) of the picture into discrete parts each of which can stand uniquely for a particular individual or a particular property.[7] Each part of the icon encodes location values along represented spatial axes as well as features instantiated at that location.[8]

A sentence like (1) lacks these properties. No part of (1) represents part of Dylan (nor do parts of the name 'Bob Dylan'). Distance relations among parts of (1) also do not track distance relations in what is represented. And finally, some parts of (1) represent individual items or properties without being holistically bound up with other items or properties. For example, 'checkered' represents a surface texture without representing the object that has that texture, its specific shape, its location in the scene, etc. Discursive representations like (1) have

---

[6] This distance-preservation principle is not a necessary feature of icons. Even icons that tend to obey it, such as photographs, do so in a way that is indexed to particular spatial coordinates. For example, two spots right next to each other on a (two-dimensional) photograph might represent parts of the scene that are remote on the z-axis of the three-dimensional scene—think of a photo taken at the top of the Grand Canyon pointing straight down, such that a part of the photo representing the ground right at one's feet might be adjacent to a part representing the bottom of the canyon. It also seems at least conceptually possible that there could be a format in which parts of representations correspond to parts of what they represent, but the way parts are organized fail to preserve structural relations such as distance. One might argue that a section of a pie chart, for example, has spatial parts that correspond to parts of what it represents—if it represents a segment of the population in favor of a certain policy, greater support for the policy is represented by more parts in the relevant segment. Yet spatial relations between parts of a segment may not express anything about the relations between the people in favor of the policy. In addition to the isomorphism of distance preservation, computations over icons in human minds also tend to exhibit a "second-order isomorphism" (Shepard & Chipman 1970), where computational relations between representations mirror relations between objects or stages of objects; see the discussion of mental imagery in Section 3.1.

[7] Fodor (2007) also argues, following Kosslyn (1980), that one can freely segment icons spatially and preserve semantic significance, such that icons have no canonical decompositions of any sort. Nothing below hangs on whether this claim is true.

[8] These properties of icons hold of parts down to some level of fineness of grain; there will inevitably be parts (e.g., individual molecules) that fail to represent anything, let alone some part of the scene or holistically bound clusters of features (Fodor 2008, 173n6). This suggests that icons must break down into primitive representational parts, though what counts as a primitive part will be relative to a particular type of representation (e.g., the primitive parts of some photographs may be pixels, each of which corresponds to some part of the scene and represents values along color and spatial dimensions).

canonical decompositions into a relatively small number of constituents rather than consisting of an array of parts that holistically bind featural information.

The core difference between iconic and discursive representations is that the latter do, and the former do not, break down into recombinable parts each of which can stand uniquely for a particular property, location, or individual.

The term 'iconic' has been used for decades in cognitive science to refer to representations with the properties just mentioned (Neisser 1967; Kosslyn 1980; Kosslyn et al. 2006; Fodor 2007; Carey 2009). Burge, for example, writes that "perceptual representation has a structure relevantly like that of pictorial representation" such that "[j]ust as one cannot draw a line without drawing its length, shape, and orientation, one cannot visually represent an environmental edge as such without representing its length, shape, and orientation, as such" (Burge 2014b, 493).

Dretske (1981; 2000) argues that perceptual representations are analog in that they contain "non-nested" information. The information that s is F "nests" the information that s is G if there's some analytic or nomic relation between s's being F and s's being G. For example, a representation of some determinate shade of red "nests" information about the category *red* as well as simply *color*. For Dretske, analog representations always carry *non*-nested information (Dretske 1981, 136ff). Dretske's condition on analog representations carrying non-nested information captures the fact that icons compose features holistically; if each part of an icon represents multiple properties at once, then a representation of one feature will have to also represent a value along another feature dimension, and thus will necessarily carry non-nested information.[9]

[9] There are many other senses of 'analog' in the literature, too many to detail here. Goodman's (1976) notion of "density," i.e., that between any two atomic representations there lies a third, is arguably far too stringent (though Haugeland [1998, 82–83] pointed out that it is "everybody's aboriginal intuitive idea of analog systems"). Goodman's account nonetheless captures (*inter alia*) the fact that distance relations between parts of icons correspond to distance relations in what's represented and that properties represented in icons vary along continuous (though not necessarily dense) feature dimensions; hence it will typically (but not always) be true that between any two parts of an icon lies another part and that between any two values along a particular dimension lies a third value. David Lewis (1971) argues that analog representations use magnitudes to represent magnitudes, which is true of icons (Carey 2009, 118–135). Icons may be best understood as supersets of coordinates along analog feature dimensions, thereby capturing all these conditions on analogicity (Quilty-Dunn 2017). If parts of icons vary along multiple continuous dimensions at once—e.g., part of a photograph might simultaneously have values along dimensions of hue, saturation, brightness, and the spatial x- and y-axes of the image—then they can be modeled as coordinates. In that case, parts of icons will map continuous magnitudes to continuous magnitudes, values will be holistically bound together in each set of coordinates, and it will often be that between any two

The best way to substantiate the claim that parts of icons correspond to parts of the scene and represent multiple features holistically is simply to demonstrate its explanatory utility. In the following section, I will describe some evidence that supports the thesis that perceptual systems output mental representations that are iconic in this sense. One might object to the notion of iconicity developed here. A successful version of this objection that still sought to appeal to iconicity in explaining perception would require an alternative notion of iconicity with independent empirical substantiation. Block (ms.) writes that nobody has adequately defined iconicity, and as noted above, my aim is not to provide such a definition. But even if nobody has adequately *defined* iconicity, the *phenomena* discussed in the following section that are explained by appeal to iconicity indirectly provide a characterization in terms of parts that represent parts and encode features holistically.

§3. Evidence for Perceptual Icons

I'll now argue that there is considerable non-demonstrative evidence in favor of the hypothesis that at least some of the representations outputted by perceptual systems are iconic (i.e., are "perceptual icons"). It's important to note at the outset that accepting this hypothesis does not necessitate that *all* perceptual representation is iconic and is therefore compatible with perceptual pluralism. Even such a staunch defender of iconic format in perception and perceptual imagery as Kosslyn still affirms that "there are two sorts of representations

values along a continuous dimension there lies a third. There may be mental representations that are analog in some sense but are not iconic, since they use magnitudes to represent magnitudes but fail to bind them to spatiotemporal values, such as analog magnitudes in numerical cognition (Clarke ms.). Though these analog representations encode features one-by-one, they simply do not compose features together at all (e.g., numerosity can be represented independently of location or other features). There is good reason to think that when analog representations *do* compose features like color, shape, and location together, however, they do so in a manner that obeys the principles of holistic binding and parts correspondence that are characteristic of iconicity (e.g., in ensemble perception and mental imagery). Since the perceptual object representations discussed in Section 4 below compose features together non-holistically, they tell against any analog-based account of perceptual object representation; there is no independent reason to suppose analog feature dimensions can bind together without doing so holistically. Intuitively, while you might be able to have separate analog representations for color and shape, an analog representation that simultaneously represents both—e.g., an image—does so by means of the same parts of the representation and is therefore holistic. Finally, I note that Haugeland's (1998) distinction between analog and digital representation eschews format differences entirely, but while his theory is of interest, I cannot devote adequate space to discussing it here. The usefulness of format-based distinctions will be clear in what follows.

underlying images, one 'perceptual'…and one discursive" (1980, 142), the latter constituting "a list of facts in a propositional format" (1980, 144).

The evidence for perceptual icons is broad and diverse, and too large to be chronicled here with any reasonable degree of completeness. One such example is topographic mapping in sensory cortices, where (e.g.) spatial location and features like edge orientation are holistically bound together (Kosslyn 1994).

Another is high-capacity sensory memory stores (Sperling 1960), such as "iconic memory." Sperling briefly showed subjects rows of (e.g., 9) letters. Subjects could name only 3 or 4. But when a row was cued after the letters disappeared, subjects could name nearly all the letters in that row; this suggests that a high-capacity representation is briefly stored early in visual processing. The high capacity of early sensory memory stores can be explained by icons that lack constituents for separate individuals and can thus store additional items without requiring additional vehicles (Fodor 2007). Moreover, holistic feature-binding requires icons to encode many features at once (Dretske 1981).[10] Bronfman et al., for example, found that color properties of rows of letters are represented "spontaneously and without cost" (2014, 1401) in a task that involves encoding the letterforms of only a single row; the fact that color comes for free with shape in this way is a prime example of high-capacity holistic feature binding. There is controversy about whether all the information in such studies is represented consciously (Phillips 2011; Block 2014a; Ward et al. 2016), but there is general consensus that the information is *represented*, which suggests iconic format whether or not it phenomenally overflows cognitive access (cf. Gross & Flombaum 2017; Quilty-Dunn ms.). Indeed, the apparent richness and fine-grainedness of perceptual phenomenology may constitute another form of evidence for nonconceptual perceptual icons (Tye 2006).

I'll now focus in detail on two examples: mental imagery and ensemble perception. The purpose of the following discussion is twofold—first, to argue that there are perceptual icons, and second, to show that perceptual icons obey the principles outlined in the previous section, especially that they lack separate symbols for separate properties and individuals.

3.1—Mental imagery. The imagery debate has been raging since the early 1970s, with Kosslyn, Shepard, and others defending the iconic interpretation, and Pylyshyn, Dennett and others defending the discursive interpretation. Phenomena like mental rotation (Shepard and Metzler

[10] Fragile visual short-term memory is another sensory memory store and also seems to be iconic, as suggested by its high capacity and holistic binding of shapes to locations (Pinto et al. 2013).

1971) and image scanning (Kosslyn et al. 1978) seem to show that differences in processing speed are proportional to spatial differences in what is represented, suggesting that the processing involves manipulating or "scanning" iconic mental images with functional analogues of the spatial properties of the represented scene.

Finke and Pinker (1982), for example, showed participants a picture containing four dots; then the picture disappeared and was replaced by an arrow pointing in a specific direction, and participants were asked whether the arrow was pointing at one of the previously presented dots. Reaction times increased proportionally to the degree of distance between the arrow and the dot. The iconic interpretation of this effect is that the subjects form a mental image of the dots and compare the orientation of the arrow. The proportionality between distance and reaction time is explained by appeal to the functional–spatial structure of the icon itself.

This explanation invokes iconic representations with parts that correspond to parts of the scene. In order for reaction times to increase when reporting on greater distances between the arrow and the dot, subjects must need to access intermediate location values one after another such that a greater number of intermediate location values takes more time. There must therefore be parts of the representation that correspond to locations in the scene such that scanning over a longer distance requires scanning over more parts of the image.

Moreover, parts of the representation not only correspond to parts of the scene, they also encode both the location and any shape present at that location. The functional relationship between the part of the representation that represents the arrow and the parts that represent each dot is explained by the fact that each of those parts also represents a certain location value, thus situating representation of shape within a larger functional–spatial array. One cannot access the shape of the arrow without accessing its location, suggesting holistic binding. Iconic explanations of image-scanning therefore invoke the holistic character of icons.[11]

[11] Pylyshyn (2002; 2003) has argued strenuously for a purely discursive model of mental imagery. I cannot engage with his arguments at length here (see Kosslyn et al. 2006). It's worth noting, however, that so-called "iconophobic" models of mental imagery have perennially been on the defensive, while the most exciting empirical developments have been driven by iconic models of mental imagery despite constant (and useful) methodological critique from Pylyshyn and others. The predictive success and resilience of iconic models is a powerful reply to such critiques (Prinz 2002).

Imagery occurs in topographically mapped cortical areas also used in online vision (Kosslyn et al. 2006, chapter 5). Topographically mapped areas of visual cortex are also the loci of the informational persistence involved in iconic memory (Duysens et al. 1985), suggesting a commonality of format. Imagery can also enhance perceptual discrimination in the same sort of task as online perceptual learning (Grzeczkowski et al. 2015). Moreover, vision and visual imagery compete for resources (Pearson et al. 2015). The thesis that visual imagery and early vision share the same representational format explains these interactions.

3.2—Ensemble perception. People can only explicitly individuate about four individual objects at once (Pylyshyn 2003). Nonetheless, there is evidence that we extract statistical regularities in scenes by computing over many more than four objects. This capacity is known as ensemble perception or perception of summary statistics.

In a pioneering study, Ariely (2001) showed participants as many as 16 circles of varying diameters. Participants were then shown a probe circle and asked whether it was identical to one of the 16 just seen. They were more likely to answer "yes" the closer the diameter of the probe was to the mean diameter of the set (independently of whether the probe actually was a member of the set). This result requires participants to encode the average diameter of the set of circles even though they do not encode each individual circle as such. Furthermore, they were also above chance when asked explicitly whether a probe was greater or smaller in diameter than the average of the set.

Ensemble perception suggests that participants represent an array of items without deploying a discursive representation for each individual item that can be stored and used for report. If participants deployed discursive representations for each individual item, then they should succeed at identifying an object as a member of the set. But Ariely found that "observers were unable to distinguish test spots that were in the set from those that were not" (2001, 159). Similar effects have been found for color (Haberman, Brady & Alvarez 2015), location and motion direction (Hubert-Wallander & Boynton 2015), and even facial emotions (Haberman & Whitney 2012).

Ensemble perception can be explained by supposing that perceptual icons function as inputs to processes that extract statistical averages. This explanation in terms of iconic inputs to ensemble coding can explain why participants lack information about whether an individual item was a member of the presented set as well as why ensemble coding can average over so many items—the items were encoded iconically and were not segmented out and represented

by distinct discursive representations (see also Fodor's [2007] discussion of "item effects"). Ensemble perception also proceeds independently of loading visual working memory with discursive representations (Epstein & Emmanouil 2017). Moreover, the fact that ensembles seem to be storable in iconic memory (Bronfman et al. 2014) provides independent support for their iconicity.

Moreover, people with prosopagnosia (or "face blindness") perceive ensemble properties for faces such as average emotion and identity, with performance comparable to control subjects with normal face-recognitional abilities (Yamanashi Leib et al. 2012). The same subjects who are unable to identify individual faces, therefore, nonetheless perceive the average identity of a crowd of 18 faces. If ensembles are computed on the basis of iconic representations that fail to deploy discursive symbols for individual items, this discrepancy can be explained. Prosopagnosics may be unable to deploy discursive representations for individual faces (thus explaining their failure to recognize individual faces) but retain the capacity to iconically represent an array of faces, which can then be averaged over.[12]

There is good (tentative, abductive) reason to suppose that there are perceptual icons. But as we will see below, not all representations delivered by perceptual systems are iconic.

§4. Perceptual Object Representations

The most striking example of discursive format in perception occurs in object perception. Kahneman, Treisman, and Gibbs (1992) showed the existence of an "object-specific preview benefit" (OSPB) in a kind of experiment known as an object-reviewing paradigm. In the object-reviewing paradigm, participants are presented with two objects (e.g., circles) in which features are previewed (e.g., letters). The features then disappear and the objects move in different directions. Then a feature flashes in one of the objects and participants answer whether it's one of the previewed features or not (or simply name the feature). It doesn't matter for the task which object the feature was originally previewed in. If, however, the same feature was previewed in that particular object originally, participants are quicker at identifying a match. For example, in Figure 2, participants would be quicker to recognize the 'T' as a match if it were presented in the same object in both the preview and target displays. This is true even though the objects are qualitatively identical (thus the effect is not feature based) and move

[12] This explanation posits that some aspects of face perception deliver and operate over discursive representations—for more discussion, see Quilty-Dunn 2017.

locations after previewing (thus the effect is not location based). The effect is genuinely object specific.

Kahneman et al. posited the existence of "object files" to explain the coherence of scene-level visual perception across changes in retinal stimulation, leading to the prediction of an OSPB (Kahneman and Treisman 1984). Object files are representations of particular objects that do not represent objects via any particular feature, but rather bind features by attributing them to the same object.

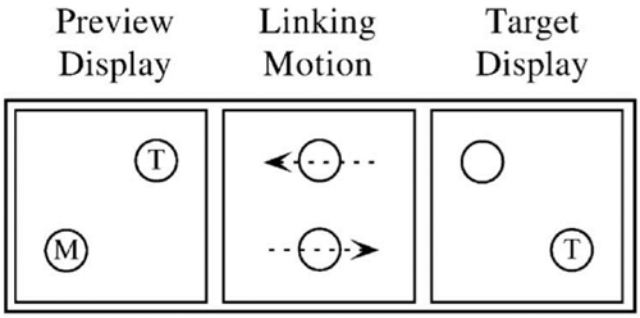


Figure 2—An object-reviewing paradigm (adapted from Mitroff et al. 2005)

Object files are linked to the "visual indexes" or "FINSTs" studied by Pylyshyn, Scholl, and others. While the notion of object files arose from studying the OSPB, the notion of visual indexes came from studying multiple-object tracking (MOT). In the classic MOT paradigm, participants foveate on a fixation cross while roughly eight objects (e.g., squares) populate the rest of the stimulus (Pylyshyn & Storm 1988). Some small number (e.g., four) of the objects will flash, indicating that they are to be tracked, while the rest are to be ignored. Participants are remarkably good at tracking three or four objects, even when the objects are qualitatively identical and intersect each other's paths and even when they are occluded by hidden barriers.

Visual indexes and object files are plausibly manifestations of the same underlying capacity. Object files need spatiotemporal addresses of objects to maintain identity over time and across changes, which would be supplied by a visual index. MOT leads properties of objects to be encoded and stored (Bahrami 2003) and boosts the OSPB (Haladjian & Pylyshyn 2008). These results suggest that visual indexes and object files are bound together in perceptual representations of objects.

In order to avoid confusion, I will refer to perceptual object representations (PORs). PORs are representations of objects that select individuals through the deployment of a visual index and store information about the objects they represent.

The mere existence of PORs places pressure on a wholly iconic model of perception. Object perception requires a segmentation of the percept into constituent PORs, and thus rule out a naïve view of perception on which the visual system delivers a single pure, unsegmented icon. This view is obviously implausible, however, since other forms of segmentation like figure-ground segregation or other forms of perceptual grouping are incompatible with it as well. Instead, anti-pluralists can allow that the visual system segments the visual field, but still require that every output of perceptual segmentation processes is an icon.

Carey (2009) cites evidence that, when shown two groups of cracker that are then hidden in buckets, children crawl toward the bucket with more overall cracker; however, this only occurs when the number of individual crackers is three or less, suggesting that object representations are involved (Feigenson et al. 2002). However, it is unclear why the mere ability to represent and sum over surface area implicates iconic format, since surface area can be represented in any format. This experiment therefore doesn't seem directly to probe representational format.

There is a wealth of evidence that does bear directly on the format of PORs. It strongly suggests that PORs are not icons.

(i) Visual indexes. PORs incorporate an index-like representation of individual objects. The evidence for this hypothesis is that MOT works despite dramatic changes in features. For example, Zhou et al. (2010) found that changing color, luminous flux, and various shape properties did not disrupt MOT. Bahrami (2003) introduced a change detection task, where objects in an MOT paradigm moved around and an object might change its shape or color. Subjects often failed to register changes despite successfully tracking the objects (especially when a "mud splash" appeared on the screen during the change), though detection of changes were better for tracked objects than untracked objects (see also Scholl et al. 1999; Saiki 2003).

This evidence suggests that, while features are bound to objects in MOT, the binding is not holistic. The representation of the object can persist even though features are changing or lost entirely. If the representation of a feature is syntactically separate from the representation of the object, then we would predict that they could detach from one another in the way suggested by these results. This syntactic separation precludes holistic binding and instead requires a distinct constituent that stands for individual objects apart from various features. Thus PORs are not icons.

This argument is not meant to deny that an icon could represent changing features—motion pictures would be a salient counterexample. But it is not obvious that such an icon

would *explicitly* and continuously represent the unchanging identity of an object without the help of some additional representational apparatus, such as a label-like representation persisting across the various changes and securing explicit reference. Nor is the argument meant to deny that an icon could ever lose features (e.g., a color image might somehow become black and white). However, the hypothesis that a POR is a holistically bound iconic representation of an object and its various features would not straightforwardly predict the free loss of even basic low-level features amid successful continuous tracking. Instead, we should expect that features and objects, once bound together, will generally be stored together or lost together. While the change and loss of features is not logically incompatible with an iconic model of PORs, therefore, it is more readily predicted and more easily explained by a discursive one.

The foregoing argument also doesn't require that indexes are label-like in the strong sense of being genuine mental analogues of bare demonstratives such as 'this'. Pylyshyn (2008) holds that view, and it is indeed incompatible with an iconic model of PORs. But what matters is only that the binding between indexes and feature representations is not holistic. The evidence calls for a representation of an object that comes apart from representations of color, luminosity, and various shape properties. This alone is sufficient to reject the view that a POR is an icon. An icon that encodes the color, shape, and location of an object does so holistically, and PORs lack this sort of holistic binding. This is true even if the indexical components of PORs always encode certain properties (e.g., location, trajectory, and even topological shape [Zhou et al. 2010]).

(ii) Abstract features. The fact that the indexical components of PORs are not iconic may be compatible with features' being encoded via an icon. The evidence tells against this view as well. Before looking at the evidence, however, note an *a priori* problem. This hybrid view holds that the indexical components of PORs are not iconic, and that featural components are; thus PORs comprise multiple formats at once. But how? If these representations are in genuinely distinct formats, the proponent of the hybrid view must offer an account of how they compose.

This *a priori* point is not necessary to defeat the hybrid view, however. The object-reviewing paradigm can be used to determine how features are stored in PORs, which in turn can shed light on representational format. A series of experiments have showed abstract properties to be explicitly represented in PORs (Quilty-Dunn 2016). Henderson (1994) found an OSPB for letters previewed in objects even when the letters differed in case from preview

to test displays. Gordon & Irwin (1996) used entire words, e.g., 'bread' appeared in the preview display and 'BREAD' in the test display. Given that there are no shapes in common from 'bread' to 'BREAD', the identity of the word must be stored in a format that abstracts away from shape properties.

If features were represented iconically in PORs, then they should be bound together holistically. In that case, a POR should represent high-level properties like the semantic identities of words in a way that is bound up with representations of low-level properties—one should not find a high-level property represented independently of low-level properties. That's not to say that icons can't represent high-level properties; rather, they should only do so holistically, just as Figure 1 represents Dylan, but in a way that is bound up with low-level properties (fi.e., with his specific appearance in that photograph). Burge, for example, writes regarding face perception that "higher-level facial attributives never float free from low-level geometrical attributions" (2014a, 578). These object-reviewing results nonetheless show non-holistic encoding of high-level properties that float free from low-level properties.

Burge (2014a) appeals to a distinction between specific and generic low-level properties that ground perceptual attribution of high-level properties. For example, there might be a generic banana-shape property that the visual system uses to ground perceptual attribution of the kind *banana*. One might therefore object that the studies just cited only show high-level properties floating free from specific low-level features, not generic low-level features.

A later study by Gordon & Irwin (2000) found the OSPB even when preview stimuli were words and test stimuli were corresponding pictures. For example, the previewed features might be the words 'apple' and 'bread' and the test stimulus would be a match if it were a picture of either an apple or a loaf of bread (see Fig. 3). Gordon & Irwin found that response times were quicker if the picture of the apple occurred in the same object in which 'apple' was previewed. There are no low-level features in common between the word 'apple' and a picture of an apple—not even highly generic low-level features. The POR needs to store a representation of the category *apple* in a way that completely abstracts away from generic as well as specific low-level features. This capacity would be explicable if what's being stored is a distinct symbol that has the content <apple> and is bound to other features of the object only in a separable, non-holistic fashion. That is, the capacity is explicable if the category is encoded in a discursive rather than iconic format.

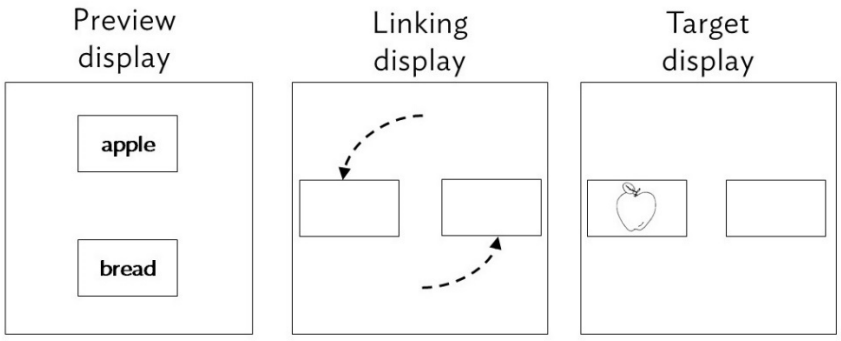


Figure 3—Illustration of Gordon & Irwin (2000)

If discursive representations of categories are bound into PORs non-holistically, then we should expect category representations bound into PORs in one modality to be accessible by another modality. This is precisely what Jordan et al. (2010) found. Jordan et al. used pictures as preview features, after which objects would move to either side of the screen. Subjects then heard a sound come from one side and had to indicate whether the sound matched either of the previewed pictures. For example, suppose objects 1 and 2 are in the center of the display and the previewed pictures are a telephone (1) and a cat (2); 1 and 2 then move to the left and right, respectively. Hearing either a telephone ring or a meow from either side would be a match but hearing a dog bark would not. Jordan et al. found an OSPB: subjects were quicker to identify a sound as a match if it came from the same side as the corresponding object. That is, subjects would be faster if they heard a meow from the right than if it came from the left, since object 2 was previewed with a picture of a cat and moved to the right. The effect is not due to post-perceptual activation of lexical representations like 'cat', since subjects engaged in articulatory suppression (i.e., repeated words in their head during the study), thus loading verbal working memory.

This result vindicates the idea that discursive symbols that stand for categories and abstract away from low-level features are stored in PORs without using a modality-specific format. The authors draw this very conclusion, hypothesizing that PORs "store object-related information in an amodal format that can be flexibly accessed across senses" (2010, 500).

One might try to explain these results through a classical empiricist strategy—namely, by appeal to associative connections between modality-specific icons. Perhaps a visual icon of a cat, for example, is associatively linked to lexical representations like 'cat' and auditory representations like the sound of a meow, such that seeing the shape of a cat causes the activation of an auditory representation. This explanation fails for two reasons.

First, associative learning involves a "gradual strengthening of something" through repeated exposures (Gallistel & King 2009, 220), while a single presentation of a feature in an object generates the OSPB. There are cases of one-shot associative learning (i.e., learning an association in a single exposure that is then modulable through extinction). But the classic examples of one-shot learning of associations—taste aversion (Logue 1979) and avoidance learning (Seligman 1970)—involve high-arousal affective states such as pain and fear. They also involve species-specific prepared learning (e.g., rats will associate a painful shock with audiovisual stimuli but not gustatory stimuli and will associate a noxious feeling with gustatory stimuli but not audiovisual stimuli [Garcia & Koelling 1966]; see also Bolles 1970). There is no independent reason to suppose that this sort of one-shot associative learning underwrites the OSPB, which involves neither high-arousal affect nor biologically prepared learning.

Second, an associative explanation makes a clear prediction: the OSPB should be found for items associated with the previewed item. Gordon and Irwin (1996) specifically tested this by presenting words as preview features and associated or non-associated words as test features and running a lexical decision task where subjects indicated whether the test feature was a word or non-word. While they found a general priming effect (i.e., reaction time was quicker for discriminating test words associated with a previewed word), they found *no* object-specific benefit for associated items. That is, while reading the word 'doctor' in an object did cause general facilitation for the word 'nurse' over unrelated words like 'bread' (as an associative model predicts), the associated word meaning was not bound into the object file. Therefore the effects that show an OSPB for categories cannot simply be associations between iconically represented features.

These results show that the thesis that PORs encode features in an iconic format is empirically untenable.

(iii) Low-level features. As a last retreat, one might admit that indexes and abstract categories are represented discursively in PORs but hold on to the idea that low-level features are represented iconically. This view preserves anti-pluralism in name only. It concedes that amodal discursive symbols are deployed by the visual system alongside icons and therefore seems to be a version of perceptual pluralism. The concessive anti-pluralist could still insist that this view preserves the claim that every output of the visual system is at least *partly* iconic.

Unfortunately, even this view is not consistent with the evidence. If low-level features are represented in an icon in each POR, then low-level features are bound holistically and thus should not come apart from one another. As mentioned above, Bahrami (2003) found that

participants often failed to store information about the shape and color of tracked objects in an MOT paradigm (though not as often as they failed to store features of untracked objects). Bahrami also found differences in when shape and color were lost. For example, in trials without a distracting mud splash, color changes were detected significantly more often than shape changes. Thus there were trials where color was preserved and shape lost in a POR.[13]

Green and Quilty-Dunn (forthcoming) argue that low-level features are not encoded iconically in PORs on the basis of independent feature storage in object representations in visual working memory. They cite evidence that subjects can, e.g., store the color of a triangle in a POR without its orientation, and vice versa (Fougnie & Alvarez 2011). Green and Quilty-Dunn argue that this result shows non-holistic feature binding, and thus non-iconic format, in low-level feature representations bound into PORs.[14]

---

[13] It should be noted that, in general, change blindness results do not demonstratively prove loss of stored information. It is consistent with Bahrami's results that subjects merely failed to compare successfully stored information with the information in the post-change display, as Bahrami himself notes (2003, 962), following Simons (2000). Thanks to an anonymous referee for pressing this point. However, the independent evidence discussed below for separate feature dimensions being represented by separate symbols seems to provide at least some reason to favor an interpretation of Bahrami's results as showing a genuine loss of stored information. It's also worth noting that increasing encoding time and using precision-sensitive tests had no impact on independent feature loss in the Fougnie and Alvarez (2011) study discussed below, which suggests that independent feature loss in PORs is not simply an artifact of some performance constraint but genuinely involves the independent storage and loss of features (see also Park et al. 2017). Finally, the Bahrami study found enhanced change detection for tracked items; thus subjects did seem to successfully access their object representations. This provides some reason to think the results may not be a matter of access failure (though comparison failure is still possible). See note 17 for more discussion of the relation between object correspondence and change blindness.

[14] Block (ms.) appeals to a distinction between properties that are integral and those that are not, i.e., groups of properties that must be instantiated together (like hue and saturation) and those that need not (like height and speed). I'm skeptical of this distinction, since it seems to me that every object with a certain height must also have a certain speed—even if that speed is zero—and likewise for the other "separable" features Block mentions. Indeed, Burns (1987) found holistic binding of size and brightness in iconic memory, two features that should be separable by Block's lights but are physically integral. That worry aside, however, the notion of integrality is tied to feature types rather than to syntactic modes of composition and thus differs from the holisticity of icons. Block claims that some icons can fail to be represent integrally, as when an icon encodes the width of a volume of liquid but no (determinate) height even though width and height must be coinstantiated. But these sorts of cases involve features that fail to be encoded together at all rather than features that are encoded together and then come apart individually. As argued above, what matters for the holistic character of iconicity is not that features are always encoded together but rather that, when they are encoded together, they are represented by means of the same parts of the representation and thus should not be expected to come apart as readily as if they had been encoded by means of separate vehicles. Block's appeal to icons that lack integrality therefore cannot be used to accommodate the evidence discussed here.

Other evidence consistent with this hypothesis shows that storage limits in PORs are individuated by type of feature. For example, Wang et al. (2017) first showed participants colored triangles at various orientations, took them away for a delay period, and finally displayed a second array of colored triangles that did or didn't change one of the colors or orientations. They also varied the number of different feature values (such that an array of six triangles could have six different colors and only two orientations or vice versa). They found that increasing the number of feature values along one dimension significantly damaged performance on the change detection task for that feature, but not for the other.

On an iconic model, storing both the color and orientation of an object is simply a matter of storing one and the same iconic symbol. It's not obvious why the storage capacities for color and orientation would differ if both features are stored by means of the same symbol. On that sort of model, storing color and storing orientation would be accomplished by storing the very same representation, and thus *ceteris paribus* one would expect aspects of their storage (such as required encoding time, capacity, and duration) to be equivalent (as the evidence once seemed to suggest [Luck & Vogel 1997]). But if instead features in different dimensions are stored by means of discrete symbols, then the storage capacity for color (represented by one symbol) and for orientation (represented by another) could easily vary independently. The fact that storage capacities differ for distinct feature dimensions is therefore better explained by separate symbols for separate features that can be stored independently from one another. Green and Quilty-Dunn (forthcoming) argue that a large body of evidence supports a "multiple-slots" model of object-based storage in VWM, where representations along different feature dimensions are stored in dimension-specific slots whose storage capacities vary independently. The idea of storing representations of features in separate working-memory slots is hard to square with a model on which features are holistically bound into a single iconic symbol.

There is no compelling positive evidence for iconic format in PORs and a plethora of compelling evidence against iconic format in PORs. Combined with the arguments in favor of perceptual icons in the previous section, there is good reason to think that the visual system outputs both iconic and discursive representations. In other words, perceptual pluralism is true.

§5. PORs in Perception

One might object that PORs are not genuinely perceptual, and hence don't bear on the format of perception. In that case, the moniker 'POR' is a misnomer; for the rest of this section I will instead use the neutral term 'object file'.

The view that object representation is always post-perceptual was famously defended by Elizabeth Spelke, who argued that perception proper represents only a "continuous layout of surfaces in a state of continuous change" and not coherent "units" (1988, 229). Spelke's argument, however, simply assumes that perceptual representations are exclusively iconic. Her argument would rule out *any* form of segmentation as being genuinely perceptual, and therefore seems to assume an overly impoverished view of perceptual representation. Moreover, much of what drives Spelke's argument is the fact that object files are accessible across modalities and represent relatively high-level properties like solidity. Similarly, Carey argues that object files are in "core cognition" rather than perception (though see Carey 2011) because they "cannot be stated in the vocabulary of perception" (2009, 63) since they do not reduce to "perceptual or sensori-motor primitives" (2009, 67). But the thesis that object files have a discursive format can explain these properties without positing that these representations are in any sense post-perceptual.

Carey also argues that object files have a "rich conceptual role" (2009, 94). She cites evidence that "infants as young as 2 months old represent physical relations between objects such as inside and behind, and their representations are constrained by knowledge of solidity—a property of real objects but not of 2-D visual objects" (2009, 103). She also mentions that infants expect objects to be "subject to the laws of contact causality" and that they "represent objects as the goals of human action" and "represent self-moving agents as the cause of motion of inanimate objects" (2009, 103). Carey concludes that object files figure in inferences and "play a central conceptual role" (2009, 103).

Some of this evidence seems to be explained in terms of the fact that object files encode properties like solidity that influence looking times and other behavioral measures. This capacity does not in itself require a central conceptual role (though it arguably would if it were true that perceptual content must reduce to transduced primitives). Other evidence does seem suggestive of genuine inferential transitions. But in all these cases object files function as *premises* in central-cognitive inferences, not *conclusions*. A representation might wholly belong to the visual system and yet be fed into cognition to function as a premise in inferences, and perceptual representations in the same discursive format as cognition would be poised to do

so. The highly constrained "conceptual role" of object files therefore does not suffice to show that they are post-perceptual.

Block (ms.) pursues a different strategy. According to Block, the outputs of perception are iconic, and include iconic perceptual object representations; object files are, for Block, only present in visual working memory (VWM) and are largely discursive.[15] Block does not explicitly concede that evidence of the sort discussed in the previous section establishes that object files are discursive, but his main line of response is to insist that experiments involving the object-specific preview benefit (OSPB) and other evidence cited above probe post-perceptual object files.

Object files were not originally posited as constituents of VWM. Instead, they were posited to explain proprietarily visual phenomena, and were considered "mid-level" constructs of the visual system, after the most basic feature detection but prior to late vision or post-perceptual processing (Kahneman et al. 1992). A well-known problem in vision science concerns the correspondence of two temporally contiguous and qualitatively distinct retinal inputs (Ullman 1979). Object files were posited partially to solve this correspondence problem by allowing changes in retinal input due to (i) the movement of objects, (ii) changes in the features of objects, or (iii) saccades, i.e., movement of the eyes, to be coherently integrated by appeal to representations of enduring objects (Kahneman et al. 1992, 179). Segmenting the world into coherent, enduring units that can gain, lose, or change features while retaining their identity allows visual processing to make sense of retinal input. Without object representations, vision would be as William James imagined it to be for infants, a "blooming, buzzing confusion."

[15] Block argues that the iconic elements of object files are mainly limited to spatial properties. This view is *prima facie* implausible since, if perceptual object representations are entirely iconic and object files in VWM merely add a discursive overlay (like adding a caption to an image), then we would expect low-level properties generally to be encoded iconically in VWM. The evidence cited above tells against this prediction. It's unclear why spatial properties would be preserved in VWM in their original iconic format without, say, color; without some independent motivation this hypothesis seems *ad hoc*. Moreover, there is evidence that spatial properties like length, gap presence, and orientation come apart from one another in VWM, suggesting that even spatial properties are not encoded iconically (Hardman & Cowan 2015; see also Green & Quilty-Dunn forthcoming). Online object perception also does not seem to be iconic, given that color and shape come apart from one another in MOT (e.g., Bahrami 2003). Furthermore, as mentioned above, icons in perception don't merely represent spatial properties—they bind spatial properties holistically to other properties such as color (Bronfman et al. 2014). If spatial properties were encoded iconically in online object perception, therefore, we should expect them to bind holistically to other properties, but they don't.

For Block's proposal to be correct, OSPB-based experiments cannot actually probe the outputs of mid-level vision (or at least, not in the same format), despite the fact that they were designed to do so. For if they did, then the evidence detailed in the previous section would show that perceptual object representations are discursive. It is crucial for Block, then, that representations that solve the correspondence problem and representations held in VWM don't have the same format. I'll argue shortly that the evidence suggests that they do. First, however, it will be helpful to consider the relation between object perception and VWM.

There is evidence that object files can be held in VWM (Hollingworth & Rasmussen 2010). There is, however, no reason to conclude from this that object files are not constructed by proprietary mid-level-visual processes or that the OSPB only probes object files after a transformation in format upon entering VWM. Top-down manipulation of representations in VWM is an instance of cognitive processing. But it does not follow that *storage* of representations in VWM is an instance of cognitive processing. And it certainly does not follow that representations stored in VWM are *formed* through post-perceptual cognitive processes, or that their formats are transformed upon entering VWM. As Burge writes, the "primary function" of working memory "is to preserve perception already formed" (2007, 501). Without independent reason to multiply types of object representations, it's plausible that the same mid-level-visual representations that solve the correspondence problem can be held in VWM without transforming their formats and can be studied in the form in which perception delivers them by using the OSPB.

For Block, there are iconic perceptual object representations and only some of their iconic aspects are inherited by object files in VWM. It is not obvious, however, how we can know anything about these iconic perceptual representations if not through the OSPB. One might appeal to MOT and its various effects, including the tunnel effect, wherein objects are perceived as moving continuously behind occluders (Flombaum & Scholl 2006). But as mentioned above, MOT facilitates feature storage (Bahrami 2003) and boosts the OSPB (Haladjian & Pylyshyn 2008), strongly suggesting that the vehicles of MOT (i.e., visual indexes) are constituents of object files. And as argued above, MOT shows non-holistic binding (Scholl et al. 1999; Bahrami 2003; Zhou et al. 2010). MOT thus cannot provide a method for studying putatively iconic perceptual object representations. MOT-based evidence provides good reason to think that the object representations formed in online perception and the discursive object files stored in VWM are one and the same.

Block (ms.) argues that iconic object representations are studied via object-based attentional effects such as inhibition of return (i.e., the inhibition of attention back to an unchanged, previously attended object at its original location)—but the inhibition of return also makes use of VWM resources (e.g., Castel et al. 2003; see below for further discussion). Moreover, the OSPB and MOT are typically taken by researchers to be paradigm cases of object-based attention (Scholl 2001). Block also appeals to the influence of spatial information on object representation as evidence for iconicity, but as noted above, the mere representation of spatial information is not evidence for iconic or discursive format.

The view that discursive object representations are exclusively post-perceptual is unmotivated. But there is also more positive evidence that the object files held in VWM display the hallmarks of being genuinely perceptual. These hallmarks are (a) being informationally encapsulated from cognition, and (b) being integrated in perceptual processes.

One form of evidence for the informational encapsulation of object files is the divergence between cognitive and perceptual criteria for object individuation and tracking. For example, we know that an object can expand and contract while maintaining its identity, but tracking is disrupted when that occurs (vanMarle & Scholl 2003). This is a case of informational encapsulation; the object perception system operates on its own proprietary store of information that excludes information stored in central cognition. This encapsulation also goes in both directions, since the information used for visual tracking is not accessible to cognition. It's a surprising result that visual tracking is limited in this way, not a mere verification of something we already (cognitively) knew to be the case.

Another, even more striking case of informational encapsulation was found by Mitroff et al. (2005). Mitroff et al. ran a typical OSPB paradigm, except the motion of the objects intersected in a way that was visually ambiguous between two objects "bouncing" off one another and two objects "streaming" through one another (see Fig. 4). In addition to testing for an OSPB in identifying a match between features presented before and after the ambiguous motion, they also asked subjects to judge whether the objects had bounced or streamed. Remarkably, they found a sharp divergence—the OSPB showed that the object files had bounced even though subjects judged the objects to have streamed through each other. This result is another clear example of informational encapsulation. The processes that output object files are stimulus driven and informationally encapsulated, and thus seem to be genuinely perceptual. Even if the OSPB paradigm probes object files only once they're stored

in VWM, the fact that the paradigm taps into *encapsulated* object representations provides strong evidence that it taps into the outputs of perception prior to cognitive influence.

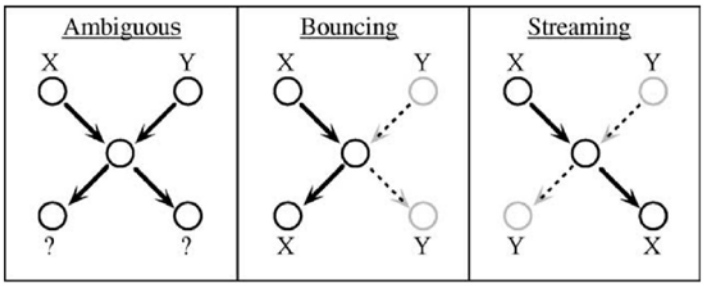


Figure 4—From Mitroff et al. 2005

This experiment also motivates strongly against the claim that object files are transformed in format upon entering VWM. If an object file's format is transformed and its contents are subject to post-perceptual cognitive processes, then it should not show encapsulation from post-perceptual judgment. And if it fails to show encapsulation from post-perceptual judgment, then we should not find the OSPB directly contradicting post-perceptual judgment. The Mitroff study thus provides strong evidence against the thesis that OSPB probes merely the "remnants of perception" (Block ms.) after intervention by post-perceptual cognitive processes; on the contrary, the OSPB probes object representations as they are delivered by encapsulated perceptual processes independently of cognition. OSPB-based evidence thus cannot be dismissed as tapping into representations that have been transformed by post-perceptual cognitive processes.[16]

Another hallmark of a representation's being genuinely perceptual is that it is integrated into perceptual processes, which object files are. One example is the demonstrated relationship between object files and visual indexes (Haladjian & Pylyshyn 2008). Another perceptual process which operates on object files is the guiding of saccades. Upon walking into an unfamiliar room, your eyes might wander around and thus "saccade" to various parts of the

[16] An anonymous referee raises the worry that the Mitroff et al. (2005) study may undermine the claim that object files underlie perceptual phenomenology, which may in turn threaten their status as perceptual. While the result shows that object files can occur unconsciously (Quilty-Dunn forthcoming), it doesn't show that they *always* occur unconsciously. As Mitroff et al. note, it's compatible with their results that object files diverge from conscious experience only rarely, when stimuli are ambiguous in ways that concern the solidity constraint on object persistence (2005, 88–90). The hypothesis that object files are a crucial part of explaining aspects of visual phenomenology (e.g., solving the correspondence problem) doesn't entail that they are sufficient for visual phenomenology. It's compatible with this hypothesis that in ambiguous cases other processes may bias phenomenology in a way that contradicts the encapsulated perceptual processes governing visual object persistence.

room. Perceived information about the scene is used to guide where we saccade and to maintain a stable percept across saccades. This phenomenon is undeniably perceptual. It is hard to imagine what seeing would be like if the visual system did not maintain information across saccades—our percepts would be an incoherent series of snapshots. Visual perception that did not store information across saccades would also fail to be useful for ordinary action. Our eyes are often moving as we perform an action, and in order to visually guide our actions effectively we must be able to maintain coherent visual percepts amid changes in where our eyes are pointing from one moment to the next. The ability to store and integrate visual information across saccades is critical for the coherence of visual phenomenology and the visual guidance of action. It is extremely implausible that all object-based representations and processes that endure across eye movements are post-perceptual cognitive ones.

Episodic information used to guide saccades is referred to as "transsaccadic memory" (Irwin & Gordon 1998). Representations in transsaccadic memory are directly used for low-level visuomotor tasks like tracking a moving object; an object might move while you saccade to it, and the ability to correctively saccade to its new location requires a transsaccadic memory store. While the guidance of saccades is sometimes under voluntary top-down control, the *information stored about the scene* is perceptual and is directly accessed and operated over by low-level sensorimotor processes. The representations stored in transsaccadic memory must be used by the visual system to solve the correspondence problem, i.e., to render visual phenomenology coherent across saccades.

Object files are the constituents of transsaccadic memory (Irwin & Andrews 1996; Irwin & Gordon 1998; Henderson & Siefert 2000; Gordon & Vollmer 2010). Schut et al. (2017a), for example, found that corrective saccades are facilitated by previous fixations despite interim changes in location and features. They conclude that "corrective saccades are executed on the basis of object files" (Schut et al. 2017a, 138).

Like object-reviewing paradigms, tests of transsaccadic memory can shed light on representational format. Irwin (1992) presented participants with an array of letters at one fixation and, after saccading, presented a partial report cue to a subset of letters, similar to the Sperling (1960) experiments discussed above. This experiment provides a clear test of the hypothesis that the correspondence problem is solved by means of iconic format, since if the visual system stores iconic representations across saccades, we should find the same behavioral signature found in tests of iconic memory. Unlike in the Sperling experiments, however, participants only showed storage of three or four letters—the same limit for discursive object

representations. This result falsifies the claim that icons are used in deriving object correspondence across saccades.[17] Since object correspondence needs to be computed by the visual system (and not merely by some post-perceptual process—cf. Block ms.), then there must be non-iconic representations in the visual system.

Pollatsek et al. (1984) had participants fixate on the center of the screen with an object in the periphery, then saccade to the object and name it. Changing the features of the object during the saccade slows down response time, showing that features are encoded in transsaccadic memory. Pollatsek et al. switched out objects with a different exemplar of the same basic-level category, with sometimes substantially different low-level features (e.g., a young cat with textured fur facing forward vs. an older cat without textured fur facing to the left, or dogs of two different breeds—see Fig. 5). They still found significant facilitation in naming the object despite the large change in low-level features.



Figure 5—Stimuli from Pollatsek et al. 1984

Rayner et al. (1980) used a similar paradigm with words instead of pictures and found that words could vary in case across saccades and still facilitate naming. Pollatsek et al. (1990) used similar stimuli to Pollatsek et al. (1984) but varied the location of objects pre- and post-saccade. They still found significant facilitation despite the change in location and low-level features. This suggests that the abstract identity of the picture is encoded in transsaccadic memory without holistic binding to low-level features (see also Henderson & Siefert 2000).

[17] As discussed above in note 13, change blindness is compatible with storage of information, since failure at change detection can arise from failure to compare past and present stimuli (Mitroff et al. 2004). The basic function of transsaccadic memory, however, is to compute object correspondence across saccades; this function requires not merely storage, but also successful comparison of information pre- and post-saccade such that the visual system can determine whether there is object correspondence (i.e., the same object in a different retinotopic location due to the saccade). Since the Irwin (1992) study shows a failure to integrate iconically represented information across saccades, the possibility that the results are explained by comparison failure does not undermine the point that iconic information is not used for transsaccadic object correspondence.

Objection: These transsaccadic memory effects only show storage of what Burge calls "generic" low-level features, like a generic cat shape.

The objection is correct in that (e.g.) Pollatsek et al. (1984) did not control for this alternative. But the independent evidence that object files are stored in transsaccadic memory, together with the independent OSPB-based evidence that object files encode kinds independently of generic low-level features, should bias our interpretation of these results in favor of the claim that representations in transsaccadic memory explicitly encode abstract categories.

Moreover, there is independent evidence within the transsaccadic memory literature to support the hypothesis that transsaccadic memory stores abstract categories. Gordon and Vollmer (2010) had participants saccade to a location between two objects, at which point the two objects were replaced by a single object. The task was to name this object. The object either did or didn't match the category of one of the two previewed objects and may or may not have changed color. Gordon and Vollmer found an object-specific effect: a decrease in reaction time for naming the object above(/below) the fixation point if that object was previewed above(/below) fixation. This effect was significantly reduced when color changed, *but only for objects with diagnostic colors*. Thus the correspondence between transsaccadic representations of a banana depended on the yellow color of the banana, whereas correspondence for transsaccadic representations of an object without a diagnostic color (such as a bucket) did not depend on its color. This result suggests that the representation is not maintained merely on the basis of low-level properties; which low-level properties are used for object correspondence depends on their relation to abstract categories that are explicitly represented in transsaccadic memory (see also Gordon et al. 2008; Gordon 2014). This effect is predicted by a model on which discursive symbols are bound into object files that are held in transsaccadic memory and fails to be predicted by an iconic model.

It is unclear what mechanism could solve the correspondence problem if not representations stored in transsaccadic memory. But these representations are not iconic and seem to be the same object files that are accessible to VWM. The evidence thus suggests that the representations that solve the correspondence problem (and are therefore genuinely perceptual) are identical to the discursive object files that figure in VWM.

The view that transsaccadic memory only stores post-perceptual representations that fail to share a format with perception is *ad hoc*. But it also faces a more serious problem that arises from asserting simultaneously that object files are transformed into a conceptual

discursive format after perception and yet also are used to guide saccades and solve the correspondence problem. If the format of object files has been changed to one that is not native to low-level sensorimotor systems, how can such systems access and compute over them across saccades? It seems far more plausible to say instead that the discursive format of object files is in fact native (though not exclusive) to vision, and that this continuity of format between object representations in vision and in VWM explains how constituents of the latter can be operated over by low-level visuomotor processes. If iconic object representations solved the correspondence problem, then we should find signs of iconic format in transsaccadic memory, but we don't.

As mentioned above, Block argues that object-based attentional effects such as inhibition of return probe iconic, genuinely perceptual object representations while transsaccadic memory effects and the OSPB probe post-perceptual discursive object files. This dichotomy is untenable, however, given that object-based attention uses transsaccadic memory. For example, inhibition of return operates across saccades (Ro et al. 2000; Ludwig et al. 2009). In order for attention to be inhibited for a particular object across saccades, the representation of the object must be stored across saccades and must therefore be a constituent of transsaccadic memory—i.e., a discursive object file. There is evidence that inhibition of return relies on preparation of saccades, suggesting that transsaccadic memory mediates much or even *all* object-based inhibition of return (Rafal et al. 1989); for example, inhibiting subjects' ability to saccade to a visible object minimizes inhibition of return (Michalczyk et al. 2018). It would be extremely hard to explain these results if object-based attention were not driven by object representations in transsaccadic memory. Thus even the object-based attentional effects Block favors such as inhibition of return are explained by discursive object files rather than iconic object representations. This fact provides powerful independent evidence in favor of the hypothesis that discursive object files are the vehicles of object-based attention. There is therefore no reason to posit more than a single (discursive) type of object representation. Any object representation that failed to survive across saccades would fail to be useful for human perception, cognition, and action, and the object representations that do survive across saccades appear to be discursive.

One might make a blanket assumption that all working memory systems, including VWM, fall cleanly within the borders of cognition rather than perception. Block insists that transsaccadic memory is simply VWM, and there is indeed significant experimental evidence in favor of this claim (e.g., Schut et al. 2017b; Kleene & Michel 2018). These two assumptions together entail that transsaccadic memory is cognitive rather than perceptual (Block ms.). But

in fact transsaccadic memory is central to visual processing. It allows the visual system to integrate information across saccades, a function that lies at the heart of the coherence of visual phenomenology and the usefulness of vision for action. It also underwrites basic aspects of visual attention, such as the inhibition of return. The centrality of transsaccadic memory to visual processing should lead us to doubt the overly simplistic assumption that VWM is wholly post-perceptual and cognitive.

Perceptual processes deliver perceptual representations, and these representations can be held in VWM. Once held in VWM, they can be used in specifically perceptual ways by perceptual processes; for example, they can be used by the visual system to maintain a coherent percept across saccades. The use of object representations in VWM for saccades taxes VWM capacity but not auditory working memory capacity (Schut et al. 2017b), suggesting both (a) that VWM is a modality-specific (rather than simply cognitive) store and (b) that the use of object representations to guide saccades deploys proprietarily visual mechanisms to fulfil a proprietarily visual function. Cognitive processing can also occur in VWM. VWM constitutes an interface between perception and cognition where perceptual information can be held and used for either perceptual or cognitive purposes. We should thus resist attempts to cordon off all representations in VWM as post-perceptual, on pain of not being able to make sense of core perceptual functions. The fact that the same representations delivered by the visual system into VWM can be used offline in cognitive processing only bolsters the hypothesis that perception and cognition often use the same discursive representational format.

To sum up: object files are deployed via encapsulated input-driven processes, are used in low-level visuomotor and attentional processes, and are responsible for a basic form of coherence in perceptual phenomenology and object-based visual guidance of action. If there are iconic object representations in perception that never enter working memory (or change formats upon entering working memory), we have no convincing evidence of their existence. They also seem to play no role in solving the correspondence problem. We should conclude that object files are discursive, genuinely perceptual object representations formed by the visual system that can be held in working memory without transforming their representational format.

§6. Conclusion

Perceptual pluralism offers a rich and flexible account of how we perceive the world. The fact that perception outputs discursive representations can explain how cognition is able to use

some perceptual information directly in updating beliefs and planning actions. The fact that perception outputs iconic representations can explain how some aspects of perception are fundamentally different from discursive thought.

Perceptual pluralism also opens up a potentially fruitful research program. Further work could aim to characterize the variety of formats that are outputted by perceptual processes. For example, face perception, gist perception (Mandelbaum forthcoming), structural perception (Green forthcoming), aspects of ensemble perception (Quilty-Dunn 2017) and other perceptual capacities may deliver diverse representational formats.

The pluralist picture advocated here has focused on a binary distinction between iconic and discursive formats. History may show that a more accurate picture will appeal to a range of distinct formats, or the iconic–discursive distinction may turn out to map onto the most explanatorily significant division among representational kinds. Or perhaps there are multiple discursive formats used in the mind, such that one could mount a format-based distinction between perception and cognition by appeal to different discursive representational structures. At present, however, the deployment of discursive representations in perception suggests a commonality of format between aspects of perception and cognition, thus undermining representational approaches to the perception–cognition border.

The encapsulation of PORs is consistent with an architectural approach. Developing such an approach in a full and empirically plausible way remains a task for future research. However things shake out, it is unlikely that perception can be usefully distinguished from cognition by appeal to the difference between iconic and discursive representational formats.[18]

[18] Thanks to audience members at the Rutgers-Barnard-Columbia Mind Workshop, the Philosophy of Mind Works in Progress group at Oxford, and the Thought and Sense conference at the University of Oslo for useful feedback. Thanks also to Zed Adams, Bahador Bahrami, Jake Beck, Jake Berger, Ned Block, David Chalmers, Sam Clarke, Ryan DeChant, Tatiana Aloi Emmanouil, E.J. Green, Steven Gross, Zoe Jenkin, Eric Mandelbaum, Michael Martin, John Morrison, David Papineau, Ian Phillips, Jesse Prinz, Nick Shea, Josh Shepherd, and Joulia Smortchkova for discussion and/or comments on an earlier draft. I'm grateful to an anonymous referee for *Noûs* for extensive comments that greatly improved the paper. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 681422.

References

Ariely, D. (2001). Seeing sets. *Psychological Science* 12(2), 157–162.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition* 10(8), 949–963.

Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly* 59(236), 385–404.

Beck, J. (forthcoming). Marking the perception–cognition boundary: The criterion of stimulus-dependence. *Australasian Journal of Philosophy*.

Block, N. (2014a). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences* 18(9), 445–447.

---. (2014b). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research* 89(3), 560–572.

---. (ms.). *The Border between Seeing and Thinking*. Book manuscript.

Bolles, R.C. (1970). Species-specific defense reactions and avoidance learning. *Psychological Review* 77(1), 32–48.

Braine, M.D.S., & O'Brien, D.P., eds. (1998). *Mental Logic*. Mahwah, NJ: Erlbaum.

Brewer, B. (2006). Perception and content. *The European Journal of Philosophy* 14, 165–81.

Brogaard, B. (2013). Do we perceive natural kind properties? *Philosophical Studies* 162, 35–42.

Bronfman, Z.Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: "Cost free" color phenomenality outside focal attention. *Psychological Science* 25(7), 1394–1403.

Burge, T. (2007). Psychology supports independence of phenomenal consciousness. *Behavioral and Brain Sciences* 30(5/6), 500–501.

---. (2010). *Origins of Objectivity*. Oxford: OUP.

---. (2014a). Reply to Block: Adaptation and the upper border of perception. *Philosophy and Phenomenological Research* 89(3), 573–583.

---. (2014b). Reply to Rescorla and Peacocke: Perceptual content in light of perceptual constancies and biological constraints. *Philosophy and Phenomenological Research* 88(2), 485–501.

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.

Byrne, A. (2005). Perception and conceptual content. In E. Sosa and M. Steup (eds.), *Contemporary Debates in Epistemology* (London: Basil Blackwell), 231–250.

Byrne, A., & Siegel, S. (2017). Rich or thin? In B. Nanay (ed.), *Current Controversies in Philosophy of Perception* (New York: Routledge), 59–80.

Camp, E. (2007). Thinking with maps. *Philosophical Perspectives* 21, 145–182.

---. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research* 78(2), 275–311.

Carey, S. (2009). *The Origin of Concepts*. Oxford: OUP.

---. (2011). Concept innateness, concept continuity, and bootstrapping. *Behavioral and Brain Sciences* 34(3), 152–162.

Castel, A.D., Pratt, J., & Craik, F.I.M. (2003). The role of spatial working memory in inhibition of return: Evidence from divided attention tasks. *Perception & Psychophysics* 65(6), 970–981.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3), 181–204.

Clarke, S. (ms.) Beyond the icon: Core cognition and the bounds of perception.

Cohen, M., & Dennett, D.C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences* 15(8), 358–364.

Dehaene, S. (2011). *The Number Sense*. Second edition. New York: OUP.

Dennett, D.C. (1978). A cure for the common code. In his *Brainstorms: Philosophical Essays on Mind and Psychology* (Bradford), 90–108.

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

---. (2000). *Perception, Knowledge, and Belief: Selected Essays*. Cambridge: Cambridge University Press.

Duysens, J., Orban, G.A., Cremieux, J., & Maes, H. (1985) Visual cortical correlates of visible persistence. *Vision Research* 25 (2), 171–178.

Epstein, M.L., & Emmanouil, T.A. (2017). Ensemble coding remains accurate under object and spatial working memory load. *Attention, Perception, & Psychophysics* 79(7), 2088–2097.

Evans, G. (1982). *The Varieties of Reference*. Oxford: OUP.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology* 44, 33–66.

Finke, R.A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8(2), 142–147.

Firestone, C., & Scholl, B.J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences* 39, 1–72.

Flombaum, J.I., & Scholl, B.J. (2006). A temporal same-object advantage in the tunnel effect: facilitated change-detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance* 32, 840–853.

Fodor, J.A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

---. (1981). *The Modularity of Mind*. Cambridge, MA: MIT Press.

---. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

---. (2007). The revenge of the given. In B. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind* (Oxford: Blackwell), 105–116.

---. (2008). *LOT2: The Language of Thought Revisited*. New York: OUP.

Fodor, J.A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3–71.

Fougnie, D., & Alvarez, G.A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature–store model. *Journal of Vision* 11(12), 1–12.

Gallistel, C.R., & King, A.P. (2009). *Memory and the Computational Brain*. West Sussex: Wiley-Blackwell.

Garcia, J., & Koelling, R.A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science* 4(1), 123–124.

Gawronski, B., & Bodenhausen, G.V. (2006). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology* 44, 59–127.

Goodman, N. (1976). L*anguages of Art: An Approach to a Theory of Symbols*. 2nd edition. Indianapolis: Hackett.

Gordon, R.D. (2014). Saccade latency reveals episodic representation of object color. *Attention, Perception, & Psychophysics* 76(6), 1765–1777.

Gordon, R.D., & Irwin, D.E. (1996). What's in an object file? Evidence from priming studies. *Perception and Psychophysics* 58(8), 1260–1277.

Gordon, R.D., & Irwin, D.E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(1),136–150.

Gordon, R.D., & Vollmer, S.D. (2010). Episodic representation of diagnostic and non-diagnostic object color. *Visual Cognition* 18(5), 728–750.

Gordon R.D., Vollmer S.D., & Frankl M.L. (2008). Object continuity and the transsaccadic representation of form. *Perception and Psychophysics* 70, 667–679.

Green, E.J. (forthcoming). On the perception of structure. *Noûs*.

Green, E.J., & Quilty-Dunn, J. (forthcoming). What is an object file? *British Journal for the Philosophy of Science*.

Gross, S., & Flombaum, J. (2017). Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind & Language* 32(3), 358–391.

Grzeczkowski, L., Tartaglia, E.M. Mast, F.W., & Herzog, M.H. (2015). Linking perceptual learning with identical stimuli to imagery perceptual learning. *Journal of Vision* 15(13), 1–8.

Haberman, J., Brady, T.F., & Alvarez, G.A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General* 144(2), 432–446.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe and L. Robertson (eds.), *From Perception to Consciousness: Searching with Anne Treisman* (Oxford: OUP), 339–349.

Haladjian, H., & Pylyshyn, Z. (2008). Object-specific preview benefit enhanced during explicit multiple object tracking. *Journal of Vision* 8(6), 497.

Hardman, K., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(2), 325–347.

Haugeland, J. (1998). *Having Thought*. Cambridge, MA: Harvard University Press.

Heck, R. (2000). Nonconceptual content and the "space of reasons." *The Philosophical Review* 109(4), 483–523.

Henderson, J.M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General* 123(4), 410–426.

Henderson, J.M., & Siefert, A.B.C. (2000). Types and tokens in transsaccadic object identification: Effects of spatial position and left-right orientation. *Michigan State University Eye Movement Laboratory Technical Report* 3, 1–12.

Hollingworth, A., & Rasmussen, I.P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* 36(3), 543–564.

Hopp, W. (2011). *Perception and Knowledge: A Phenomenological Account*. Cambridge University Press.

Hubert-Wallander, B., & Boynton, G.M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision* 15(5), 1–12.

Irwin, D.E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 18(2), 307–317.

Irwin, D.E., & Andrews, R.V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J.L. McClelland (eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (Cambridge, MA: MIT Press), 125–155.

Irwin, D.E., & Gordon, R.D. (1998). Eye movements, attention, and transsaccadic memory. *Visual Cognition* 5, 127–155.

Johnson-Laird, P. (2006). *How We Reason*. Oxford: OUP.

Jordan, K.E., Clark, K., & Mitroff, S.M. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition* 18(4), 492–503.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D.R. Davis (eds.), *Varieties of Attention* (Orlando: Academic Press), 29–61.

Kahneman, D., Treisman, A., & Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* 24, 175–219.

Kleene, N.J., & Michel, M.M. (2018). The capacity of trans-saccadic memory in visual search. *Psychological Review* 125(3), 391–408.

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

---. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

---. (1995). Mental imagery. In D. Osherson and S. Kosslyn (eds.), *An Invitation to Cognitive Science: Visual Cognition, Volume 2* (Cambridge, MA: MIT Press), 267–296.

Kosslyn, S.M., & Pomerantz, J.R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology* 9, 52–76.

Kosslyn, S.M., Thompson, W.L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford: OUP.

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. (1978). Visual images preserve metric spatial information: Evidence from studies of imagery scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4, 47–60.

Kulvicki, J. (2015). Analog representation and the parts principle. *Review of Philosophy and Psychology* 6(1), 165–180.

Yamanashi Leib, A., Puri, A.M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia* 50, 1698–1707.

Lewis, D. (1971). Analog and digital. *Noûs* 5, 321–327.

Logue, A.W. (1979). Taste aversion and the generality of the laws of learning. *Psychological Bulletin* 86(2), 276–296.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 309, 279–281.

Ludwig, C.J.H., Farrell, S., Ellis, L.A., Gilchrist, I.D. (2009). The mechanism underlying inhibition of saccadic return. *Cognitive Psychology* 59, 180–202.

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology* 6(4), 547–569.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs* 50(3), 629–658.

Mandelbaum, E. (forthcoming). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*.

Michalczyk, Ł., Paszulewicz, J., Bielas, J., & Wolski, P. (2018). Is saccade preparation required for inhibition of return (IOR)? *Neuroscience Letters* 665, 13–17.

Mitroff, S.R., Scholl, B.J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition* 96, 67–92.

Mitroff, S.R., Simons, D.J., & Levin, D.T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics* 66(8), 1268–1281.

Neisser, U. (1967). *Cognitive Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Park, Y.E., Sy, J.L., Hong, S.W., & Tong, F. (2017). Reprioritization of features of multidimensional objects stored in visual working memory. *Psychological Science* 28(12), 1773–1785.

Pearson, J., Naselaris, T., Holmes, E.A., & Kosslyn, S.M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in Cognitive Sciences* 19(10), 590–602.

Phillips, I. (2011). Attention and iconic memory. In C. Mole, D. Smithies, & W. Wu (eds.), *Attention: Philosophical and Psychological Essays* (Oxford: OUP), 202–225.

Pinto, Y., Sligte, I.G., Shapiro, K.L., Lamme, V.A.F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review* 20, 732–739.

Pollatsek, A., Rayner, K., & Collins, W.E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General* 113(3), 426–442.

Pollatsek, A., Rayner, K., & Henderson, J.M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance* 16(1), 199–210.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

---. (2006). Is the mind really modular? In R.J. Stainton (ed.), *Contemporary Debates in Cognitive Science* (Malden, MA: Blackwell), 22–36.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs* 34(4), 517–549.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

---. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22(3), 341–365.

---. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences* 25, 157–238.

---. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

---. (2008). The empirical case for bare demonstratives in vision. In C. Viger and R. J. Stainton (eds.), *Compositionality, Context, and Semantic Values: Essays in Honour of Ernie Lepore* (Springer), 255–274.

Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* 3(3), 179–197.

Quilty-Dunn, J. (2016). Iconicity and the format of perception. *Journal of Consciousness Studies* 23(3–4), 255–263.

---. (2017). *Syntax and Semantics of Perceptual Representation*. Ph.D. dissertation, The Graduate Center, City University of New York.

---. (forthcoming). Unconscious perception and phenomenal coherence. *Analysis*.

---. (ms.) Is iconic memory iconic?

Quilty-Dunn, J., & Mandelbaum, E. (2018a). Against dispositionalism: Belief in cognitive science. *Philosophical Studies* 175(9), 2353–2372.

Quilty-Dunn, J., & Mandelbaum, E. (2018b). Inferential transitions. *Australasian Journal of Philosophy* 96(3), 532–547.

Rafal, R.D., Calabresi, P.A., Brennan, C.W., & Sciolto, T.K. (1989). Saccade preparation inhibits reorienting to recently attended locations. *Journal of Experimental Psychology: Human Perception and Performance* 15(4), 673–685.

Rayner, K., McConkie, G.W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology* 12, 206–226.

Rescorla, M. (2009). Cognitive maps and the language of thought. *British Journal for the Philosophy of Science* 60(2), 377–407.

---. (2014). The causal relevance of content to computation. *Philosophy and Phenomenological Research* 88, 173–208.

Ro, T., Pratt, J., & Rafal, R.D. (2000). Inhibition of return in saccadic eye movements. *Experimental Brain Research* 130, 264–268.

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision* 3, 6–21.

Scholl, B.J. (2001). Objects and attention: The state of the art. *Cognition* 80, 1–46.

Scholl, B.J., Pylyshyn, Z.W., & Franconeri, S. (1999). When are spatiotemporal and featural properties encoded as a result of attentional allocation? *Investigative Ophthalmology and Visual Science* 40(4), S797.

Schut, M.J., Fabius, J.H., Van der Stoep, N., & Van der Stigchel, S. (2017a). Object files across eye movements: Previous fixations affect the latencies of corrective saccades. *Attention, Perception, & Psychophysics* 79, 138–153.

Schut, M.J., Van der Stoep, N., Postma, A. & Van der Stigchel, S. (2017b). The cost of making an eye movement: A direct link between visual working memory and saccade execution. *Journal of Vision* 17(6), 1–20.

Seligman, M.E.P. (1970). On the generality of laws of learning. *Psychological Review* 77(5), 406–418.

Sellars, W.S. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science, vol. I* (Minneapolis, MN: University of Minnesota Press), 253–329

Shea, N. (2014). Distinguishing top-down from bottom-up effects. In D. Stokes, M. Matthen, & D. Biggs (eds.), *Perception and Its Modalities* (Oxford: OUP), 73–91.

Shepard, R.N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology* 1(1), 1–17.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703.

Siegel, S. (2017). *The Rationality of Perception*. New York: OUP.

Simons, D.J. (2000). Current approaches to change blindness. *Visual Cognition* 7(1/2/3), 1–15.

Spelke, E. (1988). Where perceiving ends and thinking begins: The apprehension of objects in infancy. In A. Yonas (ed.), *Perceptual Development in Infancy: Minnesota Symposium on Child Psychology, Vol. 20* (Hillsdale, NJ: Lawrence Erlbaum), 197–233.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied* 74(11), 1–29.

Tye, M. (2006). Nonconceptual content, richness, and fineness of grain. In T. Gendler and J. Hawthorne (eds.), *Perceptual Experience* (Oxford: OUP), 504–530.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.

vanMarle, K., & Scholl, B.J. (2003). Attentive tracking of objects versus substances. *Psychological Science* 14, 498–504.

Wang, B., Cao, X., Theeuwes, J., Olivers, C.N.L., & Wang, Z. (2017). Separate capacities for storing different features in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(2), 226–236.

Ward, E.J., Bear, A., & Scholl, B.J. (2016). Can you perceive ensembles without perceiving individuals? The role of statistical perception in determining whether awareness overflows access. *Cognition* 152, 78–86.

Zhou, K., Luo, H., Zhou, T., Zhuo, Y., & Chen, L. (2010). Topological change disturbs object continuity in attentive tracking. *PNAS* 107(50), 21920–21924.