IS ICONIC MEMORY ICONIC?

Jake Quilty-Dunn
Faculty of Philosophy & Brasenose College, University of Oxford
Department of Philosophy & PNP Program, WUSTL

Forthcoming in *Philosophy and Phenomenological Research*
Penultimate draft

Abstract: Short-term memory in vision is typically thought to divide into at least two memory stores: a short, fragile, high-capacity store known as iconic memory, and a longer, durable, capacity-limited store known as visual working memory (VWM). This paper argues that iconic memory stores icons, i.e., image-like perceptual representations. The iconicity of iconic memory has significant consequences for understanding consciousness, nonconceptual content, and the perception–cognition border. Steven Gross and Jonathan Flombaum have recently challenged the division between iconic memory and VWM by arguing against the idea of capacity limits in favor of a flexible resource-based model of short-term memory. I argue that, while VWM capacity is probably governed by flexible resources rather than a sharp limit, the two memory stores should still be distinguished by their representational formats. Iconic memory stores icons, while VWM stores discursive (i.e., language-like) representations. I conclude by arguing that this format-based distinction between memory stores entails that prominent views about consciousness and the perception–cognition border will likely have to be revised.

> *"There are two kinds of visual memory: one when you skillfully recreate an image in the laboratory of your mind, with your eyes open (and then I see Annabel in such general terms as: 'honey-colored skin,' 'thin arms,' 'brown bobbed hair,' 'long lashes,' 'big bright mouth'); and the other when you instantly evoke, with shut eyes, on the dark innerside of your eyelids, the objective, absolutely optical replica of a beloved face, a little ghost in natural colors[.]"*
>
> *—Vladimir Nabokov*

§1. Introduction

As Nabokov points out, there is more than one kind of memory. There is a difference between short-term memory, where you store a new phone number long enough to dial it a moment later, and long-term memory, where you store your own phone number such that you can produce it whenever asked (Miller 1956). Within long-term memory, there is a difference between semantic memory, where you store your knowledge that The Rolling Stones are a rock band, and episodic memory, where you may store your memory of hearing *Exile on Main Street* for the first time (Tulving 1983).

There is also more than one kind of short-term memory. Sticking to vision, theorists generally accept the existence of visual working memory (VWM), which allows a relatively small number of items to be stored for a relatively short amount of time (Miller 1956; Baddeley 1986; Luck & Vogel 1997; 2013; Cowan 2001; Block 2011a; Prinz 2012; Suchow et al. 2014; Cohen et al. 2016). The majority (though by no means universal) opinion is that VWM doesn't exhaust visual short-term memory.[1] There appears to be an earlier, sensory aspect to visual short-term memory that has a higher capacity and shorter duration than VWM—this store is typically known as "iconic memory" (Sperling 1960; Neisser 1967; Block 2011a; Phillips 2011). Some argue for a third, "fragile" visual short-term memory store that is intermediate between iconic memory and VWM in processing order as well as capacity and duration (Sligte et al. 2008; Block 2011a; Pinto et al. 2013).

Any adequate characterization of a memory store must (minimally) specify its capacity, duration, architecture, and format. Capacity is the amount of information a store can hold, and duration is the length of time it can hold it. The architecture of a memory store is its basic, invariant structure (Pylyshyn 1984). For example, some have argued that VWM contains three or four "slots," such that storing an object and its various properties takes up one slot, and no

---

[1] The term "visual short-term memory" or "VSTM" for short is often used by vision scientists simply to refer to VWM. However, I use it here as a generic term meant to cover all short-term memory stores available to visual processing.

more than three or four objects can be stored at once (Luck & Vogel 1997; Cowan 2001; Zhang & Luck 2008). This hypothesis (which will be critically discussed later on) is a claim about the functional architecture of VWM. Another such claim is that iconic memory is "maskable"—any newly presented stimulus will wipe out an iconic memory, a generalization that is not true of VWM and true of fragile visual short-term memory only in certain circumstances.

Finally, a characterization of a memory store must specify the format of the representations that are stored in it. This paper concerns the format of representations stored in iconic memory. I'll argue that iconic memory stores iconic perceptual representations, or "perceptual icons."

The question of whether iconic memory stores perceptual icons is centrally important for several debates in the philosophy of mind. One such debate centers around consciousness. Ned Block (2011a) argues that phenomenal consciousness "overflows" the reportable information held in working memory. If conscious perceptual representations outstrip what we can report at a particular time, then theories that tie consciousness to reportability are untenable—these include higher-order thought theories (Rosenthal 2005), global workspace theories (Baars 1988; Dehaene & Changeux 2005), and perhaps certain attention-based theories (cf. Prinz 2012). Block's case for the overflow thesis depends in part on the claim that representations in iconic memory are informationally rich (and phenomenally conscious) perceptual icons, which diverge in format from sparser, reportable discursive representations held in VWM (e.g., Block 2011a, 572–573). Understanding the format of representations in iconic memory is therefore a prerequisite for evaluating the leading contemporary theories of consciousness.

Another relevant debate concerns the border between perception and cognition. Some theorists argue that the distinction between perception and cognition is grounded in a difference in representational format. Theorists such as Dretske (1981), Carey (2009), Kulvicki (2015), and Block (unpublished) have argued that perceptual representations are iconic or analog while thoughts are typically discursive or digital. Iconic memory may be a key piece of evidence in favor of this proposal—Dretske (1981, 149), for example, explicitly appeals to iconic memory as a source of evidence. The question of what kind of format is stored in iconic memory is thus directly relevant to larger debates about mental architecture.

Finally, iconic memory may be relevant to debates about whether perceptual representations are invariably conceptualized (Evans 1982; McDowell 1994; Heck 2000; Byrne 2005; Mandelbaum forthcoming). Fodor (2007), for example, uses iconic memory as an argument for nonconceptual visual representation. If concepts are amodal discursive representations, then icons *ipso facto* fail to be conceptual. In that case, showing that iconic memory houses icons would demonstrate the existence of nonconceptual perceptual states.

In what follows, I'll defend the hypothesis that the elements of iconic memory are perceptual icons. After clarifying the basic distinction between iconic and discursive representations in Section 2, I'll present an argument for icons in iconic memory in Section 3. Then, in Section 4, I'll consider ways in which probabilistic resource-based models of short-term memory threaten to undermine the iconicity of iconic memory. I'll defend the distinction between iconic memory and working memory from recent attacks from Steven Gross and Jonathan Flombaum's (2017) resource-based model by appeal to a distinction in representational format.

I'll conclude by considering the upshots of a format-based distinction between iconic memory and working memory for theories of consciousness and the perception–cognition border. The format-based distinction argued for below, if combined with a cognitivist view of consciousness and a format-based view of the perception–cognition border, has the striking consequence that visual representations are never conscious. Since this consequence seems false, I suggest we ought to reject one of these prominent views.

§2. Iconic vs. Discursive

Icons as I will understand them obey two principles:

PARTS: Parts of icons correspond to parts of what they represent.

HOLISM: Parts of icons represent multiple properties simultaneously.

Consider the following picture of a blue cube.

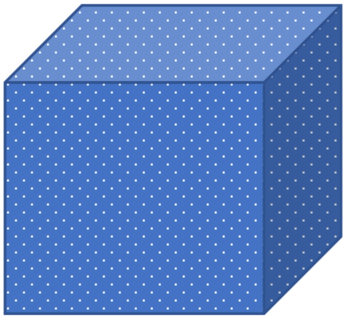

Figure 1—Blue dotted cube

Figure 1 is structured differently than the following sentence.

(1) This is a blue dotted cube.

Figure 1 obeys PARTS; parts of the image correspond to parts of the cube. For example, a part of Figure 1 corresponds to the back top right corner. Parts of (1), on the other hand, don't correspond to parts of the cube. Figure 1 also obeys HOLISM; any part of the image that you point at represents a color and a location simultaneously, and perhaps an aspect of the texture (e.g., a dot or two) as well. A single part of (1), however, represents the individual object (viz., 'This'), while another part represents a color ('blue'), another represents a texture ('dotted'), and another represents a shape ('cube'). (1) acquires its accuracy conditions from the mode of combination of these separate constituents, while Figure 1 acquires its accuracy conditions from the way features are holistically bound in each part, together with the spatial arrangement of those parts (Kosslyn et al. 2006; Fodor 2007).

These two principles are not arbitrarily related. Given an isomorphic relation (i.e., one-to-one correspondence) between parts of a representation and parts of the scene, any part of the representation that encodes a feature such as color will also correspond to a part of the scene. In that case, that part of the representation will not merely represent the color (as a single word or concept might) but will also minimally represent a location in the scene. Any other features instantiated at that part of the scene, insofar as they are represented in the icon at all, will also be represented by the same part that represents the color. It is possible that there may be representations that satisfy PARTS but fail to have a one-to-one correspondence between parts of the representation and parts of the scene (i.e., fail to be isomorphic to what they represent). These non-isomorphic representations may satisfy PARTS without satisfying HOLISM. But paradigm cases like Figure 1 seem to have such a one-to-one correspondence and, not coincidentally, are accurately described by both PARTS and HOLISM. I will restrict use of the term 'icon' to these sorts of isomorphic representations.

Throughout cognitive science iconic mental representations that satisfy PARTS and HOLISM are posited to explain a variety of phenomena, such as core cognition (Carey 2009, 452; 459) and indeed iconic memory (Dretske 1981, 149; Fodor 2007, 112ff). Perhaps the most influential example, however, is mental imagery, where phenomena like mental rotation and the "scanning" of mental images are explained by appeal to iconic mental images (Kosslyn 1980; 1994; Kosslyn et al. 2006; Pearson et al. 2015). While the use of icons to explain imagery phenomena is controversial (Pylyshyn 2003), so-called "iconophilic" explanations typically appeal to both PARTS and HOLISM.

In Shepard and Metzler's (1971) famous mental rotation experiments, for example, participants were presented with a pair of objects and indicated whether they were two differently shaped objects or the same object at two different orientations (see Figure 2 for an example). The authors found a nearly linear correlation between degree of rotation and reaction time—that is, the amount of time it took participants to correctly identify a match between two instances of the same object increased as a function of the degree of difference in orientation. Shepard and Metzler proposed that participants engaged in "mental rotation," a manipulation of a mental representation of the object with stages corresponding to the stages of the physical rotation of the object. This description of the phenomenon posits a "second-order isomorphism" (Shepard 1978, 131) whereby relations among elements of the computational process correspond to relations among elements of the physical process.

Figure 2—Example of mental rotation stimuli (from Wexler et al. 1998, with permission from Elsevier)

One can offer an iconophilic explanation of this second-order isomorphism. If parts of a representation correspond to parts of the object and features are bound holistically in the representation, then the shape of the object is not represented separately from other features such as its orientation. The shape of the whole object is encoded by means of parts of the representation that correspond to parts of the object; parts of the representation encode parts of the shape at particular locations. The arrangement of these parts (corresponding to the arrangement of spatial locations) that encodes the entire shape thus cannot fail to encode its orientation. The resulting holistic binding of shape and orientation entails that a participant cannot access the object's shape independently of its orientation, which hampers the ability to recognize a match in shape across different orientations. Instead the participant must perform some operation to transform the represented orientation to match that of the other object in the pair, at which point the participant can identify a match (or not) between orientation-bound shapes. This can be accomplished by performing an operation functionally analogous to physically rotating the object, wherein the image runs through intermediate orientation values until reaching an upright orientation that matches that of the other object.[2] Since

---

[2] One might wonder why the system would need to run through intermediate orientation values instead of simply changing the orientation immediately to a specific value. Since the orientation of the object is holistically bound up with other properties like shape, the capacity to iconically represent some object at an arbitrary orientation presupposes the capacity to represent its shape at that orientation—but the very point of changing the orientation is to acquire the capacity to represent its shape at the desired orientation. Without this capacity the system could instead implement a stepwise algorithm for representing an object at adjacent degrees of orientation (i.e., a mental-rotation algorithm), thereby moving in a stepwise fashion from the initial orientation to the desired orientation. The capacity to implement this algorithm wouldn't require the ability to represent objects at arbitrary orientations but only adjacent ones, which would be much more computationally feasible—especially assuming

running through a greater number of orientation values requires more computational steps and therefore more time, one would predict that reaction time would be a function of degree of rotation. Appealing to icons that satisfy PARTS and HOLISM thus supplies a principled explanation of mental rotation.

The debate about whether iconic format should be invoked to explain imagery phenomena has been long and hard fought (see, e.g., Pylyshyn 2002). Despite its controversial nature, the mental imagery debate provides a useful case study of iconicity as it figures in scientific explanation. I'll assume in what follows that an icon-based explanation of iconic memory should follow the same principles, and is therefore committed to both PARTS and HOLISM.

§3. Iconic memory

The original experiment that motivated the existence of iconic memory was due to George Sperling (1960). Sperling presented participants with rows of letters (e.g., three rows of three letters each) and asked them to report as many letters as possible. Participants stored an average of about four letters, suggesting that the report was constrained by VWM. In another condition, after the presentation of the letters Sperling cued an individual row for participants to report. In this "partial-report" condition subjects could report nearly all the letters in the cued row (this advantage is sometimes called "partial-report superiority"). Since the cue occurred after the presentation of the letters, so the classic explanation goes, subjects must have stored all or nearly all the letters such that a particular subset of them could be attended in response to the cue and thereafter represented in VWM.

In his influential textbook *Cognitive Psychology*, Neisser (1967) dubbed the earlier, informationally rich representation an "icon" and referred to the high-capacity visual memory store in which it is housed as "iconic memory." Iconic memory has a shorter duration than VWM—Sperling found that partial-report superiority went away after about a quarter-second, whereas information can be stored in VWM for several seconds or even longer.[3] The basic effect

---

a high degree of featural similarity between representing an object at one orientation and an adjacent one. See Quilty-Dunn & Mandelbaum forthcoming for discussion.

[3] The difference in duration between iconic memory and VWM is a complicated issue. More recently, Landman et al. (2003) found that retro-cuing a particular item in a display and asking subjects whether that item changed its orientation remained effective (i.e., showed a higher capacity than "post-cuing", or introducing a cue after the probe display appears and thereby tapping into VWM) for 1.5 seconds. Sligte et al. (2008) found that even retro-cues that appeared *four seconds* after the display remained more effective than post-cues. Sligte et al. posit an intermediate "fragile visual short-term memory" store. I will generally talk about iconic memory in a way that is

of partial-report superiority also suggests that iconic memory has a higher capacity than VWM, since it suggests storage of the entire array of 9–12 letters rather than just the three or four that would plausibly make it into VWM (Cowan 2001—cf. Gross & Flombaum 2017).

The evidence *prima facie* suggests that representations stored in iconic memory are iconic and satisfy PARTS as well as HOLISM (Dretske 1981; Fodor 2007). The limit of VWM to about three or four items is an instance of an "item effect" (Fodor 2007, 111): processing limits are sensitive not merely to the amount of information represented but specifically to the number of items. Item effects are to be expected from discursive representations. If distinct individuals are each represented by discrete vehicles, then increasing the number of represented individuals increases the number of vehicles. Adding more items to a list, or more sentences to a paragraph, thereby demands more resources (such as ink and paper in a literal list or paragraph, or computational symbols and memory space in a computational architecture). This item-driven increase in vehicles should *ceteris paribus* exhibit some effect on processing, such as reaching some item-based limit on memory capacity. A capacity limit tied to the number of represented items therefore suggests that the relevant items are represented *via* a discursive format.

Icons, however, don't require that more individuals be represented by means of more vehicles. Instead of corresponding to individuals as such, parts of icons correspond to parts of the scene. If some scene is represented by an icon, therefore, then for any arbitrary part of that scene, there will be a part of the icon that represents it whether or not there is an object (or part of an object) present at that location.

Imagine an iconic representation of a dark wall, upon which you could place a number of light stickers. The icon will encode every visible part of the surface of the wall whether or not there is a sticker there. If you place thirty stickers on the wall or three, the icon will represent them just as easily. Thus PARTS suggests icons should not be expected to display item effects as discursive representations would. Increasing the number of items in an image might increase the amount of variation within the image, which might increase the demand on representational resources. If the wall is uniformly dark, for example, then if there are thirty light stickers present there will be many more luminance edges than if there are only three stickers, and that might suck up more representational resources. However, this effect would not be due to the number of items *per se*, but rather to discontinuities in luminance; if such

---

meant to include fragile visual short-term memory, in part because they seem to have the same format and in part because they may not really constitute distinct memory stores; see below for discussion.

discontinuities were controlled for, then the number of items would place no additional demands on the icon.

The explanation on offer makes use of both the PARTS principle and the HOLISM principle. Since parts of icons correspond to parts of the scene (rather than to discrete individuals as such), icons can encode a large number of items without increasing the demand on representational resources. And since parts of icons encode information holistically, any part of an icon will represent not only some location in the scene but also properties that are instantiated at that location (such as shape, size, and color). It follows that an iconic representation of a scene could encode a large array of letters or other items that appear in that scene without segmenting out individual items or sorting them into high-level categories. The hypothesis that icons are stored in iconic memory therefore provides a good explanation of Sperling-like results.

These same points apply to fragile visual short-term memory as well. Fragile visual short-term memory also has a far higher capacity than VWM (Sligte et al. 2008; 2010). There is reason to think iconic memory and fragile visual short-term memory involve the same underlying memory capacity but iconic memory simply involves the addition of persistent retinal afterimages (Sligte et al. 2008; see also Block 2011a, 573). Though my main interest is in iconic memory, it is very plausible that fragile visual-short term memory makes use of the same representational format as iconic memory.

The hypothesis that icons are stored in iconic memory makes substantive predictions. Since icons obey HOLISM, then iconic memory should involve encoding a multiplicity of features, even those that are not necessarily relevant for a particular task. This prediction has been verified by a recent experiment from Bronfman et al. (2014). Bronfman et al. showed subjects four rows of six letters each; before presenting the letters, they cued a particular row, and then after presenting the letters, they cued a particular letter in the cued row; the task was to report the particular letter. This aspect of the experiment implements an augmented version of Sperling's partial-report paradigm. In addition, the letters varied in color throughout the display.

On some trials, after reporting the cued letter, subjects were asked to make a judgment about whether the colors of the letters, either in the cued row or in the three uncued rows, seemed high or low in "color diversity"—i.e., whether their colors all clustered near the same part of the color wheel (low diversity) or were spread out (high diversity). Despite the fact that the initial cue and task demands drew attention to the shapes of letters in a particular row, subjects nonetheless encoded the range of colors in the other rows. This result suggests that iconic memory consists of icons that not only encode a large number of items, but encode their various features such as color and shape "spontaneously and free of cost" (Bronfman et al. 2014, 1398). Though there is significant controversy about whether this holistically encoded information for each item is *consciously experienced* (Block 2014; cf. Phillips 2016; 2018; Ward et al. 2016), my argument here is simply that it is *encoded* and *stored* and this fact is best explained by appeal to iconic format.

It is possible that partial-report superiority could be explained by massive parallel processing that delivers discursive representations into a very large memory store. This explanation is inelegant in that there is no independent reason to think that discursive representations can be stored in such high capacity, particularly given the fact that there is a comparatively small item limit on discursive representations in visual and verbal working memory systems (e.g., Cowan 2001).[4] It would also be hard pressed to explain why properties of uncued items (such as color in the Bronfman et al. 2014 experiments) seem to be encoded automatically.

One could add the hypothesis that parallel processes automatically output discursive representations of properties of uncued items, but this addition seems more like an *ad hoc* epicycle rather than an independently motivated hypothesis. Moreover, icons whose structure mirrors the structure of the scene can implicitly encode many more features; e.g., an array of pixels that are ordered spatially can also thereby encode edges and therefore contours of shapes without adding additional symbols. A discursive model of iconic memory would have to posit that these properties are all represented by means of discrete symbols, thus positing a psychologically implausible explosion of discursive symbols.[5]

There may also be independent neurobiological evidence in favor of the thesis that iconic memory stores perceptual icons. The sensory "informational persistence" (Coltheart 1980) that underlies iconic memory utilizes early, retinotopically mapped visual cortical areas of the brain (Duysens et al. 1985; Irwin & Thomas 2008). These areas of the brain overlap with the loci of visual images (Pearson & Kosslyn 2015). Since mental images seem to be iconic, there is therefore some independent reason to think that iconic memory stores perceptual icons.

This inference is not simply that since two representations are activated in overlapping brain areas, therefore they have the same format. In retinotopically mapped visual cortex, the parts of the cortex (such as columns of cells in V1) that instantiate representations of features

---

[4] Note that this point does not require endorsing a "slot-based" over a "resource-based" model of working memory as long as the resources available to VWM limit its capacity relative to iconic memory (see Suchow et al. 2014).

[5] Thanks to Sam Clarke for suggesting this last point.

(such as edge orientations) correspond to particular parts of the retina; thus the neural instantiations of representations of features coincide with neural instantiations of representations of particular locations in the scene that reflect light to particular parts of the retina. This sort of neural architecture seems to be exploitable for instantiating iconic representations that satisfy PARTS and HOLISM—e.g., that represent edges at particular orientations and particular locations in visual field.[6]

§4. The challenge from flexible resources

Gross and Flombaum (2017) have recently provided a compelling defense of an alternative model of visual short-term memory. They argue against the idea of any principled capacity limit on VWM, raising the possibility that iconic memory and VWM may not really be distinct.

Their argument begins from the rejection of a classic picture of VWM, according to which its architecture consists of 3–4 discrete slots that can be "filled" with representations of one object each (Luck & Vogel 1997; Zhang et al. 2008; Adam et al. 2017; Xu et al. 2018). Gross and Flombaum argue that instead that VWM is governed by a continuous resource that can be differentially allocated to represented objects and features, thereby improving storage of those objects and features.[7] Moreover, they argue that objects and features are represented by means of a probability density. Allocation of resources may improve storage by increasing the probability that a particular object is present or that it has some feature.

Gross and Flombaum use this picture to argue that iconic memory may not in fact have a higher capacity than VWM. Instead, iconic memory might involve a relatively flat probability density (i.e., low probabilities assigned to the presence of any particular feature). The role of the cue after cessation of the stimulus may be to allocate resources to a particular

---

[6] *Contra* Clark (2009) and Kosslyn (1994), the exploitability of retinotopy for instantiating iconic format does not suggest that iconic representations should be reductively characterized in neural terms. As Clark (2009) points out, the neural properties of V1 don't perfectly satisfy PARTS. But while Clark takes this to undermine that V1 instantiates iconic representations, we should instead hold that iconicity is a functional, psychological-level notion rather than a neural one.

[7] The metaphysics of memory resources is murky. For present purposes, what matters is that, in interpreting a number of highly noisy signals, one can reduce the noise on some signal and store the result only at the cost of failing to reduce noise for other signals—one cannot reduce noise on all incoming signals at once and store all of them independently. The fact that noise reduction somewhere precludes noise reduction somewhere else captures the essence of the idea of a flexible memory resource; it may arise due to constraints on the normalization of firing rates in neural populations (Bays 2015) or some other neural-level limitation.

subset of the information represented, which would shift the probability density to be high for cued features and much lower for uncued ones.[8]

The basic assumption of this model is well-supported: VWM is not simply an array of non-competing object slots, but rather constitutes a flexible resource that can be allocated across different objects (Bays & Husain 2008; Ma et al. 2014; Suchow et al. 2014; Bays 2015; Schneegans & Bays 2016; Park et al. 2017). Crucially, the flexible resource that underwrites VWM storage modulates the *precision* of stored items. For example, the difference between storing one item and storing four is not simply a matter of filling object-specific slots that encode features like color and orientation. Instead, storing more items causes a decrease in the precision of represented features; subjects' responses on a continuous color wheel, for instance, will still be roughly accurate but will be noisier (Bays et al. 2009). Taking up VWM resources by adding another item causes a decrease in the resources allocated to the other items, resulting in a decrease in precision for stored features across objects. The "flexibility" of VWM resources consists in the fact that a common resource base is allocated across different objects and feature domains.

Gross and Flombaum construe the modulation of resources in terms of probability densities. Allocating more memory resources to an item increases the represented probability (and reduces the estimated standard deviation) that the item has some feature. Gross and Flombaum's probabilistic model, and the resources-not-slots picture it exploits, raise two salient problems for an iconic model of iconic memory. First, they argue that iconic memory may not in fact have a higher capacity than VWM, which challenges a fundamental motivation for distinguishing the two stores and positing iconic format in iconic memory. Second, they suggest that iconic memory may not even store actual representations at all. Instead, they suggest, partial report superiority may simply arise by enhancing the earliest, pre-representational stages of perceptual processing rather than cuing already-stored full-blown perceptual representations.

Gross and Flombaum cite evidence (e.g., Bays & Husain 2008) showing that increasing the number of items to be remembered in VWM causes a decline in precision. This suggests that the limitation on storage in VWM is not a sharp, item-based capacity limit but is rather due to the allocation of a flexible resource shared across items. If VWM does not in fact have a sharp, item-based capacity limit, then the argument that iconic memory exceeds VWM in

---

[8] Gross and Flombaum also argue that Sperling's (1960) results might be explained in terms of hierarchical processing from explicitly represented shape features to full-blown letter representations. I assume that iconic memory represents analog magnitudes and grant for the sake of argument that letter identities are not represented as such prior to VWM.

capacity (which undergirds the claim that its elements are iconic) is weakened. Instead, it may be that a single visual short-term memory store assigns probabilities to a large number of stimuli, and the effect of the cue in iconic and fragile visual short-term memory experiments simply affects the allocation of resources within that single store. In that case, the cue simply shifts around probability densities; while the memory store initially represents all *n* items at some low probability, it continues to represent all *n* items but raises the probability of some small number and drops the rest.

It's extremely important to this story that, though the probabilities are nil for the uncued items, those items are still *represented*. Otherwise, it's not true that the same number of items is represented before and after the cue. Suppose instead that the standard view of visual short-term memory is correct, and the number of represented items drops after the cue. Gross and Flombaum could still reasonably argue that even though the number of items drops after the cue, the capacity does not drop. On the flexible resources view they endorse, short-term memory capacity is defined not in terms of items but rather in terms of the amount of available resources. Resources can indeed be depleted by representing more items, but also by representing the same number of items more richly and precisely. It thus makes sense to say that representing ten items weakly and imprecisely and representing three items precisely and with high probability are two ways of manifesting the exact same memory capacity. In a meaningful sense of information, the two representations could carry the same total amount of information and thus draw on the same memory capacity.

Though this is a fair use of the term 'capacity', there is another sense in which the two representations also clearly differ in how much they represent. And that is not simply a difference in amount of total information, but rather concerns the number of items that are explicitly represented. A memory store that has to distribute probabilities over hypotheses about the color, shape, etc., of ten items has to be able to encode more hypotheses than a memory store that only concerns three items, even if a difference in precision entails that both memory stores have the same capacity. The PARTS principle-based explanation of iconic memory thus remains relevant even if the introduction of a cue doesn't signal a move from a higher-capacity store to a lower-capacity one. Since every part of an icon corresponds to some part of the scene, an icon can encode hypotheses about the perceptible features of ten objects just as easily as three—not so for discursive representations, even if they encode those hypotheses with lower probabilities.

All of which is to say that, if the shift in probabilities induced by the cue in iconic memory experiments reduces the number of items represented, then there is still good reason to hold that the representations maintained prior to the cue are iconic (and thus encode a large number of items and features, perhaps at low probabilities) and the representations after the cue are discursive (and thus encode a small number of items and features, though at higher probabilities and thus without a shift in the total amount of information stored).

Is it really true that, after the cue, the same number of items are represented but most of them simply have very low probabilities? Gross and Flombaum point out that the resource-based model raises this possibility: “If we speak of capacity limits in memory (as opposed to in performance), they will concern the *precision* of representations, not the *number* of items that can be represented. Concerning the latter, no limitation may be assumed at all!” (2017, 375) It is correct that a wealth of evidence points toward resource-based models of VWM and that these models don’t build in any limits on the number of items. But the evidence Gross and Flombaum cite doesn’t provide positive reason to think that VWM represents the same number of items as iconic memory with different probability assignments.

For example, they cite Bays and Husain’s (2008) results that increasing the number of items decreases precision across all stored items. This result does indeed seem incompatible with a simple slot-based model on which working memory capacity is simply a function of the number of represented items and doesn’t involve a shared flexible resource. But it doesn’t follow that VWM can represent the same number of items as iconic memory; it only follows that the number of items is not the sole limit on VWM capacity. Bays and Husain argue for a model of VWM on which it stores representations of features but not objects to which those features belong. However, memory for features is better when they are bound into objects. For example, Fougnie et al. (2010) found that memory for color and orientation improved when the features were integrated into three objects (colored, oriented triangles) rather than diffused across six objects (oriented triangles and colored circles; see also Olson & Jiang 2002). This superiority holds even when the six objects are displayed in overlapping pairs at three locations, and thus seems to be a genuinely object-based rather than location-based effect. The presence of object-based representations in VWM opens up the possibility that VWM capacity can be limited by the number of objects in a way that iconic memory is not, even if it is also constrained by a flexibly allocated resource (Brady et al. 2011; Markov et al. 2019).

A more recent study places pressure on this claim, however. Schneegans and Bays (2016) showed subjects 1, 2, 4, or 8 colored dots at various locations, followed by a 100ms mask and 900ms retention interval (thus ruling out iconic or fragile visual short-term memory). A single probe dot then appeared whose color matched one of the previously shown dots, and subjects moved it to the location at which it had originally appeared. Increasing the set size led a monotonic decrease in accuracy, increase in reaction time, and increase in the likelihood that subjects would erroneously move the probe to the location of one of the other dots. The fact that increasing objects leads to a monotonic decrease in precision—even up to *eight* objects—fits uncomfortably with the idea that the number of objects imposes any sharp limit on VWM capacity.[9]

It is still possible that iconic memory can store even more items. If Sperling's (1960) and Phillips' (1974) results are taken at face value, then iconic memory may store dozens of items. VWM may not be able to store that many items even with extremely low precision. Even if there is no sharp item limit on VWM, there may be a qualitative difference in the number of items explicitly stored with some non-zero probability estimates.

It is not clear that results showing storage of dozens of items in iconic memory should be taken at face value, however. This brings us to Gross and Flombaum's second challenge, viz., that the effects of early retro-cuing may simply modulate the pre-representational registration of sensory information. According to this challenge, the earliest stages of post-retinal processing do not yet involve genuine mental representation, but instead constitute a non-representational registration of proximal stimulation that is at most a mere elaboration of retinal transduction (Burge 2010, 315ff). In that case, the cue in Sperling-type experiments affects "what gets represented (consciously or unconsciously) in the first place" rather than "the selective transfer of representations one already has" (Gross & Flombaum 2017, 365–366). While there may be a great deal of information registered at this early stage, it does not involve genuine representation, and therefore does not actually require positing a memory store that houses representations (iconic or otherwise). According to this challenge, results purporting to show iconic storage of dozens of items (e.g., Phillips 1974) don't actually show anything about storage capacity. Instead, they show that the transition from rich sensory registration to

---

[9] Adam et al. (2017) showed subjects six items and had them report color or orientation for all six objects, one by one. They found that subjects remembered three items and gave answers for the remaining three that were uniformly distributed among possible answers, strongly suggesting guessing rather than merely imprecise storage. This result suggests a hard cap of three items. It's not immediately obvious how to render this evidence consistent with the evidence from Schneegans and Bays that subjects store as many as eight items with variable precision. One salient difference is that Adam et al. asked subjects to report the color of each square on a color wheel using a mouse, while Schneegans and Bays simply had subjects use their finger to move colored test items (e.g., a red item appearing in the middle of the test display) to corresponding locations (e.g., where the red stimulus had appeared in the original memory display). The relative ease of Schneegans and Bays' task may be responsible for the difference. This explanation is admittedly hand-wavey, however, and perhaps hard to square with Adam et al.'s finding that the result is not due to subjects' simply reporting the three best remembered items first and subsequently losing less precise memories of the remaining items (see their Experiment 2, in which the order of report is randomly determined by the computer). Perhaps the best model of working memory resources will entail (i) that resources are doled out in object-specific as well as feature-specific ways, and (ii) that doling out an ordinary amount of resources to three items exhausts remaining resources, creating a virtual limit that can only be overcome through drastically lowering precision on each item and using an extremely easy behavioral measure like Schneegans and Bays' (2016). Exploring this sort of hybrid model would require a separate paper.

genuine perception—and, only thereafter, storage in short-term memory—can be affected by a cue.

How can we determine whether a piece of information is actually encoded in a genuine mental representation as opposed to the registration of the stimulus? Representations are the elements of computational operations; the transition from registration to genuine representation involves the transformation of stimulus information into a form that can be computed over by mental processes. Therefore, showing that a piece of information is computed over in perception would provide some reason to think that it is genuinely represented rather than merely registered.

Another method would be to show that the information does not merely concern the low-level physical energies that are transduced (Pylyshyn 1984, Ch. 6), but instead pertains to higher-level perceptible kinds such as faces, objects, or causal relations. Since higher-level properties aren't available in the proximal stimulus, it is plausible that such properties must be represented rather than merely registered.

We can therefore look for an answer to Gross and Flombaum's challenge by seeing (i) whether information in early visual processes concerns large number of items in respect of higher-level properties and (ii) whether that information functions directly as an input to later visual processes. If evidence for this sort of early representation exists, then we thereby have independent evidence that early retro-cuing can modulate pre-existing representations rather than merely modulating the initial formation of perceptual representations, thus answering Gross and Flombaum's second challenge. And if the number of items encoded by such representations is significantly higher than eight—the highest number of items Schneegans & Bays (2016) found could be stored in VWM at very low degrees of precision—then we have *prima facie* evidence that these early icons have a higher representational capacity than VWM, thus answering Gross and Flombaum's first challenge.

The experimental literature on ensemble perception provides relevant evidence. Ensemble perception involves extraction of statistical properties of collections of items, such as the average size of a group of circles (Ariely 2001) or the variation of different sizes (Solomon et al. 2011). In one common paradigm, subjects are shown a set of items (e.g., circles of varying sizes) and then shown a single probe item and asked whether or how it differs from the average of the set (e.g., whether it is bigger or smaller than the average). Subjects succeed at such tasks even when they are at chance in indicating whether the probe item was the same size as any individual item in the set.

Notably, ensemble perception works on very large set sizes, such as 16 (Ariely 2001). Some have argued that ensembles are computed by sampling three or four presented items (e.g., Myczek & Simons 2008). However, if ensembles were computed by sampling three or four items rather than computing over all items in parallel, then increasing the heterogeneity of sets should reduce accuracy. For example, if a set of 16 circles contains eight circles of one size and eight circles of another size, then sampling four circles and averaging them would likely produce a result close to the average of the set. If instead all 16 circles are distinct sizes, then it becomes more likely that an arbitrary sample will be skewed and thus that the average will be less accurate. Instead, Utochkin and Tiurina (2014) found that heterogeneity does not affect accuracy, suggesting that all items are computed over. Ensemble perception also works even when subjects are under working memory load (Epstein & Emmanouil 2017). These results suggest that large numbers of items are explicitly represented in early stages of visual perception prior to storage in VWM.

Moreover, while Gross and Flombaum's resource-based model allows that there may be no limit to the number of objects in VWM, it still predicts that performance on tasks that make use of stored objects should decline as the number of objects increases. Adding more items requires more VWM resources, lowering precision across items and thereby damaging performance. Robitaille & Harris (2011), however, found that increasing the number of items actually *increases* accuracy and speeds up reaction time in estimating average size and orientation. The fact that ensemble perception improves with more items suggests that, unlike later representation in VWM, there is virtually no cost to adding more items for the early representations that function as inputs to ensemble computations. Indeed, Dakin's (2001) work on averaging of orientation involves displays with *as many as 1,024 items*.

The ensemble perception evidence cited so far pertains to size and orientation. Size in particular may not be the sort of property that is transduced as such and thus may require genuine representations. Indeed, Whitney and Yamanashi Leib (2018) take ensemble perception of size to be notable given its status as a "mid-level" property. Nonetheless, one might still insist that ensemble computations of these properties could operate over mere sensory registrations. However, there is a wealth of evidence showing that ensemble perception operates not only on low-level features but also on higher-level properties. For example, the same sorts of paradigms discussed above have been used to show ensemble perception of the average happiness or sadness in a set of facial expressions (Haberman & Whitney 2007; 2009). This capacity is even exhibited by subjects with "face blindness" who have impaired processing of faces (Yamanashi Leib et al. 2012). In order to compute the average of a set of faces, it is not enough to operate on mere sensory registration. That sensory registration must first be transformed into representations of individual faces that can then be averaged over.[10]

Ensemble perception therefore provides good evidence for genuine representations early in vision that explicitly encode a very large number of items. It also provides evidence of holistic property binding. Ensembles are computed automatically (Oriet & Brand 2013; Epstein & Emmanouil 2017) and across disparate content domains such as orientation (Dakin 2001), brightness (Bauer 2009), hue (Demeyere et al. 2008), position (Hubert-Wallander & Boynton 2015), depth (Wardle et al. 2012), motion direction (Hubert-Wallander & Boynton 2015) and many others (Whitney & Yamanashi Leib 2018). In order for averages to be processed automatically across all these domains, the initial stages of perceptual processing must include explicit representations that encode this range of properties. A simple, unified explanation would posit that early vision involves the construction of icons that holistically encode low-level properties for a large number of items. Ensemble perception processes can operate on these early icons to derive summary statistics of the scene without preserving details about individual items.

Since the ensemble perception literature provides independent evidence to think that these early icons exist, there is at least some independent motivation to think that iconic memory stores iconic representations. The thesis that iconic memory stores icons predicts that we should find such independent evidence for rich icons in early perception. Unfortunately, there is not much direct evidence on the relation between iconic memory and ensemble perception. However, Bronfman et al. (2014) provide evidence that ensemble perception of color diversity can be computed on the basis of information in iconic memory. This result suggests that the iconic inputs to ensemble perception are in fact stored in iconic memory. Rensink (2014) showed that representations stored in iconic memory can serve as inputs to visual search—e.g., identifying a single vertical bar among nine tilted bars. Since the capacity to identify the odd stimulus out of a group plausibly uses summary statistics about the group, Rensink's results suggest that elements of iconic memory are usable for ensemble coding.

Another indirect piece of evidence concerns the holistic encoding of properties in iconic memory. If the representations that function as inputs to ensemble coding are icons, and thus satisfy HOLISM, and if those representations are stored in iconic memory, then we should predict that representations in iconic memory satisfy HOLISM. Burns (1987) presented

---

[10] One might object that faces do not need to be represented *as such* in order for facial properties (such as angle of eyebrows, shape of mouth, etc.) to be averaged over. Even if faces are not represented as such in early vision, however, there must be representations that encode information about facial properties of some sort. It is extremely implausible that all the information required to calculate the average level of happiness in a group of faces is explicitly available in the earliest stages of visual processing that merely register retinal stimulation.

subjects with a 3x3 grid of objects that varied along two dimensions, such as grayscale squares of varying size and brightness. Along these two dimensions, stimuli differed at three "levels," e.g., three progressively larger sizes and three progressively brighter shades. The task was to indicate the locations of objects with a particular stimulus value. A particular session of trials would concern just one dimension (e.g., brightness), and a cue indicated that trial's "level"—i.e., upon hearing a low tone subjects would indicate the locations on the grid of the "Level 1" stimuli (the dimmest squares) and likewise for a middle tone (Level 2, brighter) and a high tone (Level 3, brightest).

Burns ran three different conditions: one where the levels were correlated along the two dimensions, such that the dimmest squares were also the smallest and the brightest were the largest; one where the dimensions were orthogonal, such that the relation was random; and one control condition where the irrelevant dimension didn't vary at all. She found that performance was better than control in the correlated condition and worse in the orthogonal condition. This suggests that even though size and brightness are completely distinct features, and even though only a single feature was relevant for a given session, the variation in the irrelevant dimension either helped or hindered performance depending on whether it tracked variation in the cued dimension. This experiment (and the later replication and modification by Burns & Hopkins 1987) provides powerful evidence for the hypothesis that in iconic memory low-level features like size and brightness are "not immediately encoded in terms of independent psychological dimensions but are initially perceived in terms of wholistic objects" (Burns 1987, 396), and therefore vindicates the hypothesis that iconic memory stores icons.

Pinto et al. (2013) also found that fragile visual short-term memory (which arguably overlaps with iconic memory, as mentioned in Section 3) is obliterated only by masks that share the same type of feature in the same type of location. If the original display consisted of four circles on the left side of the visual field, then masks consisting of rectangles on that same side, or circles on the other side, were ineffective compared to masks consisting of circles on that side. The stored representation that is obliterated by the mask appears to specify shapes and locations in a holistic format.

We can resist Gross and Flombaum's suggestion that retro-cueing is simply a matter of shifting which sensory registrations are transformed into perceptual representations. Instead, we have reason to believe that retro-cueing affects which aspects of icons held in iconic memory are used to form later representations that enter into VWM. Early icons are available to be inputs to ensemble perception, so it makes sense to posit that they are available to be briefly stored in iconic memory. The fact that features are bound holistically in iconic memory provides independent support for this hypothesis. And the evidence from the ensemble perception literature that shows that these icons encode very large numbers of items also pushes back against Gross and Flombaum's suggestion that iconic memory and VWM may not differ in capacity.

It is still logically possible that VWM can store just as many items as iconic memory, albeit with sufficiently low precision that performance fails to be accurate. For Gross and Flombaum, this raises the possibility that we should simply do away with the idea of multiple short-term memory stores in vision:

> [W]e have challenged the claim of successive stores of declining capacity. One can reject this claim by rejecting only the claim of declining capacity, leaving in place the claim of successive stores. But why do so, rather than just posit transitions from noisy signals to probabilistic representations, without any transition from a first such store to a second? The latter view is simpler and requires fewer resources.
>
> (Gross & Flombaum 2017, 382)

I've argued that there is indirect evidence in favor of a genuine capacity decline from iconic memory to VWM. Putting capacity aside, however, focusing on representational format provides an independent way to distinguish iconic memory and VWM. Some of the evidence discussed above for early icons concerns high capacity (emphasizing PARTS), but some concerns the holistic representation of features (emphasizing HOLISM). Even if early icons fail to have a higher capacity than representations in VWM, they may still differ in format. One crucial question, then, is whether features are also represented holistically in VWM. If not, then we can draw a distinction between the early storage of holistic icons and the later storage of non-holistic discursive representations even if the former fail to encode more information than the latter.

The evidence strongly suggests that representations in VWM do *not* encode features holistically (see Green and Quilty-Dunn forthcoming for a review). For example, Fougnie and Alvarez (2011) showed subjects colored triangles at various orientations. Subjects then indicated the color followed by the orientation (or vice versa) of a cued triangle after a delay period. In cases where subjects were very far away from the correct value on one dimension (i.e., when they lost information about that feature), they were nonetheless typically able to produce accurate responses on the other dimension. That is, storage of color in VWM doesn't necessarily correlate with storage of orientation (nor vice versa). Very similar results were found by Bays et al. (2011). These effects imply a lack of holistic binding in representations in VWM, and therefore suggest that representations in VWM have a discursive format and employ distinct symbols to represent distinct feature dimensions. Since these dimensions are represented by means of distinct symbols (i.e., not in an icon that satisfies HOLISM), the fact that one is lost should not be expected to tell you whether the other is lost as well.

One might object that an icon could encode features separately, as in a line drawing of a triangle at some orientation that carries no chromatic information. In that case, the effect may be due to an encoding failure rather than separate feature dimensions being represented by means of separate symbols. This possibility would preserve the iconicity of VWM since it could still be true that, when multiple features like color and orientation are successfully encoded in a representation of an object, they are holistically bound into an icon. However, Fougnie and Alvarez varied encoding times and found no difference in the separability of color and orientation. The effect is not due to encoding failure. Rather, even when both features are encoded into a representation in VWM, they are not holistically bound and can easily be lost separately.

A discursive model of representations in VWM makes a further prediction, namely that separate feature dimensions might have separate memory stores. This prediction seems to be true. Wang et al. (2017) found that increasing the variation in color of an array of triangles diminished storage of the colors of the triangles but not storage of their orientations; and likewise, increasing the variation in the orientations of the triangles diminished storage for orientation but not for color (see also Wheeler & Treisman 2002; Olson & Jiang 2002; Markov et al. 2019).

Green and Quilty-Dunn (forthcoming) appeal to this sort of evidence to argue that VWM consists of discursive object files that represent separate feature dimensions by means of separate representations. On their "multiple-slots" model, representations of features in the same dimension compete for storage in a dimension-specific "slot". While resources can be allocated flexibly across objects and feature slots, resource competition is higher within feature-specific slots than across them. This model and the evidence that supports it require a discursive model of the elements of VWM.

The foregoing discussion points toward a format-based distinction between iconic memory (and its close relative, fragile visual short-term memory) on the one hand and VWM on the other by appeal to the iconic format of representations stored in the former and the discursive format of representations stored in the latter. One might argue that the relation between iconic memory and VWM involves a smooth change in probability density, wherein precision is increased for some subset of items and decreased for the rest. The smoothness of this transition may push against the idea of a sharp difference in format between two separate memory stores and instead suggest a picture closer to Gross and Flombaum's.

However, Pratte (2018) showed that information in iconic memory is either completely present or completely absent. Instead of smoothly decaying or transitioning into VWM, representations in iconic memory "die a sudden death" (Pratte 2018, 77; see also

Zhang & Luck 2009). A format-based distinction between iconic memory and VWM predicts just this sort of sharp discontinuity between information stored in the two memory stores. A representation in VWM is not a modulated version of a representation from iconic memory. It is a distinct token vehicle with a distinct representational format.

This format-based distinction between iconic memory and VWM fits well the mainstream capacity-based distinction between these memory stores, but it does not logically require it. The holistic character of feature binding in iconic memory and the non-holistic character of feature binding in VWM suggest a distinction that cuts across the capacity-based distinction targeted by Gross and Flombaum. And ensemble perception provides evidence for high-capacity early representation that escapes Gross and Flombaum's methodological concerns, suggesting that the PARTS principle is true of the elements of iconic memory but not for VWM. There is very good reason to posit a distinction between iconic memory and VWM and to hold that iconic memory stores high-capacity icons. In other words, iconic memory is iconic.

§5. Conclusion: Consciousness and Short-term Memory

I noted at the outset that the question whether iconic memory stores perceptual icons has significant upshots for debates about consciousness and the perception–cognition border. As we've seen, the evidence strongly pushes in favor of an affirmative answer to this question. This provides a solid foundation for arguments such as Block's (2011a) and Lamme's (2003) for the thesis that phenomenal consciousness overflows cognitive access, since iconic memory does indeed store unconceptualized icons that store more information than discursive representations in VWM.

It is still possible that consciousness only enters the picture in VWM. This position interacts in interesting ways with the perception–cognition border, however, given the hypothesis that iconic memory and VWM differ in their format. On a popular approach to the perception–cognition border pursued by Block (unpublished) and others (Dretske 1981; Carey 2009; Kulvicki 2015), the distinction between perception and cognition tracks the distinction between iconic and discursive formats. If this approach is correct, and if only representations in VWM are conscious, it follows that *visual representations are never conscious*.[11]

---

[11] An anonymous referee for this journal suggests an alternate description: visual representations are only conscious when conceptualized. This re-description sounds less unacceptable than the formulation in the text. But it seems to me inapt. One could think of a representation's being conceptualized as its being accompanied by a concept, in which case saying a conceptualized representation is conscious might mean that the representation, together with its accompanying concept, is conscious. But the circumstance described here is different. Supposing a format-based distinction between iconic memory and VWM, it follows that iconic

Since VWM only stores discursive representations and perceptual icons are held in iconic memory, then if consciousness is limited to the elements of VWM, consciousness cannot extend to proprietarily visual representations.

This conclusion seems distinctly odd, not least because a paradigm case of conscious experience would seem to be our distinctly visual experience of (e.g.) redness. One may be inclined simply to bite the bullet and deny that this experience is properly visual. Instead, however, there are a few ways to avoid this implausible-seeming conclusion. One could deny, as Block does, the claim that consciousness is limited to the elements of VWM and argue for phenomenal overflow. Or one could deny the claim that perception and cognition are distinguished by their representational format and allow that the discursive representations in VWM are just as visual as the icons in iconic memory (Quilty-Dunn forthcoming). Another option is to endorse a higher-order thought theory of consciousness (Rosenthal 2005), on which working memory resources may be required for consciousness, but the representations that become conscious may extend outside the reach of VWM simply by being represented by higher-order thoughts. This possibility highlights a flexibility of higher-order thought theories of consciousness not possessed by other theories. However, higher-order thought theories have independent problems that render them unacceptable to many theorists (Block 2011b; Prinz 2012).

If we want to retain the idea—common among experimental researchers especially—that consciousness involves the use of perceptual representations in working memory, then we should reject the thesis that representational format is the key to the perception–cognition border. If we don't reject that thesis, then we should wholeheartedly embrace the controversial view that phenomenal consciousness overflows cognitive access. But if what I've argued above about representational format and short-term memory is true, then we can't both accept a format-based approach to the perception–cognition border and deny that the content of phenomenal consciousness overflows visual working memory.[12]

---

representations aren't present in VWM at all. In that case (given the assumption that the contents of visual consciousness are limited to the contents of VWM), the conscious representation in VWM does not include the iconic representation as a constituent. Thus it's not that the visual representation is conscious and conceptualized; it is at most an input to a process of conceptualization, the output of which is conscious. Thus on the constellation of views under discussion visual icons themselves are never conscious.

[12] Thanks to E.J. Green and the Oxford Philosophy of Mind Work-in-Progress Seminar for discussion, to Sam Clarke, Steven Gross, Zoe Jenkin, Nick Shea, and Joulia Smortchkova for comments on an earlier draft, and to an anonymous referee for this journal for helpful comments. 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

Adam, K.C.S., Vogel, E.K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology* 97, 79–97.

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

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge: CUP.

Baddeley, A. (1986). *Working Memory*. Oxford: OUP.

Bauer B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *Psychological Research* 59, 171–186.

Bays, P.M. (2015). Spikes not slots: Noise in neural populations limits working memory. *Trends in Cognitive Sciences* 19(8), 431–438.

Bays, P.M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854.

Bays, P.M., Catalao, R.F.G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision* 9(10), 1–11.

Block, N. (2011a). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences* 12, 567–575.

Block, N. (2011b). The higher order approach to consciousness is defunct. *Analysis* 71(3), 419–431.

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

Block, N. (Unpublished). *The Border between Seeing and Thinking*. Book manuscript.

Brady, T.M., Konkle, T., & Alvarez, G.A. (2011). A review of visual working memory capacity: Beyond individual items and toward structured representations. *Journal of Vision* 11(5), 1–34.

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

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.

Burns, B., & Hopkins, A. (1987). Integral and separable stimulus structure in iconic memory: A replication and extension of Burns (1987). *Perceptual and Motor Skills* 64, 263–270.

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

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

Clark, Austen. (2009). Location, location, location. In D. Dedrick and L. Trick (eds.), *Computation, Cognition, and Pylyshyn* (Cambridge, MA: MIT Press), 281–302.

Cohen, M.A., Dennett, D.C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences* 20(5), 324–335.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics* 27(3),183–228.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 87–185.

Dakin, S.C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A* 18(5), 1016–1026.

Dehaene, S. & Changeux, J-P. (2005). Ongoing spontaneous activity controls access to consciousness A neuronal model for inattentional blindness. *PLoS Biology* 3(5), 910–927.

Demeyere N., Rzeskiewicz A., Humphreys K.A., & Humphreys, G.W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia* 46(11), 2861–2864.

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

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

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.

Fougnie, D., Asplund, C.L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision* 10(12), 1–11.

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

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.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology* 17(17), R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception & Performance* 35(3), 718–734.

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

Hubert-Wallander, B., & Boynton, G.M. (2015). Not all summary statistics are made equal: Evidence from extracting summary statistics across time. *Journal of Vision* 15(4), 1–12.

Irwin, D.E., & Thomas, L.E. (2008). Visual sensory memory. In S.J. Luck & A. Hollingworth (eds.), *Visual Memory* (Oxford: OUP), 9–41.

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

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

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

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

Lamme, V.A.F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences* 7(1), 12–18.

Landman, R., Spekreijse, H., & Lamme, V.A.F. (2003). Large storage capacity of integrated objects before change blindness. *Vision Research* 43, 149–164.

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

Luck, S.J., & Vogel, E.K. (2013). Visual working memory capacity: From psychophysics to neurobiology to individual differences. *Trends in Cognitive Sciences* 17(8), 391–400.

Ma, W.J., Husain, M., & Bays, P.M. (2014). Changing concepts of working memory. *Nature Neuroscience* 17(3), 347–356.

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

Markov, Y.A., Tiurina, N.A., & Utochkin, I.S. (2019). Different features are stored independently in visual working memory but mediated by object-based representations. *Acta Psychologica* 197, 52–63.

McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.

McKelvie, S.J. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery* 19(3-4), 1–106.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 101(2), 343–352.

Myczek, K., & Simons, D.J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics* 70(5), 772–788.

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

Olson, I.R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong–object" hypothesis. *Perception & Psychophysics* 64, 1055–1067.

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research* 79, 8–16.

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., & Kosslyn, S.M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *PNAS* 112(33), 10089–10092.

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, W.A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics* 16(2), 283–290.

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

Phillips, I. (2016). No watershed for overflow: Recent work on the richness of consciousness. *Philosophical Psychology* 29(2), 236–249.

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.

Pratte, M.S. (2018). Iconic memories die a sudden death. *Psychological Science* 29(6), 877–887.

Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford: OUP.

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

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

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

Quilty-Dunn, J. (Forthcoming). Perceptual pluralism. *Noûs*.

Quilty-Dunn, J., & Mandelbaum E. (Forthcoming). Non-inferential transitions: Imagery and association. In A. Nes & T. Chan (eds.), *Inference and Consciousness* (London: Routledge).

Robitaille, N., & Harris, I.M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision* 11(12), 1–8.

Rosenthal, D. (2005). *Consciousness and Mind*. Oxford: OUP.

Schneegans, S., & Bays, P.M. (2016). No fixed limit in visuospatial working memory. *Cortex* 83, 181–193.

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

Shepard, R.N. (1978). The mental image. *American Psychologist* 33(2), 125–137.

Sligte, I.G., Scholte, H.S., & Lamme, V.A.F. (2008). Are there multiple visual short-term memory stores? *PloS One* 3(2), e1699.

Sligte, I.G., Vandenbroucke, A.R.E., Scholte, H.S., & Lamme, V.A.F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology* 1(175), 1–10.

Solomon, J.A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision* 11(12), 1–11.

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

Suchow, J.W., Fougnie, D., Brady, T.F., & Alvarez, G.A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, and Psychophysics* 76(7), 2071–2079.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Clarendon Press.

Utochkin, I.S., & Tiurina, N.A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica* 146, 7–18.

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.

Wardle, S.G., Bex, P.J., Cass, J., & Alais, D. (2012). Stereoacuity in the periphery is limited by internal noise. *Journal of Vision* 12(6), 1–12.

Wexler, M., Kosslyn, S.M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition* 68, 77–94.

Wheeler, M.E., & Treisman, A.M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General* 131, 48–64.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology* 69, 105–129.

Winkielman, P., Schwarz, N., Reber, R., & Fazendeiro, T.A. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch and K.C. Klauer (eds.), *The Psychology of Evaluation* (Mahwah, NJ: Erlbaum), 75–89.

Xu, Z., Adam, K.C.S., Fang, X., & Vogel, E.K. (2018). The reliability and stability of visual working memory capacity. *Behavioral Research Methods* 50(2), 576–588.

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

Zhang, W., & Luck, S.J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature* 453, 233–235.

Zhang, W., & Luck, S.J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science* 20(4), 423–428.