CHAPTER TWENTY-SEVEN

# Let's Get Rid of the Concept of an Object File

*Ned Block*

In a typical vision textbook you will see the term "object file" defined as follows: "An object file is a visual representation that "sticks" to a moving object over time on the basis of how and where that object moves, and stores (and updates) information about what that object looks like" (Scholl and Flombaum, 2010, p. 655). Object files are said to function in working memory (Green and Quilty-Dunn, 2021; Quilty-Dunn and Green, 2021) and to ground singular thought (Murez and Recanati, 2016). One claim of this article is that although thought and working memory often preserve some perceptual information, what are called the object files of both singular thought and working memory are fundamentally different from what are called the object files of perception. Indeed, there is reason for doubt that the object files of perception can even *ground* singular thought. The object files of working memory and singular thought enclose the perceptual materials from perceptual object files in a cognitive envelope and in addition transform the perceptual information, often *misrepresenting* some aspects of the stimulus in order to make other aspects of the stimulus easier to use for a specific task. That is the problem for grounding singular thought.

A second thesis of this article is that the object files of perception (that is, perceptual object representations) are iconic in format, contrary to the claims of the "pluralists" who take them to be discursive (Quilty-Dunn, 2020b). The object files of thought and working memory by contrast are conceptual and partly discursive. The term "object file" ambiguously denotes two fundamentally different kinds of entities. We would be better off without the term.

Terminology: I will often call the object files of perception "perceptual object representations" and the object files of working memory "working memory object representations," though I also use the term "object file" when making contact with other writers who use that term.

*Contemporary Debates in Philosophy of Mind*, Second Edition. Edited by Brian P. McLaughlin and Jonathan Cohen.

## 1. Perceptual Object Representations Are Iconic

I will start with the second of the two theses just mentioned, that perceptual object representations are iconic in format. There is a great deal of evidence for the iconic nature of *all* perceptual representations, including representations of object-perception. I won't go over the evidence for perceptual representations that are not object-representations, since it isn't controversial For reviews of some of the evidence about the iconic nature of perceptual representation, see Chapter 5 of (Block, 2022), from which this article is derived, or Quilty-Dunn (2019).

I am going to present two kinds of evidence for the conclusion that perceptual object representations are iconic, direct and indirect. But I'm not saying that these items of evidence "refute" the claim that perceptual object representations are discursive. Pluralists who combine discursive perceptual object representations with iconic perceptual representations of space and spatial features may be able to accommodate these results. The main issue is which view better explains the data, my view that all perceptual representations are iconic or the pluralist view that some are and some are not.

One line of direct evidence for the iconic nature of object perception exploits apparent motion, a phenomenon discovered in the early twentieth century (Wertheimer, 1912). Apparent motion occurs if a subject is shown A in Figure 27.1, followed by B, then A again, then B again, and so on. Subjects report seeing motion. At high rates of flicker between A and B, motion will be seen without intermediate stages. (This is called "phi.") At slower flicker rates, subjects see the trajectories of the moving objects with intermediate stages clearly visible. Subjects report seeing objects of one color or shape transforming into objects of another color or shape. (That phenomenon is called beta motion.) It should be said that subjects do not confuse apparent motion with real motion, but apparent motion still looks like motion Sperling et al., 1985).

Most subjects will see the motion in D in Figure 27.1 rather than the motion in C because the primary determinant of the motion is the visual system's drive to minimize the distance between the items. The effect on apparent motion of path length

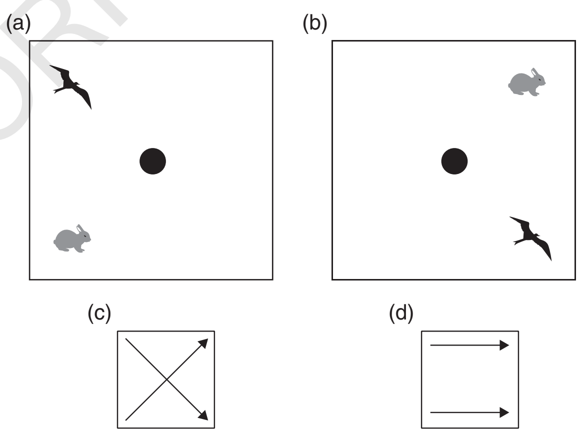


**FIGURE 27.1** IF A AND B ARE QUICKLY ALTERNATED, ONE SEES APPARENT MOTION, USUALLY AS DEPICTED IN D. THANKS TO SUSAN CAREY FOR THE FIGURE (CF. CAREY, 2009, P. 73).

has been estimated to be 15 times the strength of the effect of the shapes of the items involved (Flombaum and Scholl, 2006). The visual system "prefers" not to see a bird turning into a rabbit, but that "preference" is balanced against the stronger "preference" for shorter distances of motion. (This is sometimes called the principle of spatiotemporal priority.) So, subjects will see a bird crossing the screen from left to right, gradually changing into a rabbit at the top right, and the opposite transformation on bottom. The larger the difference between the paths, the more likely the subject is to see the shorter motion (Nakayama et al., 1995). However, if the paths are roughly equal, shape counts.

Path length and shape work together in an integrated manner. The direction of motion depends in a smooth way on the distance between the items. See Figure 27.2 in which the gradual nature of this type of transition is graphed. The gradual transitions are indicative of the analog mirroring of iconic representation. The integration of smoothly varying spatial factors with factors involving object representations suggests that these are not fundamentally different kinds of representations, as would be expected if object representations in perception are discursive whereas other representations are iconic. It would be possible to combine discursive representation of objects with a spatiotemporal representation system, but to the extent that spatial and spatiotemporal effects saturate object representations, that view is less attractive.

The apparent motion stimuli just described are ambiguous in the sense that there are two very different representations that the visual system will compute in different contexts. When stimuli are ambiguous in this sense, cognitive and conceptual factors can affect which representation the visual system computes. This is a ubiquitous kind of cognitive penetration. So, one should not be surprised if cognitive information influences which kind of motion the subject sees.

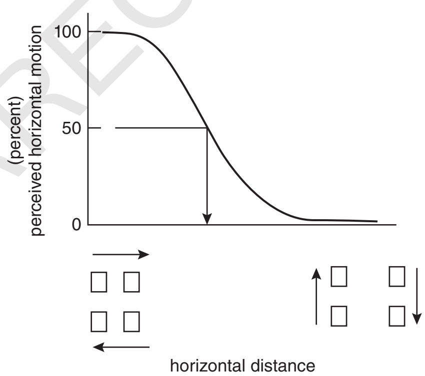


**FIGURE 27.2** THE LIKELIHOOD OF SEEING HORIZONTAL (RATHER THAN VERTICAL MOTION) IN APPARENT MOTION DISPLAYS THAT ARE VARIANTS OF THE ONE IN THE PREVIOUS FIGURE. THE HORIZONTAL AXIS SHOWS HORIZONTAL DISTANCE, WHEREAS THE VERTICAL AXIS GRAPHS THE LIKELIHOOD OF PERCEIVING HORIZONTAL MOTION RATHER THAN VERTICAL MOTION (FOR EXAMPLE, THE BAT ON THE TOP LEFT TURNING INTO THE RABBIT ON THE TOP RIGHT AND THE CORRESPONDING TRANSFORMATION ACROSS THE BOTTOM OF THE SCREEN). WHAT THE GRAPH SHOWS IS THAT AS HORIZONTAL DISTANCE GETS GREATER, SUBJECTS ARE LESS LIKELY TO SEE HORIZONTAL MOTION. FROM NAKAYAMA ET AL. (1995). THANKS TO KEN NAKAYAMA FOR THIS FIGURE.

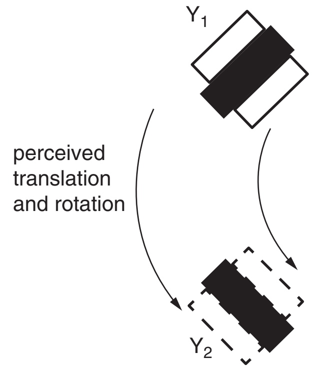


**FIGURE 27.3** A CLOCKWISE ORIENTED BAR CAN BE SEEN TO ROTATE TO A COUNTER-CLOCKWISE ORIENTED BAR IN APPARENT MOTION. FROM NAKAYAMA ET AL. (1995). THANKS TO KEN NAKAYAMA FOR THIS FIGURE.

So why does apparent motion constitute evidence for iconic object-seeing as opposed to just iconic seeing of shapes? One relevant manipulation uses pairs of white bars that protrude from their black background and differ in orientation by 90º between the left and right displays, as in Figure 27.3. Subjects see the bars as rotating back and forth (instead of birds changing into rabbits). Note that the bars appear to rotate gradually. That is, the subject sees the intermediate orientations. The fact that subjects see intermediate stages of rotation suggests that the representations are part of a system that mirrors rotation operations on actual objects—again the analog mirroring characteristic of iconic representation. See Figure 27.3.

The display is viewed via an apparatus that allows for independent manipulation of what is sent to each eye. Whether the white bars emerge from the background in the manner of objects is manipulated by changing binocular disparity cues. If the bars look like parts of a squarish shape instead of like protruding objects, then there is a visual experience of vertical motion but no visual experience as of rotation (Nakayama et al., 1995). If there is no apparent object, then there is no rotation. To the extent that shapes are involved, they are not 2D shapes, since the 2D outline is the same whether or not the display looks like parallel bars.

What makes these representations perceptual is that the bars look like they are moving and rotating. What suggests they are iconic is the presence of smoothly varying intermediate stages of rotation and translation.

In his contribution to this volume, E.J. Green argues that any theory—iconic or noniconic—would have to predict the apparent motion observed by Nakayama, so it provides no evidence for the iconic theory. But flickering images of the sort used by Nakayama need not produce apparent motion; indeed, if the flicker rate is sufficiently high or low, or the distance is too great, there is no apparent motion. Further, in some conditions, the first stimulus will be seen as expanding into the second stimulus. And the greater the distance between the two stimuli, the longer the time gap required to see motion, mirroring typical speeds in the actual world (Korte's Law). The fact that the expansion, rotation, and translation is observed at all—in any circumstance—is not surprising on the iconic account but is surprising on the discursive account.

To avoid misunderstanding, I am not saying that iconic and discursive elements cannot be combined in a single representation. An iconic depiction of the shape of a street can be combined on a paper map with the name of the street. But notice that this is possible because the name itself has spatial properties: its location, orientation, and size. Indeed, the name of Doyers Street, a storied 200 foot long curved street in Chinatown in southern Manhattan, is often curved like the street on maps of Manhattan. A paper map of Manhattan uses spatial properties and relations instantiated on the paper to represent spatial properties and relations on the island of Manhattan, but brain representations do not represent space with space. Will the advocates of discursive perceptual object representations say that the putative discursive object-representations brain representations have an analog of spatial properties comparable to the spatial properties of the name "Doyers Street"? We can call this suggestion the Doyers Street gambit.

A brain map of Manhattan uses an analog of spatial extent realized in the place cells and grid cells of the brain. Although some startling discoveries have been made recently about place cells and grid cells, how the brain represents space is still largely a mystery. But we can refer to the analog of space in the brain as "place–grid–cell–space." Advocates of the Doyers Street gambit could say that the discursive object representations also instantiate place–grid–cell–spatial properties, just as the name "Doyers Street" instantiates real spatial properties. This is an interesting and adventurous hypothesis, but I know of no evidence for it.

A defender of the Doyers Street gambit might say that the Nakayama result just described is evidence for it. To take this claim seriously we would need independent evidence for both the Doyers Street gambit and discursive perceptual object representations. As we will see in the second half of this article, the evidence that has been offered for discursive object representations applies to working memory, not perception.

The apparent motion results are direct evidence for the iconicity of object perception because they exhibit the smooth variation indicative of analog mirroring. I now turn to indirect evidence that perceptual object representations are iconic. More specifically, I will consider evidence that object representations in perception are so tightly integrated with other iconic representations in perception, notably spatial representations, as to put pressure on pluralism. The Doyers Street gambit is one way of resisting that pressure, but perhaps there are others.

The first type of evidence I will consider involves object-based attention. (See Scholl, 2001 for a review.) Perceptual attention can be divided into three types, depending on what is attended to: object-based attention, in which what is attended to is an object; spatial attention, in which what is attended to is a region of space; and feature-based attention, in which what is attended to is a property of objects or regions of space. The word "attention" is used in many different ways, including speaking of attention to items that cannot be perceived directly. But the kind of attention being discussed here is perceptual in that it is tightly integrated into perceptual systems and it obeys perceptual regularities such as a phenomenon known as divisive normalization (Bloem and Ling, 2019). I'll give an example below.

Subjects show faster and more accurate processing for features belonging to the same object than for features belonging to different objects, showing that perceptual object representations are involved in the control of attention. One type of experiment that shows this is illustrated in (a) in Figure 27.4. If subjects see a cue at C, they are

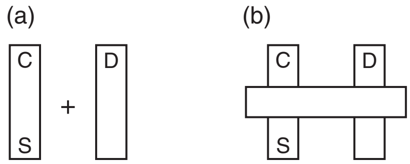


**FIGURE 27.4** IF CUED TO C, SUBJECTS ARE FASTER TO DETECT TARGETS AT S THAN AT D EVEN THOUGH S AND D ARE EQUIDISTANT FROM C. THANKS TO BRIAN SCHOLL FOR THIS FIGURE. SEE SCHOLL (2001).

faster at detecting a target on the same object at S (for "same") than an equidistant target on another object, D (for "different"). And this holds whether or not there is an occluder, as in (b). The fact that even an occluded object is subject to object-based attention indicates that the subjects are seeing the occluded objects as objects. This is not in itself evidence for iconicity, but that is coming in the next paragraph.

Here is the evidence for iconicity: Object-based attention is a matter of degree. Objects such as the vertical rectangles of Figure 27.4 show less of an object effect if the rectangles are altered so as to be less "good" as objects, for example if the bottom horizontal bar of the rectangle is deleted (Marino and Scholl, 2005). If there was a radical format difference between object-perception and other perception, one would not expect such gradual effects. The difference between discursive and iconic representation is not a matter of degree.

Pluralists may postulate links between spatial attention and discursive object representations. But results of the kinds just described put pressure on them to justify the extra assumptions involved in such explanations.

Another feature of object-based attention that should trouble pluralists is that attention "spreads" within an object from a cue at one end of the object (as in Figure 27.4) (Richard et al., 2008; Zhao et al., 2013). Spreading suggests representational analogs of the spatial extent of the object that mirror the spatial properties of the object.

A similar point about the integration of perceptual object representations with spatial representation applies to a phenomenon known as inhibition of return. Inhibition of return was demonstrated in a paradigm in which there are three boxes, a central box and two flanking boxes. One of the flanking boxes (say the one on the right) is cued (e.g., it suddenly brightens), so attention is drawn to it. Then the central box is cued. If a target is presented in the right box within 150 ms, there is a detection advantage (due to the residual attention to the right box), but if a target is presented in the right box after 300 ms, there is a disadvantage in detection. The upshot—now verified in many paradigms—is that the attention system is inhibited from attending to *something* that has recently been attended for as long as 3 seconds.

But what is that something? Is it an area of space, a scene, an object, or what? The answer is areas of space and objects both show inhibition of return, not surprising since there is both object-based attention and spatial attention. The object-based effect is exhibited when what is inhibited is a return of attention to the object in which the cue originally occurred (Tipper et al., 1999). This is verified by varying other properties such as location, showing an independent effect of the same object. Object perception and spatial perception function similarly, a puzzling fact if they are fundamentally different in format.

With inhibition of return as with object-based attention, there is a gradient of effects within an object, with the strongest effect at the cued location within the object, and weaker effects in the same object but further away from the cue (Klein and Ivanoff, 2008). This shows integration of object-based effects with spatial effects, again providing evidence against the view that there is a difference in kind between object perception (allegedly discursive) and other perception (iconic). Again, object perception is integrated seamlessly with spatial attention, something that would call for explanation if object representations and spatial representations involved representations of different formats as the pluralists claim.

A different kind of support for iconicity in object-perception involves visuospatial neglect, a syndrome in which subjects fail to attend to objects on one side of the body. The point of discussing visuospatial neglect is that it reveals that object representations, spatial representations, temporal representations, and numerical representations are tightly coupled in overlapping systems, counting against the claim that object representations have a different kind of format from other perceptual representations. Of course representations of different formats can be linked, but the question is whether postulating such linkages has independent support.

In one kind of visuospatial neglect, subjects ignore or fail to consciously see the left side of the visual field. Patients fail to eat the food on the neglected side of their plates, fail to dress the neglected side of the body, and so on. This kind of neglect is based in one form of egocentric perception. When left-sided neglect patients are asked to bisect a horizontal line, they put the bisecting mark to the right of the midpoint. Interestingly, some left-sided neglect patients show the same effect for imagined lines. The size of the rightward drift in bisecting lines is proportional to the length of the line, i.e., larger displacements with larger lines. This dependence on degrees is indicative of iconic representation. These effects are all matters of degree and interact with many spatial visual features, again providing problems for pluralism. Interestingly, for very short lines, there is an effect in the opposite direction, the "crossover effect" (Zorzi et al., 2012), and this crossover effect also appears in other perceptual activities that use the same system.

Neglect often involves inattention to, and perhaps lack of perception of, one side of space. But it often applies to one side of individual objects that have salient axes, showing again that the control of attention depends on spatial aspects of perception, showing integration of object representations with spatial representations. That integration would require explanation if the two kinds of representations were of different formats, one iconic, one discursive. Many patients neglect the left sides of objects all over the visual field, showing some influence of allocentric spatial representation (Beschin et al., 1997; Tipper and Behrmann, 1996). And patients often neglect the initial letter or segment of a word, even if the word is presented vertically; neglect the Western Hemisphere even in an upside-down map; or neglect the left side of a face even in an upside-down photograph (Bisiach and Luzzatti, 1978; Caramazza and Hillis, 1990). Again, these effects are matters of degree and interact with spatial features.

A classic demonstration of object-based neglect involved barbells, two circles connected by a line. Neglect patients had trouble with detecting targets on the left circle, but when the barbell was rotated so that the left circle had moved to the right, many patients showed flipped results, with more trouble on the right circle. This effect was only observed if the barbell was a single object: if the line between the circles was omitted, there was no such effect (Tipper and Behrmann, 1996). Again we see seamless integration of object perception with spatial perception, a surprising result if the two have entirely different formats.

Another classic demonstration of object effects in neglect is shown in Figure 27.5. A patient who was asked to copy a picture left out the left side some of the individual objects in the picture. See Walker, 1995 for other examples. Once again we see that the perception of objects is part of a spatial representation system, something that would be in need of explanation if object representations were discursive but spatial representations were iconic.

As just mentioned, neglect also extends to "numerical space." Left-sided neglect patients asked what number is halfway between 2 and 6 skew their answers toward 6. Strikingly, the crossover effect just mentioned also applies to numerical space (Zorzi et al., 2012). Similar results apply to temporal estimation problems for neglect patients (Bonato et al., 2016). A further crossover effect is that normal subjects who were asked to estimate additions or subtractions of dots showed a leftward bias on the number line (that is, they underestimated) for small numbers of dots, but a rightward bias for large numbers (overestimating) (Zorzi et al., 2012). The explanation usually given for this kind of result is that spatial representation is co-opted for the numerosity system. There is a great deal of evidence for this. For example, if subjects are given the task of pressing one button if the number referred to by a presented digit is bigger than 5 and another button if it is less than 5, subjects are faster for 7 than for 6, faster still for 8, and faster still for 9, with similar results obtaining for digits on the other side of 5 (Dehaene, 2011). These results show that object representations, spatial representations, temporal



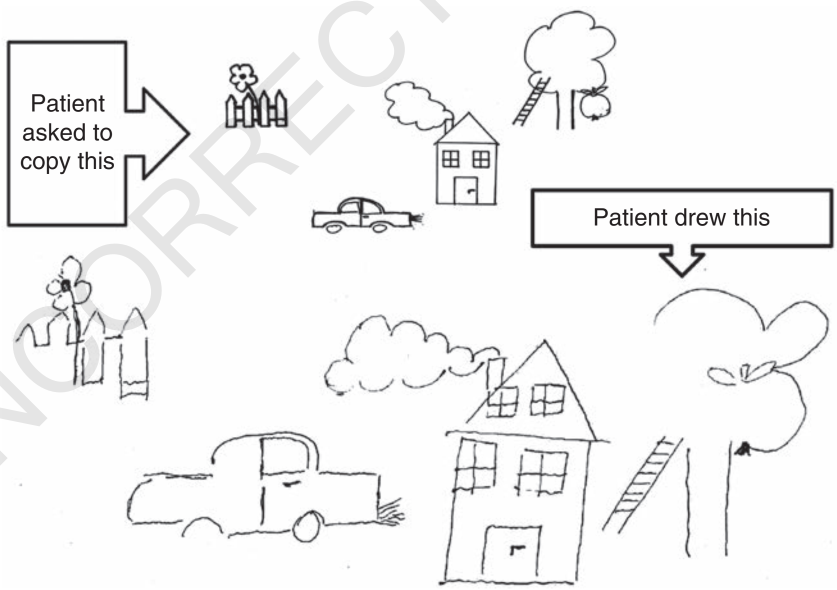


**FIGURE 27.5** A VISUO-SPATIAL NEGLECT PATIENT WAS ASKED TO COPY THE TOP PICTURE. THE BOTTOM PICTURE IS THE PATIENT'S ATTEMPT. NOTE THAT THE PATIENT LEAVES OUT SOME OF THE LEFT SIDE OF SOME OF THE INDIVIDUAL OBJECTS IN THE PICTURE. THANKS TO JAMES DANCKERT FOR THIS PICTURE.

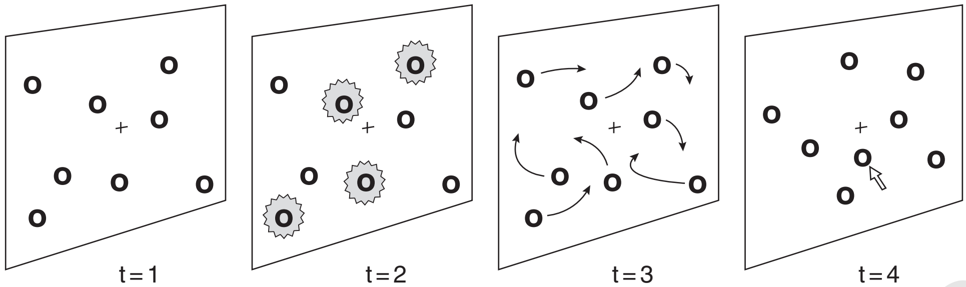


**FIGURE 27.6** SEQUENCE OF EVENTS IN A BASIC MULTIPLE OBJECT TRACKING EXPERIMENT. EIGHT DISKS APPEAR ON THE SCREEN; FOUR ARE INDICATED AS THE ONES THE SUBJECT IS SUPPOSED TO TRACK; THE INDICATIONS OF THE DISAPPEAR; AND THE DISKS MOVE RANDOMLY. AT THE END THE SUBJECT IS ASKED TO IDENTIFY THE FOUR THAT WERE FIRST CUED. MOST PEOPLE CAN TRACK FOUR OBJECTS FAIRLY RELIABLY. FROM PYLYSHYN (2007) UNDER A CREATIVE COMMONS LICENSE.

representations, and numerical representations are tightly coupled in overlapping systems, putting a burden on those who think that object representations have a different kind of format from other perceptual representations.

In the next section, I will be urging caution about drawing conclusions about perceptual representations from evidence about the remnants of perceptual representations in working memory. To the extent that inhibition of return involves working memory, that caution applies to the results just mentioned. The only way in which working memory would add perceptual features rather than subtract them is via the imposition of perceptual imagery, but it would have to be shown that imagery is involved in the experiments I am talking about.

Finally, in the multiple object tracking paradigm, not only can subjects track about four disks but also they can track the average position of the disks. See Figure 27.6. A number of disks (eight in the figure) are shown on a screen. Four of the disks blink or are otherwise indicated, and then the disks move randomly. The subject is supposed to track those objects as they move about in a random way. Most subjects can track about four objects if they do not move too quickly. Subjects turned out to be able to track the centroid of the target disks, but they could also track the centroid of the distractor disks. And they could do this even if their attention was drawn off by a difficult secondary task of counting the number of times the disks crossed some lines on the screen (Alvarez and Oliva, 2008). This experiment suggests that perceptual object representations integrate with spatial representations even though working memory is also involved in multiple object tracking. In sum, there is plenty of evidence that perceptual object representations are of a piece with other sorts of spatial perception, putting pressure on the view that object representations are different in format from other sorts of perception.

## 2. Object Files of Working Memory and Thought

The first section of this article argued that perceptual object representations (the so-called object files of perception) are iconic. I now turn to the other main claim of this article, that the so-called object files of singular thought and working memory are fundamentally different from those of perception. Indeed, there is reason for doubt that the object files of perception can even ground singular thought. I will be focusing on working memory at the expense of singular thought. A singular thought can be based on a simultaneous perception, but singular thought can also be based on working memory, and it is the latter case that is relevant to the discussion to follow.

Since we are going to be talking about working memory representations, we should first understand what they are, and that requires contrasting them with other short-term forms of perceptual memory: iconic memory and fragile visual short term memory. I now turn to that contrast.

## Iconic memory

Immediately after the presentation of a stimulus, reverberating retinal activity especially in the rods and also activity in V1 grounds perceptual representation of low-level properties, what is often called "visual persistence" (Coltheart, 1980). As a result, for a few hundred ms, there is a genuine form of memory, "iconic memory," that also represents higher level properties (Pratte, 2018).

The classic experiment demonstrating iconic memory was done by George Sperling. In the Sperling experiment, there is a brief flash of an array of letters separated into rows, e.g., 3 rows of 4 letters each (Sperling, 1960). Subjects report seeing all or almost all the letters but can recall only three or four of them once the display has gone off. However, if one row is cued by a tone within a few hundred milliseconds after the stimulus disappears (a high tone for the top row, low tone for the bottom row, etc.) subjects can recall three or four from any given row, suggesting that they did have a brief visual representation of all the letters. The ratio of total capacity (roughly 3.5 in each of three rows) to capacity without a cue is called the "partial report superiority."

## Fragile visual short-term memory

Victor Lamme's laboratory at the University of Amsterdam demonstrated fragile visual short-term memory in a series of articles (starting with Landman et al., 2003). The experimental paradigm combines the "iconic memory" paradigm of the Sperling experiment with "change blindness." This paradigm shows a greater capacity in fragile visual short-term memory than in working memory but a smaller capacity than in iconic memory.

A recent experiment (Pratte, 2018) suggests that representations in iconic memory undergo a "sudden death" decay, in which the surviving representations maintain the same level of precision rather than decaying in precision as "pool of resources" models would predict. Since the memory capacity found by Pratte decays smoothly from 33 ms to 1000 ms, and since iconic memory does not last more than a few hundred milliseconds, both iconic and fragile visual short-term memory would appear to be involved in this experiment. Since working memory *does* fit the pool of resources model, it would appear to be of a different kind than the earlier stores, as would be predicted by the claim of a format difference. This is one item of information that suggests that while perception is iconic, working memory is discursive (though it can contain iconic remnants in a discursive envelope).

## Working memory

Working memory is a kind of cognitive scratch pad that can be used to manipulate information for cognitive purposes. For example, if you want to reason from the proposition that p and the proposition that if p, then q, you must hold the premises in working memory in order to make the deduction. There can be cognition without working memory, but working memory is necessary for reasoning in which a premise is retained for later use. Presence of a representation in working memory is not "storage" but rather active maintenance.

Working memory is far more robust than either iconic or fragile visual short-term memory. Ilja Sligte found that a white screen (a so-called light mask) obliterated iconic memory but not fragile visual short-term memory or working memory. A pattern mask obliterated fragile visual short-term memory but not working memory (Sligte et al., 2008).

Working memory is generally taken to be controlled by prefrontal cortex on the outside mid-level surfaces (dorsolateral prefrontal cortex). Transcranial magnetic stimulation (TMS) is the application of an electromagnetic pulse to a brain area, creating neural noise. Transcranial magnetic stimulation to visual areas (notably V4) impaired fragile visual short-term memory, and TMS to a cognitive area, the dorsolateral prefrontal cortex, impaired working memory but not fragile visual short-term memory (Sligte et al., 2008; Sligte et al., 2010; Sligte et al., 2011). So these different forms of memory are distinct both at the psychological and neural levels.

There have been many proposals for further fractionating working memory. For example, Justin Wood has argued that working memory can be divided into a view-dependent store with a capacity of roughly four items and a more abstract view-independent store of about two items (Wood, 2009). However, it is unclear whether the view-dependent store might involve fragile visual short-term memory.

For the kinds of stimuli discussed here, working memory has a limit of three to four items. Typically, the three- to four-item limit is observed with small closed-class groups of stimuli. For large open classes of stimuli, many more items can be represented in working memory with diminished precision. See (Endress and Potter, 2015; Endress and Siddique, 2016) for an explanation of the difference between cases in which the three- to four-item limit is observed and cases where it is not.

When working memory representations do not show iconicity, one cannot be sure whether the iconicity was lost in the conceptualization process, but when they do show iconicity, the iconicity derives from perceptual remnants that are contained in the working memory cognitive envelope.

I now move to the evidence that the object representations of working memory are fundamentally different from those of perception even when the object representations of working memory involve perceptual materials. I will mention three points of difference:

1 Perceptual object representations have a higher capacity than working memory object representations.
2 Working memory object representations do not show fundamental computations of perception.
3 Working memory object representations are task specific in ways that perceptual object representations are not.

## Capacity

As just mentioned, the Sperling experiment showed that the more perceptual representations of iconic memory have roughly three times the capacity of working memory. And fragile visual short-term memory has around double the capacity of working memory.

## Fundamental computations

Perception exhibits a canonical computation, divisive normalization. One manifestation of this computation is center–surround suppression, in which perception of a central disk is suppressed by similar properties in a doughnut surrounding it. This is illustrated in Figure 27.7. When the disk and the doughnut were presented one at a time, with the first stimulus maintained in working memory, there was no center–surround suppression (Bloem et al., 2018). This result suggests that a basic computational feature of perception is absent in perceptual working memory.

## Task specificity

A recent experiment showed how two quite different perceptual representations can be converted into the same working memory representation if the subjects' tasks are appropriately similar. Yuna Kwak and Clay Curtis (2022) used two kinds of stimuli on different trials, oriented gratings (Gabor patches) and clouds of moving dots. Subjects' task was to indicate the orientation of the grating or the direction of the moving dots after a delay period. They scanned the subjects using fMRI during the delay period prior to doing the tasks. The first result was that decoding trained on the grating task also worked on the dot task and vice versa. This fact shows that the working memory representation was sufficiently abstract as to be common between the two perceptions. The second result homed in on what the actual shared representations were.

They developed a visualization technique that allowed them to transform the brain representations into a display on a screen that would have produced that brain activation. And the result was that both the representations of the grid and the dot motion transformed to an oriented stripe. The representation of the cloud of dots abstracted away from the representations of the individual dots and the representation of the

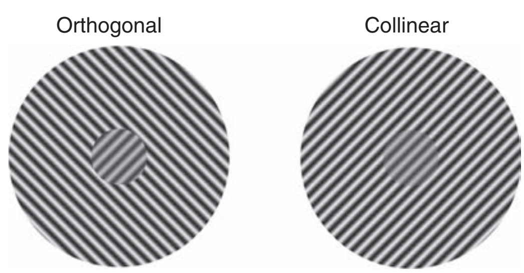


**FIGURE 27.7** ILLUSTRATION OF THE EFFECT OF DIVISIVE NORMALIZATION. THE CENTER DISK IS THE SAME ON BOTH SIDES BUT LOOKS LOWER IN CONTRAST ON THE RIGHT BECAUSE OF SURROUND SUPPRESSION THAT DEPENDS ON SIMILAR ORIENTATION OF THE DISK AND ITS SURROUND. THANKS TO SAM LING FOR THIS FIGURE.

grating abstracted away from the spatial frequency and contrast of the grating. What this experiment shows is that working memory representations depend not only on the stimulus but also on the task. A similarity in task can lead to a similarity in working memory representation even if the percepts differ. This experiment also shows that there is a sense in which the perceptual materials in a working memory representation can "misrepresent" the stimulus for the sake of usability of other information. In the case of the moving cloud of dots, there is no "stripe" in the stimulus.

I mentioned earlier that the "object files" of working memory may not even be grounded in perceptual object representations as is often claimed (Recanati, 2012). As just noted, the task specificity of working memory object representations can transform the perceptual information in a perceptual object representation. The working memory representation that is derived from perception does not just *discard* some information—it *transforms* the information for task-specific use. In the Kwak and Curtis experiment, the information about the dots and the spatial frequencies are transformed into a quite different representation.

I have described three important differences between the perceptual information in "object files" of perception and of working memory. I have omitted others that would have required more extensive discussion, e.g., fineness of grain.

I now move to a discussion of a concrete example of the difference keyed to the issue of iconicity.

## Arguments against iconic object representations that are based on perception and memory of objects

The evidence provided by E. J. Green and Jake Quilty-Dunn (Green and Quilty-Dunn, 2021; Quilty-Dunn, 2016, 2020a, 2020b) for discursive object files representations is based on the "object-specific preview benefit" or OSPB (Kahneman et al., 1992). They use the OSPB to argue that the format of object files is discursive rather than iconic.

In one version of the OSPB, two boxes are on the screen containing pictures, for example, pictures of an apple or a loaf of bread. The pictures disappear and the boxes move. Then a picture appears in one of the boxes, either of an apple, a loaf of bread, or something else. The subject's task is to name the object. Subjects are faster in naming an apple if a picture of an apple was in either one of the boxes. (So far, that is just "priming," a phenomenon whereby something just seen or appropriately related to something just seen is easier to recognize.) However, and this is the OSPB, subjects are faster still if the apple is in the very box that it started in, even if that box has changed sides.

Another version of the OSPB is illustrated in Figure 27.8. Words are presented in boxes. Then the words disappear and the boxes move as indicated for 1.5 seconds. Then a picture appears in one of the boxes which the subject is supposed to name. The result is that the subject is faster to name the apple if the box the apple is in was the one in which the word "apple" had appeared. Green and Quilty-Dunn take this result to indicate that the perceptual representation—the "object file" that underlies this ability—is a symbol that has the content apple and is bound to semantically linked information in a separable, nonholistic fashion. For example, the object file might simply be a discursive list of linked properties.





0005464032.INDD 506 09-10-2022 11:32:11

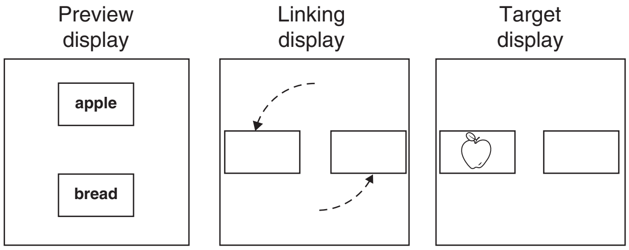


**FIGURE 27.8** VERSION OF THE OBJECT-SPECIFIC PREVIEW BENEFIT THAT SHOWS THAT OBJECT FILES CONTAIN BOTH LINGUISTIC AND PICTORIAL INFORMATION. FROM QUILTY-DUNN (2016), BASED ON GORDON AND IRWIN (2000). THANKS TO JAKE QUILTY-DUNN FOR THE FIGURE.

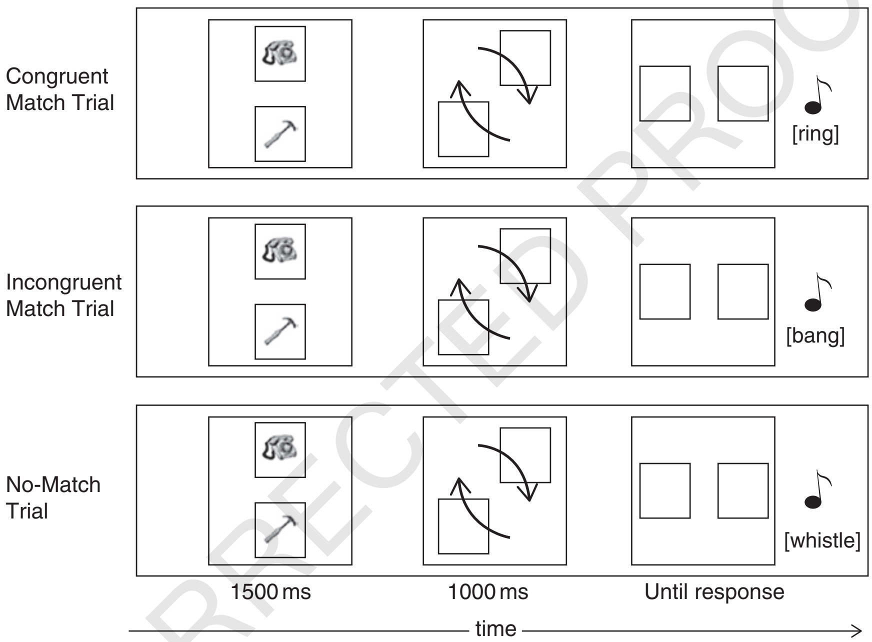


**FIGURE 27.9** VERSION OF THE OBJECT-SPECIFIC PREVIEW BENEFIT. FROM JORDAN ET AL. (2010), P. 495, WITH PERMISSION OF TAYLOR AND FRANCIS, HTTP://WWW.TANDFONLINE.COM



As Green and Quilty-Dunn note: There is also an OSPB from lowercase words to uppercase versions of the same word, for instance, from "bread" to "BREAD." That, they say, shows that the representation of the word abstracts from shape properties and so cannot be iconic. This is part of the abstractness argument against iconicity.

In another variant, pictured in Figure 27.9 (Jordan et al., 2010), two boxes are presented with pictures in them, say a hammer and a whistle. The pictures disappear and the frames then move so that the boxes can end up on a different part of the screen from which they started a second later. Then the subject hears a sound and has to say whether the sound matches one of the pictured items. Subjects are faster if the sound matches the object that was in the box that is now on the side that the sound is coming from. For example, in the top row of Figure 27.9, the sound of ringing matches the picture of a telephone. Subjects are fastest for the "congruent" situation in the top row. The sound of banging in the second row does not match but was present (bringing with it the speed increment of priming.). That row comes in second. The slowest is the bottom row in which the sound—a whistle—does not match either of the pictures.

Green and Quilty-Dunn conclude that object files involve discursive symbols that abstract away from modality-specific information in an amodal format.

How do we know that the representations involved in the OSPB are working memory representations? In the experiments pictured in Figures 27.8 and 27.9, there is a delay between the first stimulus and the last stimulus. In the experiment with the word "apple" and the picture of the apple, the delay is 1.5 seconds. In the experiment with the sounds matched to objects, the delay is 1 second. A further experiment showed that the OSPB was preserved even if the blank period lasted as long as 8 seconds (Noles et al., 2005).

However, iconic memory of the perceptual kind exhibited in the classic Sperling experiment lasts only a few hundred milliseconds, so the OSPB representations cannot be representations of iconic memory. As I mentioned earlier, there is another kind of perceptual memory, "fragile visual short-term memory" (Lamme, 2016). Fragile visual short-term memory has been shown to last up to 4–5 seconds, but never longer. In addition, fragile visual short-term memory has been shown in static displays but not to my knowledge in moving displays. Further, fragile visual short-term memory is, well, fragile, and easily overwritten. The motion in these displays may be enough to damage fragile visual short-term memory representations. These considerations strongly suggest that the kind of memory involved in the OSPB is working memory, the least perceptual of the three kinds of visual short-term memory.

I think the OSPB concerns working memory representations that have conceptualized *remnants* of perception in a cognitive envelope and that there is no evidence that the abstractness shown in the OSPB can be ascribed to perception as opposed to the cognitive aspects introduced by the conceptualization and the cognitive envelope.[1] So the crucial issue concerns whether the OSPB involves *perceptual representations of the sort that are involved in perception itself*.

The first thing to note about the OSPB is that after the picture or word disappears, *the subject is no longer in a state that seems like seeing them*. They see the boxes that are rotating, not what was originally in the boxes. I have looked at OSPB displays. Once the letters disappear one just sees the boxes moving with no awareness of the letters. The fact that the subject does not see the picture or word by itself shows that we should be suspicious of any claim that in the blank period the subjects have perceptual representations of the items that were originally in the boxes.

There is no reason to think that the subjects in this experiment have any visual phenomenology of the items in the boxes during the blank period. The iconic memory and fragile visual short-term memory mentioned above are said by subjects to be phenomenal, but I don't know of any reports of phenomenology of working memory in experiments that contrast iconic memory, fragile visual short-term memory, and working memory, such as the experiments by Victor Lamme's group in Amsterdam (Lamme, 20 03, 2004, 2006, 2016, 2018; Landman et al., 2003; Pinto et al., 2013; Pinto et al., 2015; Sligte, 2011; Sligte et al., 2008, 2009, 2010, 2011).

Further, it takes 1.5 seconds for a subject to generate a mental image. In the 1 second that the boxes are rotating as depicted in Figure 27.9 there would be no time to generate a mental image of the hammer or telephone. Both of these points suggest a difference in kind between the "object files" of working memory and the "object files" of perception.

Consider the top row of Figure 27.9. On the left we see a box with a telephone on the top and a hammer on the bottom. Then the pictures disappear and the boxes move. They move for 1 second as depicted in Figure 27.9, but as I mentioned the time lag can be as long as 8 seconds. Then a sound plays. As I mentioned, the subjects are not seeing the telephone or the hammer. They just see the empty boxes moving. If the representations of the telephone and the hammer are real perceptual representations, perhaps they would be *unconscious perceptual representations*.

Now I happen to be a fan of full perceptual representations in unconscious perception (Peters et al., 2017; Phillips and Block, 2016). But one lesson of recent work on unconscious perception is that it is harder to produce than was earlier thought. Megan Peters and Hakwan Lau (Peters and Lau, 2015) did an informal survey of people who work on perception and found that though most thought unconscious perception exists, most also thought that unconscious perception had not been demonstrated to exist.

Note the contrast with the evidence presented earlier for iconic object representations in perception. Recall the apparent motion case, in which a bird is seen to be moving and then changing into a rabbit. The trajectory and bird/rabbit shapes are consciously experienced even though they are not on the screen. And in the Nakayama experiment, the moving object is seen to rotate even when nothing is rotating on the screen.

More illumination on the difference between perceptual object representations and OSPB representations can be found in a phenomenon known as the tunnel effect, in which an object disappears behind a narrow occluder (the "tunnel") and an object emerges from the other side of the tunnel. The second object may differ in color, shape, and kind from the first (e.g., a lemon goes in and a kiwi goes out). If the tunnel is narrow enough relative to the size of the object moving through it (best results are achieved when the occluder is the width of the object) and the motion is fast and smooth enough, subjects see a single object going behind, changing shape and color, and emerging from the other side.

An early article on the effect from the days in which first-person descriptions were routinely used in perception journals, says that "an absolutely compelling impression of continuous and uniform movement can be produced ... all the observers agree that the movement behind the tunnel is as 'real' as" motion without the occluder (Burke, 1952, p. 124). As the relative length of the tunnel increases and the speed decreases, subjects can still track the moving object using a working memory representation, but they no longer experience motion. My point is that when the representation becomes a working memory object representation rather than a perceptual representation is when consciousness fades. I have seen no report of awareness of the objects in the OSPB.

In the OSPB, perceptual representations are conceptualized in working memory. As we saw in the Kwak and Curtis experiment described earlier, we can expect that conceptualization in working memory will produce a format difference that is keyed to the task. Kwak and Curtis describe a format change in the direction of abstraction.

In paradigm cases, what happens when a perceptual representation is conceptualized is that it is broadcast in the global workspace. "On top of a deep hierarchy of specialized modules, a "global neuronal workspace," with limited capacity, evolved to select a piece of information, hold it over time, and share it across modules" (Dehaene et al., 2017, p. 489). When a piece of information is held and shared in the global workspace, the perceptual information is enclosed in a cognitive envelope. Further, the perceptual information in the cognitive envelope is itself transformed as explained earlier.

One reflection of the fact that working memory representations that contain perceptual materials are more abstract than perceptual representations is a difference in "tolerance." Tolerance is a term used in the memory literature to describe whether the subject in a memory experiment regards an object as the same as one that was seen earlier. Visual long-term memory in humans is famously tolerant, especially in comparison to artificial intelligence programs that have a great deal of difficulty recognizing an object as the same one seen earlier but from a different vantage point angle (Schurgin and Flombaum, 2018). Schurgin and Flombaum showed that visual working memory is very tolerant, indeed substantially more tolerant than visual long-term memory. But perceptual representations are viewpoint-specific.

An indication that the relevant features of object representations that is exploited in these experiments are cognitive aspects of the representations is that the links adverted to via the term "match" above may involve inference. The sound of a piano is said to "match" the picture of the piano. The sound of a dog barking is said to "match" the picture of the dog. Likewise, for a "match" between a sound and a picture of a train. Matching in this sense is inferential rather than perceptual. Jordan, et al. are aware of this possibility and they tried to hamper one form of inference by asking the subjects to memorize four digits presented before each trial. After the subjects give the matching response, they were to repeat the four digits. This was supposed to interfere with a strategy of coding the pictures verbally. But the matching can be inferential even if that inference is not accomplished in a verbal system. The subject does not have to state the premise and conclusion explicitly for the process to be inferential.

Jordan et al. end up seeming to favor the hypothesis that I am suggesting, that the result concerns the working memory aspect of object files rather than their perceptual aspects:

> Alternatively, object file representations may not be intimately tied to any particular sensory modality. In this sense, object files should not be conceived of as visual or auditory, but rather as abstract amodal representations. Although no evidence to date can conclusively tease apart these alternatives, the existence of nonvisual object processing ... may support the latter hypothesis. Such multisensory information could be bound in working memory via the episodic buffer's linking of visual and verbal material. (Jordan et al., 2010, p. 501)

Jordan et al. seem to be thinking that the results reflect abstract amodal aspects of working memory rather than perception.

Another type of evidence for discursive perceptual object representations presented by Quilty-Dunn (2020b) involves transsaccadic memory. A saccade is a fast, ballistic movement of the eye, usually occurring two to three times per second. Visual processing is greatly reduced during a saccade, so the visual system must rely on memory to





0005464032.INDD 510 09-10-2022 11:32:13

encode which objects in the scene after the saccade are the same as the ones in the scene before the saccade. If I am watching a horse race, my visual system must keep track of which horse is which as I saccade back and forth between them.

There are indications that the same kind of object files that figure in the OSPB also have a role in the transsaccadic memory representations that are involved in tracking objects and guiding eye movements to them (Schut et al., 2017). As I understand him, Quilty-Dunn takes this as evidence that the object representations that are indexed by the OSPB are perceptual.

However, there is ample evidence that transsaccadic memory representations are working memory representations. For example, Irwin (1992) did an analog of the Sperling experiment for transsaccadic memory. In Sperling's experiment, subjects could recall only 3 or 4 items from an array of 12 but they could also recall 3 or 4 from any given row if cued after the stimulus had disappeared. Their iconic memory capacity was roughly 3 × 3.5, i.e., 10.5 letters. In Irwin's transsaccadic memory version, subjects saw an array of letters at one fixation but were not given the cue until after they had moved their eyes to the new location. The result was that their memory capacity was about a third of that revealed in Sperling's experiment. That suggests that the kind of memory involved is working memory, since that is a typical working memory performance for letters as stimuli.

Irwin found that a mask presented within 40 ms of the stimulus had a significant impact, but there was no effect at periods longer than 40 ms (120 ms and 950 ms), suggesting that a visual icon is present but only very briefly, being wiped out by the saccade. (In the Sperling phenomenon, iconic memory lasts 200–300 ms.) Irwin concludes (p. 311), "It appears that transsaccadic memory retains visual aspects of a stimulus but perhaps for a brief time only."

(Irwin and Andrews, 1996) used a different procedure with similar results. Subjects saw an array of 6–10 colored letters in the center of the visual field together with a peripheral target to which subjects were supposed to move their eyes. The subjects then saccaded to the peripheral target at which time the central array disappeared and the peripheral target was replaced by an indicator of one of the positions that had been occupied by a letter. Subjects were supposed to report the letter and its color. The subjects could only do this via memory of the presaccade fixation, so this task uses transsaccadic memory. They could report the letter and its color for only three to four locations, the typical signature of working memory with this kind of stimuli.

The fact that transsaccadic memory contains only some perceptual elements is widely appreciated. For example, Gordon et al. (2008) describe the Irwin and Andrews experiment as follows (p. 667):

> Contrary to what would be expected if transsaccadic memory had a very high capacity, Irwin and Andrews found that the subjects could report the color and identity of only 3–4 of the letters in the array. Interestingly, this capacity was very similar to that reported by Irwin (1992), who required subjects to report letter identity alone. Irwin and Andrews concluded that transsaccadic memory consists primarily of integrated object representations (which may include a number of object features), along with residual activity in the feature maps that underlie sensory processing. Subsequent work in which more complex stimuli were used also suggests that transsaccadic memory consists primarily of representations of a small number of objects in the scene).

The point by Gordon et al. that the result by Irwin and Andrews (1996) and Irwin (1992) both come up with the limit of three to four even though one involved reporting two properties and the other reporting just one property comports with a well-known property of working memory, namely that its limit of three to four (with certain kinds of stimuli) is a matter of three to four items, independently of the number of features of those items. (To avoid misunderstanding, recall that the three- to four-item limit in working memory applies only with certain kinds of stimuli, including alphanumeric characters.)

There is also evidence of long-term memory involvement in transsaccadic memory. Hollingworth and Henderson (2002) did an experiment in which subjects fixated naturalistic scenes while their fixations were being tracked with an eye-tracker. In one of their experiments, subjects were given a change-detection task. The experimenters decided on one of the objects in the scene as the target object. When subjects happened to fixate on it for more than 90 ms their attention was drawn to another part of the scene, and later a green square appeared, obscuring the object. Subjects had been instructed to fixate the green square and then decide as between two scenes which scene had the original object. Subjects were more than 80% correct even though numerous fixations had intervened between the original fixation and the fixation of the green square. The average number of intervening fixations was 4.6, and even with 9 fixations there was no sign of decreasing accuracy. The upshot is that there is a form of transsaccadic memory that integrates over multiple fixations. In other experiments, subjects retained object files for as long as 30 minutes. The authors conclude that there can be what they call "long-term memory object files."

As I understand Quilty-Dunn, he takes these transsaccadic memory results to indicate that the perceptual object representations before the saccade were not iconic. Here is his discussion of the analog of the Sperling experiment for iconic memory (Quilty-Dunn, 2020b, p. 826):

> 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. ... 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.

But an alternative interpretation—bolstered by the masking experiment just described in which perceptual information lasts only 40 ms—suggests the opposite, that the perceptual object representations *before* the saccade were iconic and those iconic aspects do not survive the saccade very well. The upshot would be that transsaccadic memory is a form of working memory, or even long-term memory, with remnants of perception. So, it cannot be used in this way to show that perception is noniconic and conceptual.

I have not yet addressed what many take to be the strongest argument for discursive format for perceptual object representations. In the multiple object tracking experiments mentioned earlier, subjects can track a number of objects despite radical changes in properties. This fact has been taken to suggest "syntactic separation" of the element





0005464032.INDD 512 09-10-2022 11:32:13

that tracks and feature representations. For example, E. J. Green and Jake Quilty-Dunn (Green and Quilty-Dunn, 2021, p. 672) say

> object files involve explicit indexes, akin to demonstratives. There is strong reason to believe that these indexes are syntactically separate from any feature representations used to attribute features to the object … For example, indexes are plausibly maintained across changes in the feature representations held in an object file. Subjects can reliably track objects in MOT despite significant changes in colour, shape, and size during a trial …"

My response should be clear by now: what they say may be true of the object files of working memory, but that does not show this conclusion applies to the object files of perception, i.e. to perceptual object representations.

So should we just restrict the term "object file" to working memory representations and not use it to refer to perceptual object representations? I am afraid that the term "object file"—indeed, the concept of an object file as a perceptual object representation—is firmly ensconced in the perception literature, which is why in the first sentence of this article I quoted the definition of an object file from the *Encyclopedia of Perception*. The cleanest terminological revision would be to drop the term altogether.

To sum up the argument of this paper: I mentioned three important differences between the perceptual information in the so called "object files" of perception and the "object files" of working memory (and of course, also singular thought using working memory representations). These differences are capacity, fundamental computations and task-specific computations. And of course the working memory (and subsequent thought) representations are enclosed in a cognitive envelope, unlike the representations of perception. I also argued that there is reason to believe that visual object representations are iconic and that evidence to the contrary can be explained away as dependent on working memory representations that enclose remnants of perception in a cognitive envelope. The term "object file" is used to apply to kinds of representations that are fundamentally different from one another and so the term is a source of confusion. We would be better off without it.

## Note

1 Steven Gross has described evidence that visual representation of letters is abstract enough to be common to lower- and uppercase letters. His article and my reply are forthcoming in *Analysis Reviews*.

## References


Alvarez, G., and Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attentiion. *Psychological Science*, 19, 392–398.

Beschin, N., Cubelli, R., Della Sala, S., and Spinazzola, L. (1997). Left of what? The role of egocentric coordinates in neglect. *Journal of Neurology, Neurosurgery, & Psychiatry*, 63(4), 483–489.

Bisiach, E., and Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, 14(1), 129–133. http://doi.org/10.1016/S0010-9452(78)80016-1

Block, N. (2022). *The Border between Seeing and Thinking*. New York: Oxford University Press.

Bloem, I. M., and Ling, S. (2019). Normalization governs attentional modulation within human visual cortex. *Nature Communications*, 10(1), 5660. https://doi:10.1038/s41467-019-13597-1
Bloem, I. M., Watanabe, Y. L., Kibbe, M. M., and Ling, S. (2018). Visual memories bypass normalization. *Psychological Science*, 29(5), 845–856. https://doi.org/10.1177/0956797617747091
Bonato, M., Saj, A., and Vuilleumier, P. (2016). Hemispatial neglect shows that (before) is (left). *Neural Plasticity*, 1–11.
Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology*, 4(3), 121–138. https://doi.org/10.1080/17470215208416611
Caramazza, A., and Hillis, A. E. (1990). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, 346(6281), 267–269. https://doi.org/10.1038/346267a0
Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.
Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, 27(3), 183–228.
Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics (2nd ed.)*. Oxford: Oxford University Press.
Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, 358, 486–492.
Endress, A., and Potter, M. (2015). Large capacity temporary visual memory. *Journal of Experimental Psychology: General*, 214(143), 548–565.
Endress, A. D., and Siddique, A. (2016). The cost of proactive interference is constant across presentation conditions. *Acta Psychologica*, 170, 186–194. https://doi.org/10.1016/j.actpsy.2016.08.001
Flombaum, J., and Scholl, B. (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(4), 840–853.
Gordon, R., and Irwin, D. (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., and Frankl, M. L. (2008). Object continuity and the transsaccadic representation of form. *Perception and Psychophysics*, 70(4), 667–679. https://doi.org/10.3758/PP.70.4.667
Green, E. J., and Quilty-Dunn, J. (2021). What is an object file. *British Journal for the Philosophy of Science*, 72(3). https://doi.org/10.1093/bjps/axx055
Hollingworth, A., and Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136.
Irwin, D. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 307–317.
Irwin, D. E., and Andrews, R. V. (1996). *Integration and accumulation of information across saccadic eye movements Attention and performance 16: Information integration in perception and communication*. (pp. 125–155). Cambridge: The MIT Press.
Jordan, K., Clark, K., and Mitroff, S. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition*, 18(4), 492–503.
Kahneman, D., Treisman, A., and Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24, 175–219.
Klein, R., and Ivanoff, J. (2008). *Inhibition of return. Scholarpedia*, *3*(10), 3650.
Kwak, Y. and Curtis, C. E. (2022). Unveiling the abstract format of mnemonic representations. *Neuron*, 110(11), 1822–1828.
Lamme, V. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7, 12–18.
——— (2004). Separate neural definitions of visual consciousness and visual attention: A case for phenomenal awareness. *Neural Networks*, 17(5–6), 861–872.

——— (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10(11), 494–501.

——— (2016). The crack of dawn: Perceptual Functions and neural mechanisms that mark the transition from unconscious processing to conscious vision. In T. Metzinger and J. Windt (Eds.), *Open MIND, 2 Volume Set* (Vol. 1). Cambridge: MIT Press.

Lamme, V. A. F. (2018). Challenges for theories of consciousness: Seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755).

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

Marino, A., and Scholl, B. (2005). The role of closure in defining the "objects" of object-based attention. *Perception and Psychophysics*, 67(7), 1140–1149.

Murez, M., and Recanati, F. (2016). Mental files: An introduction. *Review of Philosophy and Psychology*, 7(2), 265–281. https://doi.org/10.1007/s13164-016-0314-3

Nakayama, K., He, Z. J., and Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In S. M. Kosslyn and D. Osherson (Eds.), *Visual Cognition: An Invitation to Cognitive Science, Volume 2* (Vol. 2, pp. 1–70). Cambridge: MIT Press.

Noles, N., Scholl, B., and Mirtroff, S. (2005). The persistence of object file representations. *Perception and Psychophysics*, 67(2), 324–334.

Peters, M., and Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLIFE*, 4, e09651. http://doi.org/10.7554/eLife.09651

Peters, M. A. K., Kentridge, R. W., Phillips, I., and Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, 2017(1). https://doi.org/10.1093/nc/nix015

Phillips, I., and Block, N. (2016). Unconscious seeing—a debate. In B. Nanay (Ed.), *Current Controversies in Philosophy of Perception*: Routledge: Taylor and Francis Group.

Pinto, Y., Sligte, I. G., Shapiro, K., and Lamme, V. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-013-0393-4

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., and Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, 15(8), 13. https://doi.org/10.1167/15.8.13

Pratte, M. S. (2018). Iconic memories die a sudden death. *Psychological Science*, 29(6), 877–887. https://doi.org/10.1177/0956797617747118

Pylyshyn, Z. (2007). *Multiple object tracking. Scholarpedia*, 2(10), 3326.

Quilty-Dunn, J. (2016). Seeing objects: A case study in the under-intellectualization of perception. In J. Quilty-Dunn (Ed.), *Syntax and Semantics of Perceptual Representation*. Ph.D. thesis, City University of New York.

Quilty-Dunn, J. (2019). Is iconic memory iconic? *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12625

Quilty-Dunn, J. (2020a). Concepts and predication: From perception to cognition. *Philosophical Issues*, 30(1), 273–292.

Quilty-Dunn, J. (2020b). Perceptual pluralism. *Nous*, 54(4), 807–838. https://doi.org/10.1111/nous.12285

Quilty-Dunn, J., and Green, E. J. (2021). Perceptual attribution and perceptual reference. *Philosophy and Phenomenological Research*.

Recanati, F. (2012). *Mental Files*. Oxford, UK: Oxford University Press.

Richard, A. M., Lee, H., and Vecera, S. P. (2008). Attentional spreading in object-based attention. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 842–853.

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

Scholl, B., and Flombaum, J. (2010). Object Persistence. In E. B. Goldstein (Ed.), *Encyclopedia of Perception* (pp. 653–657). Thousand Oaks: Sage.
Schurgin, M., and Flombaum, J. (2018). Visual working memory is more tolerant than visual long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 44(8), 1216–1227.
Schut, M., Fabius, J., Van der Stoep, and Van der Stigchel, S. (2017). Object files across eye movements: Previous fixations affect the latencies of corrective saccades. *Attention, Perception, and Psychophysics* 79, 138–153.
Sligte, I. G. (2011). *A new definition of visual short term memory*. Ph.D. thesis, University of Amsterdam.
Sligte, I. G., Scholte, H. S., and Lamme, V. (2008). Are there multiple visual short term memory stores? *PloS One*, 3(2), e1699.
Sligte, I. G., Scholte, H. S., and Lamme, V. (2009). V4 activity predicts the strength of visual short-term memory representations. *Journal of Neuroscience*, 29(23), 7432–7438.
Sligte, I. G., Vandenbroucke, A. R., Scholte, H. S., and Lamme, V. A. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, 1, 175. https://doi.org/10.3389/fpsyg.2010.00175
Sligte, I. G., Wokke, M. E., Tesselaar, J. P., Scholte, H. S., and Lamme, V. (2011). Magnetic stimulation of the dorsolateral prefrontal cortex dissociates fragile visual short tem memory from visual working memory. *Neuropsychologia*, 49, 1578–1588.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74(498), 1–29.
Sperling, G., van Santen, J., and Burt, P. (1985). Three theories of stroboscopic motion perception. *Spatial Vision*, 1, 47–56.
Tipper, S., Jordan, H., and Weaver, B. (1999). Scene-based and object-eentered inhibition of return: Evidence for dual orienting mechanisms. *Perception and Psychophysics*, 61(1), 50–60.
Tipper, S. P., and Behrmann, M. (1996). Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1261–1278.
Walker, R. (1995). Spatial and object-based neglect. *Neurocase*, 1(4), 371–383. https://doi.org/10.1080/13554799508402381
Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161–265.
Wood, J. N. (2009). Distinct visual working memory systems for view-dependent and view-invariant representation. *PloS One*, 4(8), e6601. https://doi.org/10.1371/journal.pone.0006601
Zhao, J., Kong, F., and Wang, Y. (2013). Attentional spreading in object-based attention: The roles of target–object integration and target presentation time. *Attention, Perception, and Psychophysics*, 75(5), 876–887. https://doi.org/10.3758/s13414-013-0445-x
Zorzi, M., Bonato, M., Treccani, B., Scalambrin, G., Marenzi, R., and Priftis, K. (2012). Neglect impairs explcit processing of the mental number line. *Frontiers in Human Neuroscience*, 6(125), 1–12.