CHAPTER TWENTY-SIX

The Perception–Cognition Border: Architecture or Format?

E. J. Green

1. Introduction

The distinction between perception and cognition figures centrally in many philosophical debates, including debates about the justificatory role of perceptual experience, the cognitive penetrability of perception, and the perceptual basis of demonstrative thought. Nevertheless, despite its vital importance in philosophical theorizing, there is no consensus about how the perception–cognition border should be characterized.

This chapter juxtaposes two approaches to the perception–cognition (P-C) border: the *architectural* approach, which I favor (Green 2020),[1] and the *format-based* approach, advocated by Ned Block in a landmark new book.[2] Architectural approaches construe perception and cognition as separate psychological systems with restricted patterns of information flow between them. On most views, the deliverances of perception are freely available for use in thought, but cognitive influences on perception are significantly constrained. Format-based approaches characterize the border via differences in format between the representations formed in perception and cognition. Most views construe perceptual representations as depictive or imagistic, and cognitive representations as prototypically discursive or language-like.

It is difficult to compare architectural and format-based approaches in the abstract, since there are myriad varieties of both. Here I focus on my own version of the architectural approach and Block's version of the format-based approach. After introducing my view, the *dimension restriction hypothesis* (DRH), I reply to Block's objections to DRH. Next, I outline Block's view and raise two problems for it: cognition-side challenges involving iconic elements in cognition, and perception-side challenges involving non-iconic elements in perception. Finally, I consider and rebut Block's arguments that object perception is fully iconic.[3]

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

2. The Dimension Restriction Hypothesis

On classic versions of the architectural approach (Fodor, 1983; Pylyshyn, 1999), perceptual processing unfolds without any direct access to information in cognition. Perception affects cognition, but cognition cannot affect perception except by altering the *inputs* to perceptual processing, such as when we make a voluntary saccade. The P-C border consists in this constraint on information flow, often called *cognitive impenetrability*, which grounds a divide between psychological systems.[4]

DRH does not adopt this strict conception of the P-C architectural divide. I side with others who hold that certain cases, such as endogenous attention (Ling andet al., 2009) and feature-based expectation (Kok et al., 2012), arguably witness genuine cognitive penetration of perception (Macpherson, 2012; Block, 2016; Wu, 2017; although see Firestone and Scholl, 2016; Gross, 2017; and Quilty-Dunn, 2020a for alternative views).

On my view, perceptual processes are coherently affected by cognition, but only in limited ways fixed by the agent's cognitive architecture.[5] Specifically, individual perceptual processes are constrained to compute over and output a restricted range of *dimensions* or *variables*.[6] Cognition may alter, in magnitude or precision, the *values* that a perceptual process computes over for some of these variables, but not the *range* of variables it can compute over. Perceptual processes are *dimension-restricted*.

More precisely, a system S exhibits dimension restriction when there is an analysis A of S into a set of processes $P_1 \dots P_n$, such that:

1 A is a functional analysis of the sort found in contemporary cognitive science. In particular, A is both *natural* and *appropriately fine-grained*.
2 Cognitive architecture constrains each of $P_1 \dots P_n$ to compute over and output values for a bounded class of dimensions.
3 The class of dimensions that any P_i within $P_1 \dots P_n$ can compute over cannot be modified voluntarily through any internal psychological process.

DRH claims that perception is dimension-restricted in this sense, while cognition is not. Cognition *cannot* be analyzed into processes each of which is architecturally constrained to compute over and output values for a bounded class of dimensions.

Regarding (1): "Naturalness" is intended to preclude *arbitrary groupings* of subprocesses. Subprocesses that contribute to a single task function should be grouped into larger processes; processes subserving largely distinct functions should not. "Appropriate fine-grainedness" requires the analysis to capture the actual computational diversity within the system. An analysis of vision that *bottomed out* in the divide between ventral and dorsal vision would be insufficiently fine-grained, since it would fail to reflect the variety of fine-grained computations performed *within* both the ventral and dorsal stream.

DRH is a theory of the distinction between perceptual and cognitive *processes*. It does not say that the execution of a perceptual process suffices for *perception*. Perception is a person-level kind. DRH does not give conditions for perception in this sense.[7] But, crudely, I believe that the person-level kind *perception* is instantiated when perceptual processes alone issue in psychological states attributable to the whole person (Green, 2020, 330–331).

Suppose that a perceptual process computes distal size from retinal-image size and distance, capitalizing on the size–distance invariance ratio (Gogel, 1971). Then cognition might, via feature-based attention or expectation, modulate the *values* computed over for retinal-image size, distance, or distal size, affecting the process's output (cp. Cañal-Bruland et al., 2011). However, such top-down influences must be filtered through the fixed range of variables to which the process has access. If these variables don't include (say) hue, then cognition cannot direct the process to compute over hue, even though other perceptual processes compute over it.

A central aim in vision science is to analyze perceptual processes as computations over fixed ranges of variables. Examples abound. Motion perception has been argued to comprise three separate processes: first-order, second-order, and third-order motion (Lu and Sperling, 2001), and the processes are individuated by the distinct sets of variables from which they derive motion. Models of visual parsing describe how vision uses a limited range of geometrical cues to divide a complex shape into significant parts (Singh and Hoffman, 2001). A contemporary model of shape perception derives the 3D shape consistent with the retinal image that maximizes a cost function determined by compactness, surface area, symmetry, and planarity (Li et al., 2009).

What's important about these models is not whether they are ultimately correct, but the general perspective they exemplify: Perceptual systems may be analyzed into processes each sensitive to delimitable classes of variables, and this analysis is relatively stable across time and contexts.

By contrast, I claim we *cannot* delimit the ranges of dimensions that cognitive processes can compute over, assuming adequate time and motivation. Theoretical models of cognitive capacities generally characterize *strategies* available for reasoning with different variables on different occasions, such as Bayes nets (Gopnik et al., 2004) or content-neutral heuristics (Todd and Gigerenzer, 2007). For instance, Lee and Cummins' (2004) evidence accumulation model describes how we choose among options evaluable along multiple criteria: We consult the criteria in order of significance, assigning scores to each option, stopping when one option's score exceeds some threshold. Importantly, no constraints are imposed on choice domains or the range of criteria consulted. The model characterizes a strategy available for choosing either restaurants or colleges and accommodates whatever criteria one deems relevant.

One might ask whether DRH is offering *metaphysically necessary* conditions on perceptual processing. I do not construe DRH in this sense. Rather, I hold that dimension restriction is a *nomologically necessary, explanatorily deep* property of human perceptual systems.[8] It helps explain various other properties of human perception—such as the impressive speed and tractable computational costs of perception versus cognition. As Brooke-Wilson (ms.) observes, the computational costs of a process are primarily driven not by the *amount of information* to which it has access, but by the *dimensionality of the problem* it solves (roughly, the range of variables it integrates in solving the problem). By limiting the dimensions perceptual processes consult, architecture limits how many hypotheses they must consider in generating their outputs.

A key desideratum on a theory of the P-C border is that it help in adjudicating difficult cases. If we are unsure whether a state is perceptual or cognitive, we would like the theory to offer some guidance. For example, the theory should optimally weigh in on the *rich–thin debate* about perceptual content (Siegel and Byrne, 2016). It should help determine whether "high-level" properties like causation or animacy are genuinely perceived or only inferred post-perceptually. A theory providing no guidance in these cases would be less useful in philosophical theorizing.

Importantly, DRH needn't place metaphysically necessary conditions on perception to help adjudicate controversial cases. If human perception is dimension-restricted by nomological necessity, then we are licensed in using evidence about whether a process is dimensionally restricted to decide whether it is perceptual. In practice, psychologists often adopt this sort of method. Those arguing that high-level properties like causation are perceived emphasize that their recovery depends on a fixed class of cue variables. Changing these variables produces systematic changes in representation of the high-level property (Scholl and Gao, 2013; Kominsky et al., 2017).

As evidence for DRH, Green (2020) cites stark limits on how cognition can affect texture segregation (Rosenholtz, 2015) and the construction of priority maps for visual search (Wolfe and Horowitz, 2017). There are also limits on the extent to which cognitive hints about the category of a target can alter perceptual similarity relations as indexed by visual search efficiency (Cohen et al., 2017). I've argued that these limits are well-explained by DRH. I've also argued that certain evidence supporting cognitive influences on perception is well-accommodated by DRH, but not by rival versions of the architectural approach.

3. Objections to DRH

In evaluating challenges to DRH, two questions must be distinguished.

First, there is the question of which dimensions a perceptual process *does* compute over on some occasion. When a process is activated, it may compute over some but not all of the dimensions that cognitive architecture permits it to compute over. When it is not activated, it computes over none of the dimensions that architecture permits it to compute over.

Second, there is the question of which dimensions a perceptual process *can* compute over, holding cognitive architecture fixed. This is the key question for DRH. DRH holds that cognitive architecture restricts the range of dimensions that any individual perceptual process can compute over. Because changes in whether a process is active at all needn't alter the range of dimensions it can compute over, DRH allows that cognition can cause changes to which perceptual processes are active at a time, thereby affecting which dimensions perception *does* represent on a particular occasion.

Now for Block's objections. Block argues that there are two readings of DRH. The first he describes as "true but not that useful for distinguishing perception and cognition," while the second he describes as "exciting and, I believe, false" (forthcoming, 305).

Here is the allegedly true but non-useful reading:

> [T]here are features that cannot be represented in perception even though they can be represented in cognition. We can perceptually represent the colors and shapes in a painting but although we can cognitively represent the fact that it was painted by Rembrandt, we cannot represent this fact perceptually. (...) In this sense of the suggestion, it amounts to the point that there are feature dimensions that are representable by cognition but not perception. In other words, some of the features of things in the world are observable and others not. (305)

Block does not explain why this sort of view would not be useful in distinguishing perception from cognition. However, one obvious drawback is that the view could not be employed to adjudicate hard cases (e.g., causation or animacy) unless we had some independent theory of which features are observable and which aren't. In any case, this view is not the correct reading of DRH. While it is consistent with DRH that there are features we can represent cognitively but not perceptually (and I accept this claim), this is not the fundamental constraint the view imposes on perception. DRH places constraints on the dimensions that *individual perceptual processes* compute over, not the dimensions that perception *as a whole* can represent. Critically, this framework captures situations in which there are dimensions that certain perceptual processes can compute over but other perceptual processes cannot.

For instance, Green (2020) argues that the processes responsible for constructing the priority map used in visual search are architecturally prohibited from computing over cross-shaped intersections (Wolfe and Dimase, 2003), although this feature is computed over by processes responsible for forming representations of transparency (Anderson, 1997). Likewise, Sousa et al. (2009) report that perceptually represented surface slant is available to guide visual search, but the variables used to derive surface slant are not themselves available.

The claim that individual perceptual processes are dimension-restricted is thus stronger and more interesting than the claim that perception as a whole is dimension-restricted. It is also easier to investigate empirically. It is easier to explore limits on the variables that individual processes can take into account than limits on the range of variables that perception as a whole can represent.

DRH also differs from Block's true, non-useful claim because it posits an *internal architectural difference* between perception and cognition—a difference in the *computational strategies* they employ. I claim that perception is wholly analyzable into subprocesses each of which is architecturally constrained to compute over a bounded range of variables, while cognition is not.

Now for the allegedly false but exciting reading: "[C]ognition cannot introduce into an individual perception a new feature that would not otherwise be represented in that perception" (Block, forthcoming, 305). Block raises challenges to this view. First, however, notice that this reading is also not equivalent to DRH. DRH says nothing about what dimensions can be introduced into an individual perceptual episode. Cognition might, consistent with DRH, affect which dimensions are represented during a perceptual episode by affecting which processes are active during that episode.

In criticizing the view he deems false and exciting, Block appeals first to cognitive effects on our perception of ambiguous figures like the "rat-man" (Figure 1). Thus:

> Consider the rat-man ambiguous drawing. ... Suppose an observer sees it as a rat and does not see it as a face. Then someone tells the observer that it can also be seen as a face, resulting in the observer knowing that it can be seen as a face, and that in turn results in the observer seeing it as a face. In this way, cognition can introduce a new feature into a perception that was not represented before the effect of the cognitive state. (305)

AQ1 **FIGURE 1** THE RAT-MAN.

Green (2020) used the distinction drawn earlier (the dimensions a process *does* compute over versus those it *can* compute over) to explain the influence of cognitive hints on perception of the rat-man. Specifically, the effect is mediated by attention to certain locations or low-level features of the stimulus (Meng and Tong, 2004; Slotnick and Yantis, 2005). Certain stimuli possess low-level features diagnostic of multiple incompatible higher-level categories. By selectively attending to features diagnostic of one of them, cognition may affect which higher-level perceptual processes dominate during a perceptual episode, ultimately changing which features are represented during that episode. All this is compatible with the individual processes being dimension-restricted. The lower-level processes are architecturally limited to computing over a restricted range of lower-level dimensions, and the higher-level processes are likewise limited to computing over a restricted range of lower-level and higher-level dimensions.

The rat-man figure contains features diagnostic of both faces (e.g., human eyes and nose) and rats (e.g., a long stringy tail). The cognitive hint that the figure can be seen as a face causes the perceiver to attend selectively to features diagnostic of face-hood. It is a "selection effect." When these features are attended, they are prioritized for further processing and are more likely to activate higher-level processes dedicated to the categories for which they are diagnostic—for instance, processes of face analysis housed in the fusiform face area (Kanwisher and Yovel, 2006; Andrews et al., 2002). Thus, perception is more likely to represent the figure as a face, or at least a face-gestalt.

Block objects to this account as follows:

> Green seems to think that if an effect is a "selection" effect, it does not impinge on the DRH. But why? Perhaps the idea is that selection would have to operate on dimensions that are already being computed over in the perception, so no new dimensions are introduced. But Green gives no evidence that when one is seeing the stimulus as a rat, there is already a face-dimension that is being computed over. (306)

But the issue is *not* whether cognition can change which dimensions are "computed over in the perception" when viewing the rat-man. I do not hold that face-hood must already be represented when the stimulus is seen as a rat, so I owe no evidence for this claim. Rather, the view is that cognition *can* affect which perceptual processes are active during an episode, thus affecting which dimensions are represented during that episode. Face-hood is *not* represented before the hint, but *is* represented afterward, because voluntary attention affects which higher-level perceptual processes are dominant. It does not, however, alter the dimensions that these processes can compute over.

Block later wields the famous Dalmatian image against DRH (cp. Nes et al., 2021, 17), but this argument has the same problem. My account of the Dalmatian mirrors the rat-man. Following a hint that the picture contains a dog, the subject preferentially attends to geometric or textural features diagnostic of animals (Long et al., 2017; Schmidt et al., 2017). We possess tacit knowledge about such diagnostic features. We use this knowledge to allocate attention to them, thereby activating higher-level perceptual processes dedicated to the analysis of animal shapes, increasing the likelihood of discriminating the Dalmatian.

This concludes my discussion of Block's objections. I turn to Block's format-based account.

4. The Full Iconicity View

Format-based approaches characterize the P-C border via differences in the formats of perceptual and cognitive representations. Block (forthcoming) argues that certain features, including iconicity, are *constitutive of* and *necessary for* perception, but not cognition. Thus:

> The constitutive iconic format and non-conceptual and non-propositional nature of perceptual representation provide necessary conditions of perceptual representation … (48)

> Perception is constitutively non-propositional, non-conceptual and iconic, and cognition does not constitutively have any of these properties. (390)

To assess these claims, we must know what Block means by "iconic," "non-conceptual," and "non-propositional." Block's claims about the explanatory relations among iconicity, nonconceptuality, and nonpropositionality also help settle an important interpretive issue with the theory.

Start with "iconic." Block endorses an *analog tracking and mirroring* (ATM) conception of iconicity, explained thus:

> Analog tracking and mirroring obtains when there is a set of environmental properties and a set of representations of those environmental properties such that:
>
> 1 Certain differences in representations function as responses to differences in environmental properties in a way that is sensitive to the degree (and also kind) of environmental differences. (…)
> 2 Certain differences in representations function to alter the situation that is represented in a way that depends on the degree (and also kind) of representational change.
> 3 Certain relations (including temporal relations) among the environmental properties are mirrored by representations that instantiate analogs of those relations. (Block, forthcoming, 146–147)

A mercury thermometer illustrates the type of system Block has in mind. The representations are mercury volumes of varying heights and the environmental properties are temperatures. Differences in mercury height function as responses to differences in temperature, and degree of change in height is proportional to degree of change in temperature. Mercury volumes *track* particular temperatures, while relative heights of the mercury volumes *mirror* relations among the temperatures.[9]

As Block acknowledges, ATM is not the only notion of iconicity. Green and Quilty-Dunn (2021) employ a different notion characterized by two principles. First, the *parts principle:* Parts of an iconic representation represent parts of what the whole representation represents (Fodor, 2007). Second, the *holism principle:* Parts of iconic representations explicitly represent properties along multiple dimensions simultaneously, and not by composing discrete representations of each of those properties (Quilty-Dunn, 2020b; Davies, 2021). The Parts+Holism conception is well-equipped to capture the *multidimensional richness* of prototypical icons like pictures, which efficiently convey abundant information along many dimensions (Dretske, 1981, 137–138), and also the low-cost binding of dimensions in high-capacity stores like iconic memory (Quilty-Dunn, 2020c).[10]

I believe that Block's equation of iconicity with analogicity collapses distinctions among representational systems that our taxonomy of formats ought to capture, but I cannot explore these matters presently. More urgently, there is a risk of dialectical confusion. Green and Quilty-Dunn argue that *some* perceptual representations—object files—are *not* iconic in the Parts+Holism sense. Block argues that *all* perceptual representations *are* iconic in the ATM sense. Because analog representations needn't satisfy the parts principle (Clarke, 2022), one might conjecture that object files are *analog*, and thus iconic by Block's lights, but *not* iconic in the Parts+Holism sense. Indeed, Green and Quilty-Dunn granted that object files might have analog constituents (2021, fn. 16).

However, our disagreement is not purely verbal. Green and Quilty-Dunn's theory of object-file format remains incompatible with the view that object files are fully iconic when "iconicity" is understood in the ATM sense. For, while we allow that *some* constituents of object files might be analog, we deny that *all* their constituents are analog (see Section 5.2).

Having flagged these issues, for convenience I'll use "iconic" in Block's sense in what follows.

Now for "nonconceptual" and "nonpropositional." While some apply these labels to *contents* (Peacocke, 1992), Block applies them to *states* or *representations* (cp. Byrne, 2005). For Block, a representation counts as (non)conceptual or (non)propositional thanks to its functional role. Conceptual representations are composed of concepts, which Block characterizes as "representational … elements that constitutively function in propositional thought, reasoning, problem-solving, (…) and other cognitive processes" (Block, forthcoming, 140). Propositional representations are characterized by their "role in content-based transitions in cognitive processes such as reasoning, inferring, thinking and deciding" (144). Presumably, nonpropositional representations are those that do not function to participate in such processes.

Block argues that perceptual representations are nonconceptual and nonpropositional on the grounds that they lack complex logical structure, and thus (says Block) cannot participate in deductive inferences like disjunctive syllogism. For example, perception cannot represent contents like *x is red or green*. This logical noncomplexity is allegedly explained by iconic format:

> The fact that perception cannot be conjunctive, disjunctive, conditional, negative or universal is tied to the iconic format of perception. To have productive disjunction, conjunction, etc., there must be format elements that play the role of logical constants. (…)

> [A]nalog mirroring involves … a correspondence between degrees of difference in representational parameters and what is represented. It is hard to see how a representation that is iconic by that standard could be a logical constant. (Block, forthcoming, 164)

To represent logical complexity, Block thinks a representation would need constituents expressing logical constants. But it's unclear how logical constants could be encoded iconically (cp. Rescorla, 2009). This aspect of Block's account is crucial, since it purports to explain *why* iconicity, nonconceptuality, and nonpropositionality cluster together: Iconicity partially explains the other two.

Now the interpretive issue. Block's claim that "constitutive iconic format [provides a] necessary condition of perceptual representation" (48) permits two interpretations. First, that necessarily, every perceptual representation possesses *at least some* iconic constituents. Second, that necessarily, every perceptual representation consists *wholly* of iconic constituents. Call the first *Partial Iconicity*, and the second *Full Iconicity*. Block does not say which he intends, but his view of the explanatory role of iconicity reveals that he must embrace Full Iconicity. For, if perceptual representations were constitutively iconic only in the Partial Iconicity sense, their iconicity could not be used to explain why they lack logical complexity, and so couldn't explain their inaptness for deductive inference.

It is easy to develop hybrid systems combining analog constituents and logical constants. Johnson (2015) gives an example. Suppose we need a system for representing spatial relations among 2D shapes—such as the scenario in Figure 2:

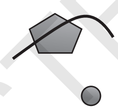


FIGURE 2 *SOURCE:* JOHNSON (2015).

We *might* produce a full depiction of the scenario that reproduces all of its spatial features. However, another option is to represent the shapes iconically while representing spatial relations symbolically, combining representations of pairs of objects in relations by conjunction, as in Figure 3:

FIGURE 3 *SOURCE:* JOHNSON (2015).

Here, "R," "S," and "T" stand arbitrarily for particular spatial relations without analog mirroring, while the shapes entering these relations are represented iconically. The system combines iconic constituents with logical connectives. We could further enrich the system with connectives for disjunction, if–then, and so on.

Crucially, if Block held only that all perceptual representations possess *some* iconic constituents, he could not rule out that they explicitly represent logical structure. Partial Iconicity is compatible with formats like that in Figure 3. To explain perception's inaptness for logical inference, Block must embrace Full Iconicity.

5. Challenges to Block's Format-Based View

I contend that the iconic/noniconic distinction is simply a distinction between broad, internally diverse categories of mental representation. It is an explanatorily significant distinction, but it does not align with the P-C border. To motivate this perspective, I raise two sorts of challenge to Block's view. *Cognition-side* challenges emphasize the role of iconic elements in cognition. *Perception-side* challenges emphasize the role of noniconic elements in perception.

5.1. Iconic Representations in Cognition

Suppose Block is correct that perception is fully iconic—an assumption I'll question below. He grants that fully iconic representations are *also* used in cognition: namely, in imagistic cognition (forthcoming, 21–22, 187–192), mental arithmetic (forthcoming, 185–186; Walsh, 2003), and map-based spatial navigation (forthcoming, 103, 141).

This raises a problem. For Block, imagistic cognition, mental arithmetic, and spatial navigation are cognitive processes, but they utilize iconic, nonconceptual representations. Indeed, there seems nothing to prevent some of these processes (or at least instances of them) from being *fully* iconic and nonconceptual. How, then, can Block use iconicity to characterize the P-C border?[11]

Block's primary reply is that while perception is *constitutively* iconic and nonconceptual, cognition isn't. Cognitive processes may *happen* to be fully iconic and nonconceptual, but such features are not *necessary* for cognition (forthcoming, 21–22, 48, 178). This position faces two challenges.

First, Block owes an explanation of *why* fully iconic episodes of imagistic cognition or mental arithmetic should be classified as cognitive rather than perceptual, given that they possess the very features allegedly constitutive of perception.[12] Perhaps the classification is supported by mere intuition, or perhaps by some further objective property that perceptual processes possess but analog mental arithmetic (say) does not. The first option fits poorly with Block's stated aim of characterizing the P-C border scientifically rather than from the armchair (forthcoming, 41–43). And if Block chooses the second option, he owes an account of what the further property is. Moreover, he must explain why this property doesn't suffice *on its own* to demarcate perception from cognition, rendering iconicity superfluous.

Second, Block's concession that the constitutive features of perception are shared with many cognitive processes undercuts his account's capacity to meet the desideratum mentioned earlier: adjudicating hard cases. We want a theory of the P-C border to help resolve disputes such as whether causation is perceived. But on Block's view, even if we learned that causation is represented iconically, this wouldn't settle the issue. For it wouldn't settle whether causation is perceived or instead recovered through imagistic cognition. Indeed, it's not clear whether learning that some mental state is iconic should even *raise our confidence* that it is perceptual.

Block (2014, forthcoming) introduces several "signature markers" of perception, such as adaptation and efficient visual search. He might reply that we can use such markers to settle difficult cases like causation (Rolfs et al., 2013). I agree that some of Block's markers are useful here (Green, 2017). However, the markers are independent of the format-based view. Proponents of the architectural approach can agree that such features increase the probability that a state is perceptual. Indeed, they are employed by scientists who remain dubious of the format-based view (Hafri and Firestone, 2021). I suggest that iconicity does not confer any additional adjudicatory power beyond the signature markers.

5.2. Noniconic Representations in Perception

I now argue that Full Iconicity is incorrect. Not every perceptual representation is composed entirely of iconic constituents. My case study will be the role of *singular constituents* in object perception.

Object files are representations that subserve our ability to perceptually track objects through motion and change. Experimental evidence for object files derives from several paradigms:

Object-reviewing: Subjects see a pair of objects, and *preview* features (e.g., letters or shapes) briefly appear within the square wireframes before vanishing. Next, the wireframes shift locations. Finally, a *test* feature appears in one of them, and the subject reports whether it matches either of the preview features. Responses are faster when there is a match, but faster still when a feature reappears within the same wireframe as before (Kahneman et al., 1992; Noles et al., 2005). This is an *object-specific preview benefit* (OSPB). According to object-file theory, when the preview feature appears, a representation of the feature is entered into the relevant file. This representation remains in the file after the feature has vanished. When a new feature appears within an object, responses to it are speeded if it matches information already stored in the relevant file. OSPBs are used to index object-file maintenance: an OSPB indicates that a single file was maintained throughout the interval between preview feature and test feature.

Multiple-object tracking (MOT): Subjects see a display of randomly moving objects and need to track a subset of them. While there is no fixed limit on how many objects can be tracked (Alvarez and Franconeri, 2007), in typical conditions subjects track around 4 objects with high accuracy, after which performance deteriorates (Pylyshyn and Storm, 1988). Most feature changes make little difference to tracking accuracy (vanMarle and Scholl, 2003; Zhou et al., 2010), and it is easier to track objects in MOT than to also retain information about their color, shape, or category (Cohen et al. 2011). Finally, OSPBs are enhanced for tracked objects in MOT, suggesting that a common system of object representation is involved in both tasks (Haladjian and Pylyshyn, 2008).

There is also evidence that object files are deployed in apparent-motion perception (Odic et al., 2012; Stepper et al., 2019), transsaccadic memory (Schut et al., 2017), and visual working-memory (VWM) tasks (Hollingworth and Rasmussen, 2010). They may also account for certain serial dependence effects, where perception of a feature is biased toward features seen recently (Collins, 2021).

How does object-file theory explain the tracking of objects through change? Green and Quilty-Dunn (2021) propose that object files comprise two components: a *singular constituent* that sustains reference to an object over time without encoding any of its features (Pylyshyn, 2007), and a *feature store* where representations of current and past features of the object (tagged accordingly) are retained. When tracking an object through change, a token singular constituent remains assigned to it while representations in the feature store are replaced. The *maintenance* of a file is determined by the singular constituent: An object file is maintained for an object throughout an interval just in case the object is denoted by the same singular constituent throughout that interval.

Singular constituents are key to explaining how object files resolve the *correspondence problem*. Suppose that multiple objects are visually differentiated at time T1 and time T2. The correspondence problem is that of determining, for each object at T1, which object (if any) at T2 is a continuation of it (Dawson, 1991). In resolving this problem, plausibly the visual system does not explicitly represent facts about identity or persistence. Instead, I suggest, it solves the problem *implicitly*, through maintenance of singular constituents. Object O2 is treated as a continuation of object O1 when the same singular constituent denotes both of them. Coreference between the perceptual representations of O1 and O2 is coreference *de jure*, in Recanati's (2020) sense. Singular constituents explain our capacity to apprehend object continuity without overintellectualizing this capacity.

Singular constituents do not possess iconic format under ATM. First, ATM characterizes representations of *properties*. Singular constituents represent particulars only, without encoding any properties, so do not fall within its scope. Second, singular constituents are not ordered by a similarity relation on their syntactic properties that mirrors a similarity relation on the objects they denote. For this to obtain, there would need to be a systematic correspondence between the syntactic properties of singular constituents and certain properties of the denoted objects. But there cannot be. Like demonstratives, singular constituents of the same syntactic type are used to pick out arbitrarily different objects on different occasions (e.g., a red square, then a blue circle), preventing such systematic correspondence. Thus, singular constituents are not iconic according to ATM. If perceptual object representations possess singular constituents, then they possess some noniconic constituents, and Full Iconicity is false.

Block grants that singular constituents would not be iconic but claims we don't need them to explain perceptual reference to particulars or the perception of object continuity during tracking. Instead, these capacities are supposedly explained by the *functional role* of fully iconic object representations (Block, forthcoming, 181). How does functional role do this explanatory work? To explain tracking, Block emphasizes the continued representation of "Spelke-object" properties like cohesion and spatiotemporal continuity:

> Does tracking behind occluders show that the perceptual representations must have a singular format element and so are not iconic? A dynamic conception of iconic representations would embody constraints on iconic object representations like those on "Spelke objects": (1) cohesion, (2) contact and (3) continuity. … These constraints were first formulated by Elizabeth Spelke on the basis of experiments on infants, but the same constraints apply to multiple object tracking in adults, suggesting that they are built into the visual system. When objects fail these constraints, multiple object tracking fails. … On a dynamic conception of iconic representation, an iconic representation can persist through changes that do not destroy its dynamic integrity. (forthcoming, 180).

It's unclear what Block means by the "dynamic integrity" of an iconic representation. Nonetheless, the idea seems to be that an object is tracked (and treated as continuing) as long as its iconic representation "persists"—that is, is maintained by the perceptual system—and that an iconic representation persists just in case it continues to represent

Spelke-object properties like cohesion and spatiotemporal continuity. The functional role of the representation disposes it to be discarded when the object no longer possesses these properties, but maintained otherwise.[13]

This proposal is empirically inadequate. Perceptual tracking does *not* systematically fail whenever objects lose Spelke-object properties (Green, 2018, 2019). Consider cohesion, which Block characterizes as the principle that "objects maintain their connectedness and their boundaries as they move" (forthcoming, 180; cp. Spelke, 1990; Carey, 2009). vanMarle and Scholl (2003) found that when closed polygons presented at the beginning of an MOT trial fragmented into pieces that moved as a tight cluster, tracking performance was unchanged from the standard MOT condition (see Wynn et al., 2002 for convergent findings with infants). Subjects could track as many noncohesive groups as cohesive objects.[14] Studies also show that tracking improves when targets exhibit common motion or color, suggesting that gestalt principles like common motion can be used to form individual noncohesive units for tracking (Erlikhman et al., 2013).

Contra Block, perceptual object representations are not systematically discarded when objects visibly lose cohesion. Neither are they systematically *maintained* whenever objects visibly *retain* cohesion while traversing continuous paths. Moore et al. (2010) ran a version of the object-reviewing paradigm in which objects abruptly changed color as they moved and found that the OSPB was eliminated under these conditions. That is, there was *no* object-specific priming effect when objects suddenly changed color between the appearance of the preview feature and the test feature (cp. Jiang, 2020). Sudden feature changes that preserve Spelke-object properties can disrupt perceived object continuity, as indexed by the OSPB.

Studies of the *flash-lag illusion* also suggest that perceived object continuity can be disrupted without cohesion/continuity violations. In this illusion, a disc proceeds along a circular path. When it reaches a certain location, a flash appears next to it. The disc is typically seen as "ahead" of the flash although the two were spatially aligned. Plausibly, this is because the spatial relation between the disc and the flash is determined using an updated record of the disc's location—not its actual location during the flash. Crucially, Moore and Enns (2004) found that when the disc abruptly shrunk alongside the flash before reverting to its original size, the flash-lag illusion was eliminated. They explain this in terms of a disruption of perceived object continuity. Despite spatiotemporal continuity and cohesion,[15] the size change is taken to mark the onset of a new object, which is taken to vanish when the disc reverts to its original size. Because the fleeting "shrunken disc" does not have its represented location updated, a record of its actual location during the flash is preserved, eliminating the illusion.

In the foregoing experiments, the visual system continues to represent certain feature bundles over time (color, shape, etc.), and the bundles are visibly spatiotemporally continuous. Nonetheless, they are not taken to correspond to a single persisting object. Why not? I suggest it is because representations of the features fail to be concatenated with a single token singular constituent throughout the episode. Block's account provides no explanation of these findings.

There is another problem afflicting any view aiming to explain perceived object continuity simply by appeal to continuity in the iconic representation of certain properties or property clusters. The evidence indicates that the perception of object continuity is not linked in any direct way to continuity in the representation of any given property cluster. Rather, it is determined flexibly by different collections of properties in different contexts.

For example, while color is *typically* irrelevant to tracking in MOT, Papenmeier et al. (2014) showed that this isn't true when the perceptual representation of spatiotemporal properties is unreliable. Subjects performed a typical MOT task. During tracking, brief disruptions occurred wherein the display was suddenly rotated or zoomed. When targets had unique colors that remained stable through the disruption, tracking was superior to a homogenous condition, where all objects were featurally identical. Performance was worst in a "swap" condition where targets traded colors with distractors during disruptions. Importantly, color-swapping did not measurably affect tracking when swaps didn't coincide with a spatiotemporal disruption. When spatiotemporal information is unreliable, the visual system turns to color in determining object continuity, although color is essentially ignored in other contexts.

These findings, alongside others discussed by Quilty-Dunn and Green (2021), suggest that perception of object continuity is a distinct capacity that cannot be reduced to continuity in the perceptual representation of any given property cluster. There is no cluster of properties such that the perception of object continuity invariably depends on continuity in the representation of that cluster. This raises serious difficulties for any view like Block's which purports to explain perceived object continuity solely via continuity in the representation (iconic or otherwise) of some particular class of properties.

Thus, there is evidence that object files possess singular constituents and so are not fully iconic. Block's iconic account of perceptual tracking is empirically inadequate, and no successor to the view looks promising. While I agree with Block that perceptual representations denoting the same object at different times share aspects of their functional role, we need an account of what *underwrites* this functional-role commonality. Singular constituents provide an account. Block has offered no viable alternative.

Before moving on, I briefly address another issue that deserves longer discussion. Block (forthcoming, 207–217; this volume) contends that object-reviewing and VWM studies probe a distinct system of object representations from MOT and apparent motion, and that the latter representations are perceptual while the former are confined to working memory. He concludes that the term "object file" is ambiguous and should be dispensed with.

First, the case against Full Iconicity just developed didn't depend essentially on object-reviewing or VWM results, so even if Block's multisystem view were correct, it would not undercut these arguments. Block (this volume) suggests that the case for "syntactic separation" between singular constituents and feature representations applies only to representations in working memory, not perceptual representations. But this is incorrect. The evidence for syntactic separation derives from patterns in the visual resolution of the correspondence problem. Block himself grants that the correspondence problem is solved in perception, so he cannot disregard this evidence. Whether there is also a distinct system object representation employed in VWM is irrelevant.

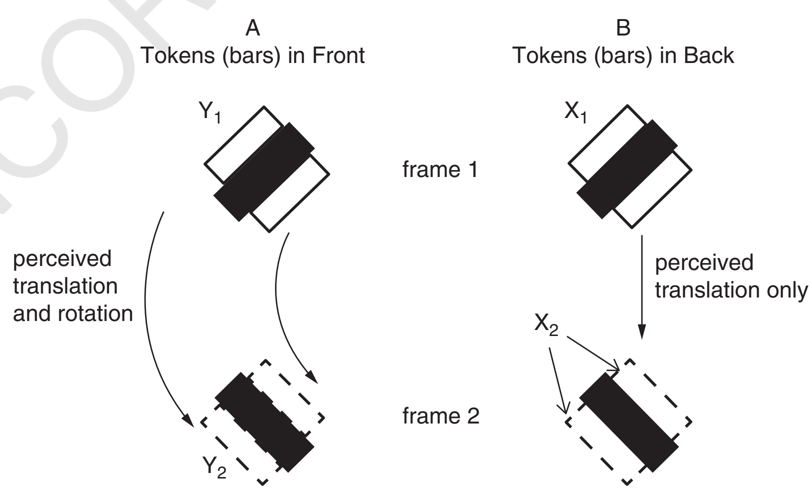
Furthermore, Block's arguments for multiple object-file systems can be resisted. He observes that, in object-reviewing studies, we lack conscious experience of preview features after they have vanished, and infers that representations of preview features are nonperceptual during the retention period between preview and test. Two replies are available. First, one might contend that preview features are encoded by *unconscious* perceptual representations during retention. Second, even if representations of preview features fail to play a perceptual role during retention, this does not show that the object representations of which they are constituents are nonperceptual. Object files are complex representations possessing certain constituents representing current features and others representing past features. Irrespective of whether past-feature constituents play a perceptual role, the file *itself* is perceptual, since it underlies the perception of object continuity throughout the retention interval. One shouldn't assume that whenever a complex representation plays a perceptual role, all its constituents must also play a perceptual role. Finally, there is both behavioral and physiological evidence that common object representations are employed in MOT and VWM (e.g., Fougnie and Marois, 2009; Drew et al., 2011).[16]

6. Block on Iconic Object Representations

I turn to Block's positive arguments for iconic object representations. Block adduces evidence that perceptual object representations are integrated with representations of spatial properties, which he takes to be uncontroversially iconic. He contends that such integration indicates that object representations are iconic too.

First, recall that the key question is not whether *there are* iconic object representations, but whether *all* perceptual object representations are iconic. Even if Block's evidence established some iconic object representations, it wouldn't establish Full Iconicity. Moreover, Full Iconicity requires not just that all perceptual representations have *some* iconic constituents, but that they have *only* iconic constituents. Establishing certain iconic constituents is not enough.[17]

Now for the evidence. Block appeals to a study described by Nakayama et al. (1995, 37–38). Subjects saw the pair of frames depicted in either Figure 4A or Figure 4B. In the former case, binocular disparity cues indicated that the white bars were in front of the black rectangle, so the bars were seen as disconnected. In Figure 4B, disparity cues



FIGURES 4A (LEFT) AND 4B (RIGHT) *SOURCE:* NAKAYAMA ET AL. (1995).

specified the white bars as behind the rectangle, so they completed amodally, forming a diamond-shaped figure. When subjects saw the frames in Figure 4A in succession, the white bars appeared both to move downward and rotate during their movement. By contrast, in Figure 4B, subjects saw nonrotational movement of a white diamond.

How do these results bear on the format of object representations? The key case is Figure 4A. Block writes:

> 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 intermediate stages of rotation and translation (i.e., vertical movement). The apparent motion results are direct evidence for the iconicity of object perception because they exhibit the smooth variation indicative of analog mirroring. (201)

However, the fact that the bars were represented as rotating through intermediate stages does *not* constitute direct evidence that they were represented iconically. Suppose that their orientations were encoded digitally via symbols bearing no nonarbitrary relation to orientation. We would *still* predict that the bars would be represented as rotating through intermediate stages, simply given the computational task of mechanisms of apparent-motion perception. These mechanisms must determine the most likely trajectory of the bars between frames 1 and 2. Since smooth rotation is the most likely trajectory (objects don't rotate discontinuously), we would *expect* a well-designed digital system to represent the bars as rotating through intermediate stages. The result would have been predicted on any model of apparent-motion perception, regardless of format.

Block's case differs from the superficially analogous case of mental rotation (Shepard and Metzler, 1971). In mental-rotation experiments, subjects take longer to match objects across orientation differences with greater angular separation between them. This result is taken to suggest that subjects rotate depictions of the objects through intermediate orientations (Block, 1983; Quilty-Dunn, 2020c). What makes this argument *prima facie* compelling is that there is no obvious reason to *expect* a discursive system to represent smooth rotation when comparing objects in mental-rotation experiments, since this is irrelevant to the shape-matching task. Precisely the opposite is true in Block's case. A discursive system of apparent-motion perception would *need* to represent the white bars as smoothly rotating, since this is their most likely trajectory, and its task is to represent their most likely trajectory.

Moreover, even if the results did mandate analog representations of orientation, a hybrid model could readily accommodate them. Perhaps object representations encode orientation in analog form, and the representation of smooth rotation involves gradual alterations of these analog constituents. This would be consistent with the view that object representations also possess non-analog singular constituents, and even with Green and Quilty-Dunn's (2021) model of object files.

Block also cites the role of object representations in guiding attention as evidence for iconicity. He argues that to guide attention, object representations must be integrated with spatial representations. This allegedly shows that objects and space are represented in a common iconic format: "[O]bject representations in perception are integrated with other representations in perception, notably spatial representations, arguing against the ... view that perceptual object representations have different formats from other perceptual representations" (201).

Block appeals to several cases of object-based attention. Drawing sweeping conclusions from such findings is risky, since, as Hollingworth et al. (2012) note, "[O]bject-based attention is likely to be composed of multiple distinct mechanisms of selection" (136). The visual system might, for instance, form both iconic and noniconic representations of objects, both of which are available for guiding attention.

In any case, object-based attention does not convincingly support iconic object representations. Instead, the evidence suggests that object representations are just one among many information sources, both perceptual and cognitive, recruited to guide attention. Given the diversity of this array, there is unlikely to be any format-based constraint on attention guidance. The fact that object representations are integrated with spatial representations in the manner needed to guide attention is *evidentially neutral* regarding their format.

Here are two examples Block discusses:

Same-object advantage (SOA): Participants see two rectangles. A brief cue appears at one end of one of the rectangles. After a brief delay, a target appears, and subjects must identify it. Responses are fastest on "valid" trials where the target appears at the cued location. More interestingly, responses are faster when the target appears at invalidly cued locations within the *same object* as the cue than when it appears in the other object, even when cue–target distance is held constant (Egly et al., 1994).

Object-based inhibition of return (IOR): When subjects are cued to attend to a location, they are initially faster to detect a target at that location (within about 150 ms), but subsequently slower to detect the target, suggesting a bias against revisiting recently attended locations. IOR can be object-based. If subjects are cued to attend to an object, then (after a delay) they are slower to detect a target appearing within that object even if it moves in the interim (Tipper et al., 1994).

Thus, object representations are sufficiently integrated with spatial representations to guide attention with respect to location. However, such integration does *not* show that they are iconic.

Sarah Shomstein and colleagues have gathered impressive evidence that attention guidance is a flexible operation sensitive not just to objects but also to high-level goals and expectations. Importantly, object-based guidance, as indexed by either SOA or IOR, can be abandoned in favor of other forms of guidance depending on the perceiver's degree of uncertainty or expectations of reward (Shomstein and Yantis, 2002; Drummond and Shomstein, 2010, 2013; Nah and Shomstein, 2020).

Shomstein and Johnson (2013) ran a version of the two-rectangle paradigm described above. In a no-reward condition, they found the normal SOA: faster identification of invalidly cued targets in the same object. However, when subjects received higher reward for identifying targets in the noncued object, *space*-based effects remained (targets were identified faster at the cued location), but the object-based effect *reversed*. Subjects were faster to detect targets in the other object than in the same object. Thus, attention was guided not only by perceptual object representations but by beliefs about reward contingencies. Similar effects can also be induced by manipulating subjects' degree of uncertainty about target location—if the cue is predictive of the target appearing in another object, then attention can be biased toward that object over the same-object location (Drummond and Shomstein, 2010; Nah and Shomstein, 2020).

According to the attentional prioritization model, object-based attention involves the combined operation of two mechanisms (Shomstein, 2012). The first is space-based and inflexible. When a location is cued, this automatically initiates a gradient of attention assigning higher priority to locations near the cue than those further away. The second mechanism is assigned flexibly given the subject's task and background knowledge. Object-based guidance reflects a *default* setting of this second mechanism. When other information isn't available, the mechanism opts to prioritize locations within a cued object (Shomstein, 2012, 165). However, object representations are but one source of information available to this mechanism, which also consults cognitive expectations and goals.

The attentional prioritization model is not universally accepted (Hollingworth et al., 2012). Nevertheless, there is compelling evidence that attentional guidance depends on the perceiver's expectations and goals. This datum suggests that mere integration with iconic spatial representations *cannot* suffice for iconicity.

Assume that, at some stage, the visual system represents spatial regions iconically, and this representation is directly involved in attention allocation. If we accept that any representation used to attentionally prioritize certain locations over others must also be iconic, then we would have to conclude that the subject's expectations about reward contingencies are iconic as well. But this conclusion is plainly unwarranted. From the Shomstein and Johnson study, all we should conclude is that the subject cognitively represents contingencies between target locations and rewards, and these representations causally interact with the spatial representations mediating attention allocation.

The object-based attention case is analogous. Mechanisms of attention guidance must have access to object representations, and those representations must specify, in some format or other, the region an object occupies. Moreover, there must be coordination between object representations and lower-level spatial representations. (Object representations may even induce "grouped arrays" within spatial representations; Hollingworth et al., 2012, 147.) But just as we are not warranted in concluding that beliefs about reward contingencies are iconic, we shouldn't conclude this about object representations. The fact that attention guidance recruits a diverse array of information sources suggests that there is no format-based constraint on attention guidance.

Block claims that *graded* effects in object-based attention offer further evidence for iconicity. First, the SOA is stronger when objects are topologically closed than when they are open (Marino and Scholl, 2005). Block writes: "If there was a radical format difference between object-perception and other perception, one would not expect such gradual effects" (203). Second, he observes that both the SOA and object-based IOR are graded phenomena. The effects are strongest near the cued location on an object and weaker further away (Hollingworth et al., 2012).

However, these effects are explainable on the attentional prioritization model without commitment on the format of object representations. The first effect suggests that mechanisms of attention guidance have access not just to information about object location but also object topology. The mechanism more heavily prioritizes locations within a cued object when it is closed. Still, there is no format-based constraint on attention guidance, so we shouldn't conclude anything about the format in which topology is represented. The second effect may be explained by concurrent operation of the first mechanism of prioritization, which is purely space-based. Neither explanation requires commitment to iconic object representations.

Block's evidence fails to demonstrate that perceptual object representations even have iconic *elements*. In any case, recall that Block needs to establish *Full* Iconicity. It is

0005464031.INDD 486 09-02-2022 16:04:54

not enough to show that object representations have iconic elements. Block's arguments that they must encode spatial features iconically in order to guide attention would only support Partial Iconicity, not Full Iconicity. I contend that Block has failed to establish even the weaker thesis.

7. Conclusion

This chapter has juxtaposed two approaches to the perception–cognition border: *architectural* approaches, which appeal to architecturally-based constraints on information flow between perception and cognition, and *format-based* approaches, which appeal to format differences between perceptual and cognitive representations. I sketched my preferred version of the architectural approach and answered objections to it. I then considered Ned Block's version of the format-based approach and argued that it faces serious difficulties.[18]

Notes

1 See also Fodor (1983), Pylyshyn (1999), Firestone and Scholl (2016), Mandelbaum (2018), and Quilty-Dunn (2020a, 2020b).

2 See also Dretske (1981), Carey (2009), Burge (2010, 2014), Tye (2020), and Clarke (2022).

3 For reasons of space, I cannot consider other views of the border here—e.g., stimulus-dependence (Beck, 2018; Phillips, 2019), phenomenological accounts (Montague, 2019), or eliminativist accounts which reject the P-C border (Clark, 2013; Lupyan, 2015). See Nes et al. (2021) for discussion of these options.

4 I am shirking certain subtleties. There are various versions of the impenetrability thesis (Gross, 2017; Stokes, 2021, ch. 4), and they differ on what constitutes cognitive penetration. Furthermore, some hold that perception is cognitively penetrable but nonetheless *encapsulated* in Fodor's (1983) sense (Quilty-Dunn, 2020a; Clarke, 2021a).

5 I construe perceptual processes as repeatable phenomena that can be re-activated on multiple occasions. By the *cognitive architecture* of a system, I mean, roughly, those aspects of its organization that remain stable through changes in the specific information the system processes, including its decomposition into basic operations and information stores (Pylyshyn, 1984). One complication is that certain architectural constraints are determined by format. In practice, architectural approaches tend to emphasize format-independent architectural constraints on information flow.

6 See Burnston and Cohen (2015) for a related approach to individuating psychological modules.

7 DRH also does not offer empirically sufficient conditions for *being a perceptual process*. It is intended only to determine whether a process is perceptual or cognitive, provided that it is one or the other (Beck, 2018). Certain motor processes may also be dimension-restricted (Mylopoulos, 2021). One might also seek further conditions to distinguish perception from nonperceptual sensory states (Burge, 2010; cf. Green, forthcoming).

8 Thanks to Adam Pautz for prompting this clarification. Block (forthcoming, 23–27) entertains the idea that iconicity is an explanatorily deep rather than essential feature of perception but endorses the essentialist view. This commits him to the dubious position that, necessarily, a creature that processed sensory signals entirely via discursive representations would not be perceiving.

9 I set aside some quibbles with ATM. For instance, it is unclear how the account captures perceptual representations of categorical, nondegreed properties—e.g., binary relations like inside/outside or attached/unattached (Hafri and Firestone, 2021).

10 Block acknowledges that perceptual icons typically obey the parts principle but argues that holism is "defective" (forthcoming, 192–198). I reject these arguments but cannot address them here.

11 See also Clarke (2021b, 23–24). One might suggest that Block *should* treat these cases as perceptual rather than cognitive—an option he sometimes floats (forthcoming, 48). However, while there is perhaps a broad notion of "perception" which includes visual imagery (Beck, 2018), this position seems less plausible for mental arithmetic.

12 When considering these cases, Block occasionally speculates that the divide he is characterizing is not really the perception/cognition divide, but rather a divide between perception and "propositional and conceptual cognition" (forthcoming, 244). If so, Block's border isn't the *perception–cognition* border after all—only a border between perception and *certain kinds* of cognition.

13 Relatedly, Hill (2021) suggests that object files are bundles of feature representations each of which is "tagged" by the time and place at which the feature was instantiated. He proposes that successive feature bundles are treated as a single object when their tags "represent successive positions on a spatio-temporally continuous path" (1406). However, as I'll explain below, spatiotemporal continuity is not sufficient for perception of object continuity.

14 While certain cohesion violations impair tracking (vanMarle and Scholl, 2003, experiment 1), these impairments are likely due to the dynamic expansion and contraction of these objects, not their noncohesion (Howe et al., 2013).

15 I should note that motion *on the screen* was not strictly continuous. The disc moved in an apparent motion sequence of 15°-step increments. However, such 15°-steps can't suffice to violate the visual system's spatiotemporal continuity constraint, since discs undergoing the same sequence were perceived as persisting when size changes were absent.

16 Block adduces some further evidence. First, there is arguably a high-capacity system of object representation probed in studies of iconic and fragile short-term memory (Landman et al., 2003). Block claims that this high-capacity system is perceptual, while the more capacity-limited system probed in VWM or object-reviewing studies is not. My reply is that perception *itself* plausibly involves multiple stages of object representation, some more capacity-limited than others. The fact that there is a system of object representation with higher capacity than object files doesn't show that object files are nonperceptual (cp. Rensink, 2000). Block also cites evidence that *center-surround suppression* is absent when stimuli are shown sequentially rather than simultaneously (Bloem et al., 2018; although see Kiyonaga and Egner, 2016). He concludes that the computation thought to support center-surround suppression—divisive normalization—is operative in perception but not VWM. However, the view that object files function in both perception and VWM does not entail that they participate in precisely the same computations in both cases. After all, the computational profile of VWM *must* differ from online perception—otherwise the system would be unable to tell whether a represented object is seen or merely remembered. There is strong evidence that perceptual representations are regularly redeployed in VWM, though their computational profile differs in the latter case (Adam et al., 2021).

17 Block grants that one might explain his evidence on a hybrid model incorporating iconic and noniconic elements, but counters: "to the extent that spatial and spatio-temporal effects saturate object representations, that view is less attractive" (forthcoming, 199). However, a complete account of object representation must accommodate *all* the evidence, not just spatial effects. If some evidence decisively supported iconic representation and other evidence decisively supported noniconic representation, a hybrid model could be the best option.

18 Thanks to Tyler Brooke-Wilson, Alex Byrne, and Selina Guter for detailed comments. Thanks also to Jake Quilty-Dunn for valuable discussion of these issues.

References

Adam, K. C. S., Rademaker, R. L., and Serences, J. T. (2021). Evidence for, and challenges to, sensory recruitment models of visual working memory. In T. F. Brady and W. A. Bainbridge (eds.), *Visual Memory*, 5–26. New York: Routledge.

Alvarez, G. A., and Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 7(13), 14, 1–10.

Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, 26(4), 419–453.

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., and Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase–face illusion. *NeuroImage*, 17, 890–901.

Beck, J. (2018). Marking the perception–cognition boundary: The criterion of stimulus-dependence. *Australasian Journal of Philosophy*, 96(2), 319–334.

Block, N. (1983). Mental pictures and cognitive science. *The Philosophical Review*, 92(4), 499–541.

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

Block, N. (2016). Tweaking the concepts of perception and cognition. *Behavioral and Brain Sciences*, 39, 21–22.

Block, N. (forthcoming). *The Border Between Seeing and Thinking*. Oxford: Oxford University Press.

Bloem, I. M., Watanabe, Y. L., Kibbe, M. M., and Ling, S. (2018). Visual memories bypass normalization. *Psychological Science*, 29(5), 845–856.

Brooke-Wilson, T. (ms.). How is perception tractable?

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

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

Burnston, D. C., and Cohen, J. (2015). Perceptual integration, modularity, and cognitive penetration. In J. Zeimbekis and A. Raftopoulos (eds.), *The Cognitive Penetrability of Perception: New Philosophical Perspectives*, 123–143. Oxford: Oxford University Press.

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

Cañal-Bruland, R., Zhu, F. F., van der Kamp, J., and Masters, R. S. (2011). Target-directed visual attention is a prerequisite for action-specific perception. *Acta Psychologica*, 136(3), 285–289.

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

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

Clarke, S. (2021a). Cognitive penetration and informational encapsulation: Have we been failing the module? *Philosophical Studies*,178(8), 2599–2620.

——— (2021b). Mapping the visual icon. *The Philosophical Quarterly.* https://doi.org/10.1093/pq/pqab054.

——— (2022). Beyond the icon: Core cognition and the bounds of perception. *Mind and Language*, 37(1), 94–113.

Cohen, M. A., Alvarez, G. A., Nakayama, K., and Konkle, T. (2017). Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of Neurophysiology*, 117(1), 388–402.

Cohen, M. A., Pinto, Y., Howe, P. D., and Horowitz, T. S. (2011). The what–where trade-off in multiple-identity tracking. *Attention, Perception, and Psychophysics*, 73(5), 1422–1434.

Collins, T. (2021). Serial dependence occurs at the level of both features and integrated object representations. *Journal of Experimental Psychology: General.* https://doi.org/10.1037/xge0001159.
Davies, W. (2021). Colour relations in form. *Philosophy and Phenomenological Research*, 102(3), 574–594.
Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, 98(4), 569–603.
Dretske, F. I. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
Drew, T., Horowitz, T. S., Wolfe, J. M., and Vogel, E. K. (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. *Journal of Neuroscience*, 31(2), 659–668.
Drummond, L., and Shomstein, S. (2010). Object-based attention: Shifting or uncertainty? *Attention, Perception, & Psychophysics*, 72(7), 1743–1755.
and ——— (2013). The timecourse of space-and object-based attentional prioritization with varying degrees of certainty. *Frontiers in Integrative Neuroscience*, 7, 88.
Egly, R., Driver, J., and Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123(2), 161–177.
Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., and Kellman, P. J. (2013). Automatic feature-based grouping during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1625–1637.
Firestone, C., and Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, 39, 1–19.
Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
Fodor, J. A. (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., and Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition*, 17(1–2), 48–66.
Gogel, W. C. (1971). The validity of the size–distance invariance hypothesis with cue reduction. *Perception and Psychophysics*, 9, 92–94.
Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111(1), 3–32.
Green, E. J. (2017). Psychosemantics and the rich/thin debate. *Philosophical Perspectives*, 31(1), 153–186.
——— (2018). What do object files pick out? *Philosophy of Science*, 85(2), 177–200.
——— (2019). A theory of perceptual objects. *Philosophy and Phenomenological Research*, 99(3), 663–693.
——— (2020). The perception-cognition border: A case for architectural division. *The Philosophical Review*, 129(3), 323–393.
——— (forthcoming). Perceptual constancy and perceptual representation. *Analytic Philosophy.*
Green, E. J., and Quilty-Dunn, J. (2021). What is an object file? *The British Journal for the Philosophy of Science*, 72(3), 665–699.
Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology*, 8, 1–12.
Hafri, A., and Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, 25(6), 475–492.
Haladjian, H. H., and Pylyshyn, Z. W. (2008). Object-specific preview benefit enhanced during explicit multiple object tracking. *Journal of Vision*, 8(6), 497.
Hill, C. S. (2021). Perceptual existentialism sustained. *Erkenntnis*, 86(6), 1391–1410.
Hollingworth, A., Maxcey-Richard, A. M., and Vecera, S. P. (2012). The spatial distribution of attention within and across objects. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 135–151.

Hollingworth, A., and Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 36(3), 543–564.
Howe, P. D. L., Holcombe, A. O., Lapierre, M. D., and Cropper, S. J. (2013). Visually tracking and localizing expanding and contracting objects. *Perception*, 42, 1281–1300.
Jiang, H. (2020). Effects of transient and nontransient changes of surface feature on object correspondence. *Perception*, 49(4), 452–467.
Johnson, K. (2015). Maps, languages, and manguages: Rival cognitive architectures? *Philosophical Psychology*, 28(6), 815–836.
Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219.
Kanwisher, N., and Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B*, 361, 2109–2128.
Kiyonaga, A., and Egner, T. (2016). Center–surround inhibition in working memory. *Current Biology*, 26(1), 64–68.
Kok, P., Jehee, J. F. M., and De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270.
Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., and Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, 28(11), 1649–1662.
Landman, R., Spekreijse, H., and Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43(2), 149–164.
Lee, M. D., and Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin and Review*, 11(2), 343–352.
Li, Y., Pizlo, Z., and Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision Research*, 49, 979–991.
Ling, S., Liu, T., and Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, 49(10), 1194–1204.
Long, B., Störmer, V. S., and Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, 17(6), 20, 1–20.
Lu, Z.-L., and Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *Journal of the Optical Society of America A*, 18(9), 2331–2370.
Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, 6(4), 547–569.
Macpherson, F. (2012). Cognitive penetration and colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24–62.
Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*, 97(2), 267–283.
Marino, A. C., and Scholl, B. J. (2005). The role of closure in defining the "objects" of object-based attention. *Perception and Psychophysics*, 67(7), 1140–1149.
Meng, M., and Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision*, 4, 539–551.
Montague, M. (2019). The sense/cognition distinction. *Inquiry*. https://doi.org/10.1080/0020174X.2018.1562371
Moore, C. M., and Enns, J. T. (2004). Object updating and the flash-lag effect. *Psychological Science*, 15(12), 866–871.
Moore, C. M., Stephens, T., and Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin and Review*, 17(5), 731–736.
Mylopoulos, M. (2021). The modularity of the motor system. *Philosophical Explorations*, 24(3), 376–393.
Nah, J. C., and Shomstein, S. (2020). Target frequency modulates object-based attention. *Psychonomic Bulletin and Review*, 27(5), 981–989.

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

Nes, A., Sundberg, K., and Watzl, S. (2021). The perception/cognition distinction. *Inquiry.* https://doi.org/10.1080/0020174X.2021.1926317

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

Odic, D., Roth, O., and Flombaum, J. I. (2012). The relationship between apparent motion and object files. *Visual Cognition*, 20(9), 1052–1081.

Papenmeier, F., Meyerhoff, H. S., Jahn, G., and Huff, M. (2014). Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 159–171.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

Phillips, B. (2019). The shifting border between perception and cognition. *Noûs*, 53(2), 316–346.

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

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

——— (2007). *Things and Places: How the Mind Connects with the World*. Cambridge, MA: MIT Press.

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

Quilty-Dunn, J. (2020a). Attention and encapsulation. *Mind and Language*, 35(3), 335–349.

——— (2020b). Perceptual pluralism. *Noûs*, 54(4), 807–838.

——— (2020c). Is iconic memory iconic? *Philosophy and Phenomenological Research*, 101(3), 660–682.

Quilty-Dunn, J., and Green, E. J. (2021). Perceptual attribution and perceptual reference. *Philosophy and Phenomenological Research.* https://doi.org/10.1111/phpr.12847

Recanati, F. (2020). Coreference *de jure*. In R. Goodman, J. Genone, and N. Kroll (eds.), *Singular Thought and Mental Files*, 161–186. Oxford: Oxford University Press.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7(1–3), 17–42.

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

Rolfs, M., Dambacher, M., and Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, 23(3), 250–254.

Rosenholtz, R. (2015). Texture perception. In J. Wagemans (ed.), *The Oxford Handbook of Perceptual Organization*, 167–186. Oxford: Oxford University Press.

Schmidt, F., Hegele, M., and Fleming, R. W. (2017). Perceiving animacy from shape. *Journal of Vision*, 17(11), 10, 1–15.

Scholl, B. J., and Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford and V. A. Kuhlmeier (eds.), *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*. Cambridge, MA: MIT Press.

Schut, M. J., Fabius, J. H., Van der Stoep, N., and Van der Stigchel, S. (2017). Object files across eye movements: Previous fixations affect the latencies of corrective saccades. *Attention, Perception, and Psychophysics*, 79(1), 138–153.

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.

Shomstein, S. (2012). Object-based attention: Strategy versus automaticity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(2), 163–169.

Shomstein, S., and Johnson, J. (2013). Shaping attention with reward: Effects of reward on space- and object-based selection. *Psychological Science*, 24(12), 2369–2378.
Shomstein, S., and Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception and Psychophysics*, 64(1), 41–51.
Siegel, S., and Byrne, A. (2016). Rich or thin? In B. Nanay (ed.), *Contemporary Debates in the Philosophy of Perception*. New York: Routledge.
Singh, M., and Hoffman, D. D. (2001). Part-based representations of visual shape and implications for visual cognition. *Advances in Psychology*, 130, 401–459.
Slotnick, S. D., and Yantis, S. (2005). Common neural substrates for the control and effects of visual attention and perceptual bistability. *Cognitive Brain Research*, 24(1), 97–108.
Sousa, R., Brenner, E., and Smeets, J. B. (2009). Slant cue are combined early in visual processing: Evidence from visual search. *Vision Research*, 49(2), 257–261.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14, 29–56.
Stepper, M. Y., Moore, C. M., Rolke, B., and Hein, E. (2019). The role of object history in establishing object correspondence. *Attention, Perception, and Psychophysics*, 82, 1038–1050.
Stokes, D. (2021). *Thinking and Perceiving: On the Malleability of the Mind*. London: Routledge.
Tipper, S. P., Weaver, B., Jerreat, L. M., and Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 20(3), 478–499.
Todd, P. M., and Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, 16(3), 167–171.
Tye, M. (2020). Filling in and the nature of visual experience. *The Harvard Review of Philosophy*, 27, 59–69.
vanMarle, K., and Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, 14, 498–504.
Walsh, V. (2003). A theory of magnitude: Common cortical metrics of times, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488.
Wolfe, J. M., and DiMase, J. S. (2003). Do intersections serve as basic features in visual search? *Perception*, 32(6), 645–656.
Wolfe, J. M., and Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3), 0058.
Wu, W. (2017). Shaking up the mind's ground floor: The cognitive penetration of visual attention. *The Journal of Philosophy*, 114(1), 5–32.
Wynn, K., Bloom, P., and Chiang, W.-C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition*, 83, B55–B62.
Zhou, K., Luo, H., Zhou, T., Zhuo, Y., and Chen, L. (2010). Topological change disturbs object continuity in attentive tracking. *Proceedings of the National Academy of Sciences*, 107(50), 21920–21924.