SENSORY BINDING WITHOUT SENSORY INDIVIDUALS

Jake Quilty-Dunn (WUSTL)

Penultimate draft; Forthcoming in A. Mroczko-Wąsowicz & R. Grush (Eds.),
Sensory Individuals, Properties, & Perceptual Objects: Unimodal and Multimodal Perspectives (OUP)

Abstract*:* The capacity for feature binding is typically explained in terms of the *attribution model*: a perceptual state selects an individual and attributes properties to it (Kahneman & Treisman 1984; Clark 2004; Burge 2010). Thus features are bound together in virtue of being attributed to the same individual. While the attribution model successfully explains some cases of binding in perception, not all binding need be understood as property attribution. This chapter argues that some forms of binding—those involving holistic iconic representations, which bind features outside the limits of attention and object files—don't fit the attribution model. The chapter then sketches an alternative *coordination model* of binding that construes icons as complex analog representations.

Keywords: analog, attention, binding problem, iconicity, perception, perceptual objects, representational format

> *"What was worked out at Ford was the practice of moving the work from one worker to another until it became a complete unit, then arranging the flow of these units at the right time and the right place to a moving final assembly line from which came a finished product."*
>
> —Charles E. Sorensen (1956), *My Forty Years with Ford*, p.116

§1—Introduction

One might think of visual processing as operating like an assembly line: subsystems work out particular subcomponents that are combined to generate a final product, which is then shipped out to consumers outside the visual system such as cognition, action, and memory. All the information that is properly *perceived* is encoded in a single ultimate percept. One point in favour of the assembly-line idea is that, when we attend to our perceptual experience, we don't seem to find distinct, redundant perceptual representations. Instead, many would argue, we find one unified perceptual experience (Bayne 2010), perhaps even transcending perceptual modalities (O'Callaghan 2014; cf. Bayne & Spence 2015).

Tempting though the assembly-line idea may be, when construed as a claim about the architecture of perceptual processing it is very likely false. Instead, various stages of visual processing output various products that are individually available to be consumed by downstream mental processes. For example, detailed early-visual representations are delivered to iconic memory, while representations of a limited number of objects are delivered to visual working memory (Sperling 1960; Luck & Vogel 1997; Quilty-Dunn 2020b; cf. Gross & Flombaum 2016). Likewise, while we can only explicitly report on a handful of objects at a time with any precision, the visual system is able to compute statistical summaries that range over dozens of objects (Dakin 2001; Whitney & Yamanashi Leib 2018). We frequently categorize objects based on visual perception of their contours, but we may plan to interact with them physically based on visual perception of their medial axes (Firestone & Scholl 2014). Instead of a single assembly line, therefore, we might invoke the metaphor of an entire factory generating a multiplicity of products that may or may not share features with one another and are shipped out to consumers separately (cp. Burnston & Cohen 2013; Henke 2021).

The heterogeneity of vision is important to keep in mind when considering general questions like, "What is the format of perceptual representation?", or "How does the visual system bind features together?" The visual system may represent the world in multiple formats, some language-like and shared with paradigmatic instances of cognition and others iconic and proprietarily visual (Quilty-Dunn 2020c). Likewise, it may have multiple distinct ways of binding features together and understanding one sort of feature binding may fail to shed light on others (Lande 2020). Even the same kinds of features—for example, shape and colour—might be bound in different ways by different visual subsystems. We might thus adopt a default pluralist attitude toward strategies for feature binding, which will be the topic of the speculative ideas developed in this chapter.

The basic problem of feature binding is the problem of how combinations of features are represented, such that combinations like *red square* and *blue circle* are distinguished from combinations consisting of the same primitive features, like *red circle* and *blue square*. Call the distinguishability of distinct feature combinations containing the same primitives the *core phenomenon* of feature binding.

When we dive deeper into the meaning of terms like 'feature' and 'combine', we find many distinct questions that fall under the heading of "The Binding Problem" (Treisman 1996). For example, the *computational* binding problem concerns what algorithms are implemented to construct representations of feature combinations. The *neural* binding problem concerns how feature binding is neurally instantiated. We can also ask, concerning a particular representation, what structural properties enable it to encode a feature combination.[1] This *representational* binding problem itself breaks down into two subproblems. First, the *semantic* binding problem concerns how the contents of complex representations unify simpler elements.[2] Second, the *syntactic* binding problem concerns how a feature combination is represented in one representational vehicle.

My topic here is the syntactic binding problem. The question I'll focus on is: what kinds of structures underlie visual representations of feature combinations? The aim of this brief chapter is modest and programmatic. Rather than arguing that existing approaches to feature binding are completely wrong, my goal will be merely to sketch and provide preliminary motivation for a pluralist approach.

There is very good reason to think an important class of feature binding involves attributing features to objects (Kahneman et al. 1992; Pylyshyn 2003; Cohen 2004; Matthen 2004). There may also be reason to think of some episodes of feature binding as accomplished through attribution to locations (Treisman 1996; Clark 2000; 2004). I'll suggest that there may be a distinctly *iconic* form of feature binding that need not be characterized in terms of attribution at all, at least syntactically (i.e., the way binding is reflected in the structure of the representational vehicle). I'll argue that the holistic character of iconic representations is best understood not in terms of attribution, but rather in terms of *coordination*. That is, icons plot vectors in multidimensional property spaces rather than selecting individuals and attributing properties to them. Binding via coordination is not intended to replace

---

[1] O'Callaghan distinguishes "feature binding awareness," or the conscious experience of an object as having multiple features, from "feature binding processes," which are processes that take representations of individual features and combine them, yielding a bound representational output (2014, 75). Thus the computational binding problem may be solved by a "feature binding process" in O'Callaghan's sense, but characterizing the structure of a bound representation doesn't presuppose this.

[2] The semantic binding problem only arises if representational contents are structured, as on Fregean and Russellian views of content. The problem of the "unity of the proposition" might be seen as a form of the semantic binding problem (Gaskin 2008; Ostertag 2019). One might argue that the problem dissolves if contents are unstructured (Stalnaker 1984).

binding via attribution. Instead, I suggest that both forms of binding are generated by different elements of the diverse collection of processes that constitutes the human visual system.

§2—Varieties of Binding

2.1—Binding as attribution. One common way of understanding binding is by appeal to property attribution. In binding through attribution, one picks out an individual, either *de re* or through quantification, and attributes multiple properties to it. A demonstrative thought exhibiting predicate conjunction like THAT OBJECT IS RED AND SQUARE is a paradigmatic example. The properties *red* and *square* are bound because they're attributed to the same individual: the referent of THAT OBJECT. Distinct constituents of the representation correspond to the individual and to its various features, and the way those constituents are organized instantiates property attribution, thereby binding the represented features together. Call this idea the *attribution model* of binding.

The attribution model provides an intuitive, satisfying answer to the syntactic binding problem. At its core is the idea that distinguishing feature combinations requires figuring out what the *things* are that have the relevant properties. That is, for sensory binding, there must be a class of sensory individuals that function as the bearers of combinations of properties. There is controversy about what sensory individuals are: are they something like locations (Clark 2000; 2004), are they material objects (Cohen 2004; Matthen 2004), can they vary across modalities (O'Callaghan 2008; Green 2019), do they fail to correspond to a particular ontological class (Green 2018), are they only represented after the deployment of attention (Treisman & Gelade 1980), etc.

Attribution-based approaches to feature binding are so prominent that it is common for theorists to explicate the very idea of binding in terms of attribution. Treisman describes feature binding as an operation in which "different properties (e.g., shape, colour, and motion) must be bound to the objects that they characterize" (1996, 171). Clark explains that the binding problem crops up under many different names, including the "Many Properties Problem" (Jackson 1977), and understands the underlying problem as a matter of "how to represent one thing as having more than one feature" (Clark 2004, 447; cp. Clark 2000, 26ff).

Despite the ubiquity of the attribution model, we ought not to characterize binding in general as simply an instance of property attribution. An immediate problem for the generality of the attribution model is that binding seems closely tied to *compositionality*, which outstrips property attribution. For example, a complex concept like RED SQUARE represents *red* and *square* together without attributing them to an individual. In this case, something like a Merge operation is deployed to create a complex representation out of simpler constituents through set formation (Chomsky 1995, 223; cp. Camp 2018). Complex concepts can plausibly be stored in and retrieved from memory without figuring in a propositional structure and can therefore bind features without property attribution. Binding through Merge thus need not involve picking out an individual and attributing properties to it.

I suggest we interpret the attribution model as providing an *explanation* of binding rather than an initial characterization of the phenomenon (cp. Cohen 2004, 472). The syntactic binding problem is how visual representations package information in such a way that feature combinations are distinguished from non-coextensive combinations that incorporate the same primitive features. Attribution of multiple features to the same individual is one way that this might be accomplished. Merge is another.

The fact that the syntactic binding problem is characterizable without appeal to attribution, together with the fact that Merge provides an example of binding without attribution, suggests that we ought to be open to the possibility of non-attributive binding in the visual system. I don't believe that, as a matter of fact, Merge provides the best model for thinking about non-attributive binding in vision. But it encourages exploration of other ways of thinking about binding beyond the attribution model.

2.2—Holistic icons. Merge is an operation that takes two separate syntactic items, A and B, and transforms them into a new, complex syntactic object with A and B as constituents (Chomsky 1995, 223). The concepts RED and SQUARE are both constituents of RED SQUARE. There are two key components to Merge as I'll understand it: (a) the constituents are bound together into one representation, but (b) they are *discrete*, i.e., even once bound into a complex, they remain separate vehicles.

The ability to form complexes out of constituents that remain discrete even once composed is characteristic of *discursive*, or language-like, representational formats. For example, the propositional thought THAT OBJECT IS A RED SQUARE binds RED and SQUARE through predication, but these constituents remain discrete.[3] Thus a purely formal logical rule could allow one to deduce THAT OBJECT IS A SQUARE through predicate deletion (Quilty-Dunn & Mandelbaum 2020).

This aspect of discursive formats furnishes concrete empirical predictions. For example, a complex like RED SQUARE that's held in working memory should be able to lose each feature independently, such that performance might decay for one feature while remaining robust for the other. Green and Quilty-Dunn (2017) cite evidence that visual object representations exhibit just this kind of "independent forgetting" of distinct feature dimensions like colour and orientation (Fougnie & Alvarez 2011; see also Markov et al. 2019). Thus Green and Quilty-Dunn argue that visual object representations, or "object files," have a discursive representational format.[4]

In Treisman's Feature Integration Theory (FIT), values along distinct feature dimensions like colour, orientation, motion, texture, etc., are initially detected independently of one another (Treisman & Gelade 1980). This initial independence creates the need for a later binding operation or "feature integration" process. Object files are posited to be the outputs of that binding operation: representations of individual objects that track them through time and space and attribute multiple features to them, binding and storing those features in visual working memory (Kahneman et al. 1992).

The role object files play in FIT is the most widely accepted instance of binding via attribution in perception. I take the fact that we sometimes bind feature-representations by integrating them in object files, and thereby attributing the features to the same object, to be as near to established fact as anything in this neck of the woods. Thus Cohen's (2004) and Matthen's (2004) point that feature binding in vision involves attributing properties to objects is well taken, as is Clark's (2000; 2004) and Matthen's (2004) appeal to predicate-argument structure.

If we assume both that binding in vision is *only* accomplished by attribution via object files and that object files have a discursive format, then it follows that binding in vision is only accomplished by representations in a discursive format. This conclusion is intuitively suspicious. If vision is really

---

[3] One might think of attribution, at least of the kind that's instantiated in propositional thoughts with predicate-argument structure, as one kind of Merge. Chomsky allows for distinct types formed through Merge in order to accommodate syntactic phenomena like headedness, and predicate-argument structure may be one such type. I don't think anything in the text hangs on this issue. What matters is that not all cases of Merge are cases of predication or attribution, such that Merge can accomplish binding without attribution. For more on the predication/attribution distinction see Burge 2010 and Quilty-Dunn 2020a.

[4] There are independent reasons to think object files are discursive. Space limitations and fear of self-plagiarism prevent me from rehearsing them here, but see Quilty-Dunn 2020a; 2020c.

couched in multiple formats, including proprietary (e.g., iconic) perceptual formats, it would be surprising if binding could only occur in one of them. We should therefore consider what binding looks like outside the context of discursive predicate-argument structure and Merge.

Perhaps the most salient alternative to discursive representational formats are *iconic* representational formats. Consider an image of a green marble cylinder (Fig. 1).

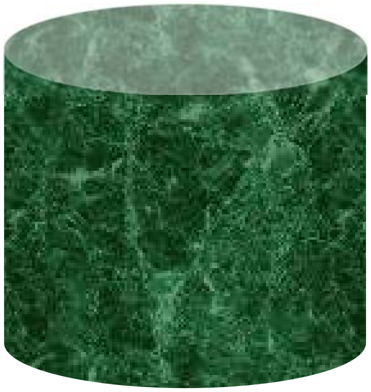

Figure 1

The first thing to notice about Figure 1 for present purposes is that it binds features together. It does not merely represent *green*, *marble*, and *cylinder*, but rather encodes the combination of those features.[5] An image of a marble cylinder and a wooden cube, for example, differs from an image of a wooden cylinder and a marble cube. Ordinary icons like Figure 1 thus exhibit what I called above the "core phenomenon" of binding: they distinguish different combinations of the same primitive features.

Another thing to notice is that icons like Figure 1 are *holistic*. That is, the multiple features instantiated by some part of the depicted scene are represented in one and the same corresponding part of the icon. In the phrase 'That is a green marble cylinder', each feature is represented by a separate constituent; but in Figure 1, the same part of the icon that represents the visible colour also represents visible texture and shape properties. Bundles of properties are not laid out in an orderly discrete fashion in an icon, as they are in a sentence. Instead, features are somehow packed into the same part of the representational vehicle. This aspect of iconic representation is at once obvious and puzzling, especially if we think of representation (mental and otherwise) on the model of language and predication.

The claim that icons are holistic doesn't entail that a visual icon mandatorily encodes every visible feature. The idea of mandatory encoding of many features is perhaps most influentially

---

[5] I use 'marble' and 'wooden' to denote diagnostic visual textures of the relevant materials, not the materials themselves.

embodied in Dretske's (1981) notion of "analog" representation. (His notion of analogicity is sufficiently unusual that I'll use 'Dretske-analog' when discussing it.) For Dretske, all representations carry *nested* information—that is, a representation that says *the cup is scarlet* also thereby carries the information that *the cup is red* and that *the cup is coloured*. But what distinguishes Dretske-analog from Dretske-digital representations is that Dretske-analog representations invariably carry *non-nested* information.[6] The sentence 'the cup is maroon' is silent on properties of the represented state of affairs beyond what can be logically inferred from the fact that there is a cup and it is maroon. But a photograph of the same scene must tell you more about it: for example, depending on the viewpoint, it will also represent the shape of the cup, whether it has liquid in it, what surfaces and objects surround it, etc.

But the holistic character of icons doesn't require that they encode many primitive features. Icons are holistic because, *when* two features are represented at a particular part of the represented scene, the same part of the icon that corresponds to that part of the scene and represents one feature will do double-duty and represent the other as well. An icon may not represent a given feature, as when a grayscale photograph fails to represent chromatic colour. But if a particular part of the scene is represented in a colour photograph, the same part of the photograph that represents a value along the brightness dimension will also represent a value along the hue dimension.

The holisticity of icons is underemphasized in discussions of iconicity. Theories of icons (at least *qua* mental representations) tend to focus on some version of the "parts principle" (Kulvicki 2015): parts of icons correspond to parts of the represented scene (Sober 1976; Kosslyn 1980; Fodor 2007; Carey 2009; Toribio 2011; Kulvicki 2015). This principle does not entail holisticity. But if we add an additional constraint to the parts principle—namely, that icons are isomorphic to what they represent, such that each represented part of the scene corresponds to exactly one part of the icon and vice versa—then holisticity follows (Quilty-Dunn 2020b). For any part of the scene, exactly one part of the icon corresponds to it. Therefore, if two properties are instantiated at that part of the scene and are represented by the icon, the same part of the icon must represent both properties.

The holistic format of icons precludes certain solutions to the syntactic binding problem. For example, icons lack separate vehicles for objects and their features. The attribution model thus seems not to apply, at least as far as the representational *vehicle* is concerned. Whether attribution might still be the best model of the representational *content* of the icon is unclear (see below). Moreover, since a

---

[6] More precisely, for Dretske, representations *carry bits of information* in an analog/digital way; thus one representation may encode some facts analogically and others digitally.

complex representation formed through Merge like RED SQUARE maintains discrete constituents, Merge also seems ill-suited to characterize icons. We've therefore run into a puzzle: how should we think about feature binding for iconic representations?

One might be tempted to dismiss the relevance of holistic icons to vision, and instead argue that we can avoid this puzzle entirely by sticking to the attribution model. But even if one is sceptical of the idea that mental representations are iconic in the sense I've outlined, the puzzle about feature binding remains, because *some* holistic icons exist. Figure 1 is an example; ample photographs and drawings are examples as well. The fact that icons exist and bind features together without explicit demonstrative-like representations of individuals, quantifiers, or Merge is a datum that needs to be explained.

Thus far I've explicated the basic idea of holistic icons and argued that (i) they exist, (ii) they bind features, and (iii) they don't seem well-suited to an attribution- or Merge-based account. I'll now argue that philosophers of perception have special reason to be interested in how icons bind features, because there is evidence that the visual system constructs and computes over icons.

§3—Holistic Icons in Vision

As discussed above, there is evidence that object files—representations that pick out individuals, track them across space and time, and attribute properties to them—are discursive rather than iconic. According to one natural interpretation of Feature Integration Theory (FIT) (Treisman & Gelade 1980), the visual system first detects features in parallel, then binds them only once attention is deployed and an object file is created, at which time the features are entered into the same object file and thereby attributed to the same object. If we combine the ideas contained in the previous two sentences, we arrive at the conclusion that there are no holistic icons in vision after all (Clarke forthcoming).

I suspect something like the interpretation of Feature Integration Theory (FIT) just sketched is common among philosophers of perception.[7] But it provides an impoverished picture of the kinds of representations constructed in vision. And it yields a clear falsifiable prediction: there is no feature binding independently of attention.

---

[7] Treisman herself continually modified FIT (e.g., Treisman & Sato 1990), and the characterization I describe here is a caricature of the nuanced views she developed over decades. This caricature is widespread, however (as Wolfe [2021] notes), and is thus worth discussing.

FIT enjoys many diverse forms of empirical support, far too many to survey here.[8] One of the most striking forms of evidence for FIT is the phenomenon of *illusory conjunctions*. In one classic experiment, subjects reported briefly presented objects, which varied in features like shape and colour (e.g., a green "N" and brown "T") (Treisman & Schmidt 1982). They found that, while subjects would occasionally misreport features that were not present (e.g., saying "red" when nothing in the display was red), they were even more likely to incorrectly integrate present features together (e.g., reporting a green "T"). And as FIT predicts, these so-called illusory conjunctions are more likely when attention is taxed and are less likely for objects that fall within the focus of spatial attention (Prinzmetal et al. 1986). Treisman and Schmidt noted that the existence of illusory conjunctions in the absence of attention is "the central claim" of FIT (1982, 108), and Prinzmetal calls it the "clearest manifestation" of FIT's notion of feature binding (2012, 215).

While illusory conjunctions provide powerful evidence that *some form* of accurate binding requires attention, this does not entail that *all* cases of binding require attention. It might be that the specific tasks used, such as verbal report and other overt responses, require the use of a certain kind of representation. As Mordkoff and Halterman put it, attention may not always be needed for binding, but rather "is needed only to select and produce the task-related, overt response" (2008, 385). If icons are truly couched in a different format from the discursive format ordinarily used for cognition and purposeful action, then they may only surface indirectly in behaviour and be masked by other representations more directly available to decision and report.

Object files are known to operate across mental systems. For example, they operate cross-modally and play a key role in cross-modal folk physics, particularly as studied in the minds of infants (Spelke 1988; Carey 2009). These findings make sense. The discursive format of object files makes them well-suited to play an interfacing role between perception and other cognitive systems including language. Another reason this kind of interfacing role makes sense for object files is that they are normally available to visual working memory (Gao et al. 2011; Quilty-Dunn 2020b), allowing them to function in a shared cognitive workspace.

---

[8] One crucial kind of evidence I regrettably cannot discuss here is visual search, which was a focus of Treisman and Gelade's (1980) original presentation of FIT and Wolfe's influential FIT-based approach to visual attention, Guided Search (Wolfe et al. 1989; Wolfe 2021). The standard findings are that search latency increases as the number of items increase and, more importantly, this item-based increase is markedly higher for conjunctions than individual features. According to Prinzmetal, this conjunction-based decline in search efficiency is "the evidence most often cited in favor of" FIT (2012, 214). But as Prinzmetal also points out, the visual search literature provides at best ambiguous support for FIT (see, e.g., Duncan & Humphreys 1989; Dosher et al. 2004; Hulleman & Olivers 2017).

Perhaps attention is normally required for correct binding of features *into object files*, thereby attributing them to a particular object. And perhaps this representational format is required for task performance, due to the role object files play at the interface of perception and cognition. But it may still be that other experimental evidence, involving different task demands, can provide evidence for binding outside the realm of attention and object files. The hypothesis that there are holistic icons in vision predicts that we should find such evidence.

I've argued elsewhere that evidence of this sort exists (Quilty-Dunn 2020b). Iconic memory, a high-capacity visual short-term memory store, seems to process feature conjunctions in a holistic fashion (Burns 1987; Pinto et al. 2013; Bronfman et al. 2014),[9] and in greater number than we can attend to or bind into object files (Landman et al. 2003). Other evidence comes from ensemble perception, the rapid extraction of statistical summaries of large numbers of objects. The representations used to compute these summaries represent many more items than can be attended to or represented via object files (Utochkin & Tiurina 2014)—perhaps by at least one order of magnitude (Dakin 2001)—and proceed independently of visual-working-memory load (Epstein & Emmanouil 2017). They also seem to encode many perceptible features at once (see Whitney & Yamanashi Leib 2018), including feature conjunctions (Boduroglu & Yildirim 2020).

For the purposes of this chapter, however, I want to focus on one probative piece of evidence from Mordkoff and Halterman (2008). Their goal was to develop a task that tested for binding outside attention but didn't require encoding the feature conjunction in a format usable for decision or report. The solution was to use a version of the *flankers task* (Eriksen & Eriksen 1974). In the standard flankers task, a target appears in the middle of the screen and is surrounded by "flankers" (e.g., two on each side). The task typically involves discriminating the target by hitting a button, and typically requires different responses for different targets. So for example, the task might be to hit a left-hand button if the target is a "1" or "2", and to hit a right-hand button if it's a "3" or "4". Like the famous Stroop effect, the standard effect is that performance is better on trials where the flankers are "congruent" with the target (e.g., target=1, flankers=2) and worse on "incongruent" trials (e.g., target=1, flankers=3).

Flankers are useful in this context because they must be perceived to affect performance even though the subject is attending only to the target. Thus the efficacy of flanking feature conjunctions could provide evidence for feature conjunctions outside attention. In the standard flankers task,

---

[9] I'm equating iconic memory and so-called "fragile visual short-term memory". The putative distinction between the two shouldn't matter here.

however, both target and flankers have a particular "response code" (i.e., right- or left-hand button press). Using the standard flankers task to probe the effects of feature conjunctions would thus raise the worry that attention is necessary for representing a feature conjunction in a format that can trigger the response code, rather than for binding *per se*.

As a result, Mordkoff and Halterman also used a *correlated* flankers task (Miller 1987). The targets and flankers were coloured shapes (i.e., feature conjunctions). One crucial feature of their task is that the conjunction-defined flankers were never used as targets (though flankers and targets shared primitive features). Conjunction-defined flankers therefore had no assigned behavioural response. Trials were balanced so that conjunctions were *correlated* with responses, e.g., trials on which flankers were yellow diamonds were also trials on which the correct response to the target was a left-hand key press. This correlation allowed them to look for an effect of feature conjunctions on performance when the conjunctions are both unattended and not assigned an overt response.

FIT predicts that, in a *standard* flankers task, conjunction-defined flankers should fail to affect performance since they fall outside the focus of attention and thus cannot be processed as conjunctions. Mordkoff and Halterman found this null effect as predicted, replicating earlier findings (Cohen and Shoup 2000). This result is also predicted, however, by the more restricted hypothesis that attention is necessary for binding features into object files that are available for decision and report, which is compatible with holistic icons representing feature conjunctions outside attention.

In the *correlated* flankers task, however, this hypothesis predicts that the feature conjunctions *should* have an effect, whereas FIT should again predict that conjunctions cannot be processed outside of attention, and thus conjunction-defined flankers should have no impact. Mordkoff and Halterman did find an impact, however: flankers that were correlated with targets grouped by a certain behavioural response improved discrimination of those targets (and impaired discrimination of others), *even when the correlation held only for feature conjunctions*.

One might object that participants attended to the flankers, perhaps because they noticed their correlations with response codes. But Mordkoff and Halterman replicated the effect in blocks of trials that interspersed the standard and correlated flankers test, eliminating the usefulness of adopting a general strategy of attending to flankers. They also replicated the effect while moving the flankers further away from targets and placing other flankers in between the targets and correlated flankers, making it yet less likely that the correlated flankers were attended to. Furthermore, the idea that subjects attend to flankers contradicts FIT's explanation of why conjunction-defined flankers fail to affect performance in the standard flankers task.

These results provide striking evidence that unattended feature combinations are bound together in vision. Equally importantly, they also show that this fact is masked in tasks that require explicit behavioural response to the relevant combinations. There seem to be multiple distinct forms of binding at work in the visual system, which differ in their relationship to cognition and report.

The attribution model is plausible in the case of object files, which bind features in a discursive format and thereby make feature combinations available to be exploited by cognition and purposeful action, including conceptualized response codes like *red square→left button*. Tasks that require this sort of availability require attention to process feature combinations. But other visual processes construct holistic icons, which encode feature combinations in a fundamentally different format. We can't expect feature combinations represented iconically to exhibit the same sort of global availability distinctive of object files. Feature combinations in visual working memory ("VWM") are not holistic (Fougnie & Alvarez 2011), but are object-based (Markov et al. 2019), suggesting that discursive object files are the vehicles of feature binding in VWM. It is thus possible that icons are not available to VWM.[10] In that case, the various cognitive and motor processes that are mediated by VWM may not be able to exploit iconically represented feature conjunctions. Thus the fact that we can only find evidence of holistic binding in iconic memory, or indirectly through influences on other processes (e.g., ensemble perception and discriminating correlated targets), is to be expected.

Two objections before moving on.

First objection: perhaps holistic icons are *neurally* implausible, because we know that early areas in visual cortex represent features independently. The idea that early vision consists solely of "specialized populations of receptors that respond selectively" (Treisman & Gelade 1980, 97) is widespread (Livingstone & Hubel 1988; Zeki 1993). Such independent areas are thought to correspond to distinct "feature maps" in primary and secondary visual cortex, a claim which has frequently been a starting point for philosophical discussions of feature binding (e.g., Clark 2000, 27ff; Campbell 2002, 30ff; Matthen 2005, 67ff; Clarke forthcoming).

---

[10] This generates another puzzle: how could icons not be available to VWM, given that visual imagery uses icons and apparently requires VWM resources (Hyun & Luck 2007)? Visual imagery, however, is not simply a matter of tokening icons. Even on iconic models, visual imagery requires fluid interaction with discursive representations of imaged objects, scenes, part-whole structures, and various other contents (Kosslyn 1980). Furthermore, icons outside of VWM might be "pointed" to by object files in VWM (Balaban et al. 2019), establishing functional interaction without making icons themselves available to VWM.

However, despite having attained the status of received wisdom, the hypothesis that distinct feature dimensions like colour and orientation are only represented independently in early visual cortex is false. Leventhal et al. (1995) found that, contrary to Livingstone and Hubel's (1988) widely cited claim that features are segregated in the primate primary visual cortex (a.k.a. striate cortex, or V1), "it appears that most cells in layers 2 and 3 are selective concomitantly for aspects of form, motion, and color" and thus "cells exhibiting the sort of stimulus selectivity required by the Livingstone and Hubel model do not exist in significant numbers in monkey striate cortex" (Leventhal et al. 1995, 1817).

More recently, Garg et al. (2019) used two-photon calcium imaging to examine whether cells in V1 show dimension-specific selectivity.[11] They found that, among cells that were strongly selective for colour, the *majority* were also strongly selective for orientation; among cells that were selective for orientation, nearly half were also selective for colour. Garg et al. conclude that "shape and color are mutually and unambiguously extracted and represented in a substantial population of V1 neurons" (2019, 1278). The widespread view that shape and colour are primarily neurally segregated in early vision appears to be wrong.[12]

Second objection: perhaps what looks like representation of feature combinations here is not really representation at all. That is, perhaps these states are merely registrations of proximal stimulus at the earliest stages of perceptual processing, before the visual system purports to represent the distal environment. Treisman argues that there may be early feature binding but that such conjunctions represent only "properties of the retinal stimuli," not "real-world properties, after constancy mechanisms have operated" (1988, 204). Likewise, Matthen argues that vision prior to attention and object individuation corresponds to spatial properties of the retina, not the distal environment (2004, 507ff).

Fortunately, there is evidence that bears directly on this issue. As mentioned above, ensemble perception seems to run statistical summaries over icons that represent more items than can be attended to. Im and Chong (2009) tested ensemble perception of average size of circles using stimuli that instantiated the Ebbinghaus illusion, in which a circle looks larger or smaller depending on whether it's surrounded by small or large circles, respectively. They found that ensemble-coding

---

[11] Two-photon calcium imaging is a technique that involves using a two-photon microscope to observe changes in fluorescence in individual cells caused by fluctuations in calcium concentration.

[12] The fact that there is widespread integration of color and form in V1 is compatible with there *also* being segregated color-specific and form-specific processing (Liu et al. 2020). In keeping with the rejection of the "assembly line" model, it would make sense for there to be both parallel feature detection and holistic icons in early vision.

processes averaged over the illusory *perceived* size, not retinal size. Likewise, several recent experiments have found that average size is computed for perceived size, integrating depth cues prior to ensemble coding (Tiurina & Utochkin 2019; Haberman & Suresh 2020; Markov & Tiurina 2021). Cells in V1 also show size constancy (Sperandio et al. 2012) and orientation constancy (Sauvan & Peterhans 1999).

§4—The coordination model

I've been arguing that there is reason to posit iconic representations in vision, and that neither the attribution model nor Merge capture how these representations solve the syntactic binding problem. I'll now propose an alternative model: icons plot coordinates in multidimensional spaces, binding values in those spaces without exemplifying the predicate-argument structure characteristic of the attribution model.

The notion of iconic representation is often discussed together with the notion of *analog* representation (e.g., Kosslyn & Pomerantz 1977; Dretske 1981, 137; Carey 2009, 135). However, many paradigmatically analog representations are not iconic, such as mercury thermometers. I lack space to discuss previous philosophical work on this topic in anything like adequate detail (e.g., Goodman 1968; Lewis 1971; Haugeland 1981; Peacocke 1986; 2019; Maley 2011; 2020; Kulvicki 2015; Beck 2019; Clarke 2020). In Goodman's (1968) seminal discussion of the analog/digital distinction, he argued that analog representations are *dense*, i.e., between any two values along some semantically significant dimension of variation in the vehicle, there lies a third value. But analog representations arguably need not be dense (Lewis 1971).

Maley (2011) argues that density is unimportant, and that what's distinctive of analog representations is that variation along some dimension of the vehicle reliably corresponds to variation along the represented feature dimension. For example, whether a clock is analog doesn't seem like it should hang on whether the second hand glides from second to second (dense) or instead ticks abruptly (non-dense). The relevant joint in nature seems to be between both kinds of clocks and truly digital clocks, which represent changes in time by updating numerals (Quilty-Dunn 2017). What makes the abruptly ticking clock analog is that variation along a dimension of the vehicle (how far the second hand has moved) functions to represent variation along the represented dimension (number of seconds elapsed), and an increase/decrease in movement of the hand corresponds to an increase/decrease in the number of seconds elapsed. The same is not true of digital clocks, where numerals vary along dimensions (e.g., on an old digital clock, number of vertical or horizontal bars contained in the numeral) but such variation fails reliably to correspond to variation in the temporal dimension.

Other ways of characterizing analog representation exist, such as Peacocke's notion of "representation of magnitudes, by magnitudes" (2019, 52) and Lewis's (1971) claim that analog representations employ primitive physical magnitudes. But a common thread, exemplified in Maley's work, is that analog vehicles vary along dimensions in ways that mirror variation along represented feature-dimensions (Beck 2019; Clarke 2020). It's this common thread that I want to use to think about binding in holistic icons.

Recent work on analog format, almost without exception, treats analog features individually. This focus may be partly due to background assumptions consistent with the assembly line model of vision, particularly in the context of FIT. One might assume that analog feature representations are deployed separately, and then an object file is deployed to attentively bind features together. But as we've seen, this model is ill-equipped to account for binding outside of attention and object files. More generally, it's odd to think that analog representations of features don't compose, or that their composition is somehow dependent on object-file-like representations. An image like Figure 1 seems to bind features together *in a distinctly analog way*—the texture, colour, and position of a part of the surface of the cylinder are represented by means of dimensions in the vehicle that correspond to feature dimensions, and they are bound together. An icon is a certain sort of *complex analog representation*. We need some meaningful notion of analog feature binding. Moreover, even the purported parallel feature detection posited by FIT requires binding outside object files for complex features that are treated as primitive: colour values contain values along multiple dimensions at once (e.g., hue and saturation), and this binding is accomplished prior to attention and attribution to objects.

So how are features bound in analog icons? One might imagine many elaborate ways that analog feature-values could be bound, but I think there is one extremely simple structure that suffices. Key to the notion of individual analog representations is the idea of variation along *dimensions*. Coordinate systems (e.g., Cartesian coordinates) allow a point to be plotted along multiple dimensions simultaneously. For example, the coordinates (3, 4) provide a representation of a point that has *both* a value of 3 along the x-axis and a value of 4 along the y-axis. Thus coordination provides a framework for conceptualizing how a single point can be plotted along multiple dimensions at once.

Now consider part of an icon. For simplicity's sake, consider a primitive part, like a pixel of Figure 1.[13] That part has values along multiple dimensions at once, including hue, brightness, saturation, and the spatial x- and y-axes of the image. We can therefore model that part of the icon as a point in a five-dimensional space, defined by a set of coordinates of the form (a_i, b_j, c_k, d_l, e_m), where each subscript denotes a particular feature dimension (e.g., hue) and each letter of a–e denotes a particular value along that dimension. Call this model *the coordination model* of binding in holistic icons.[14]

The primitive icon-part need not be understood as having a constituent structure consisting of five vehicles just because the set of coordinates we use to model it does. Instead, it merely exemplifies values along independently varying dimensions. We can distinguish *constituents* of a representation from its semantically exploitable *properties*, just as we more generally distinguish parts of objects from their properties. Consider a brown chair with four legs; while one could use 'part' to refer to both its back and its brown colour, there is clearly some relevant distinction between the chair's relation to these two things. The colour is not a part of the chair in the same sense that the back is (Olson 2017).[15] Likewise, the saturation of a pixel is not a part of it in the same sense that a pixel is part of the whole icon (cf. Lande 2020).

The core ideas behind the coordination model are familiar. The idea that aspects of perception can be understood by appeal to multidimensional feature spaces has long been discussed in the literature on perceptual *quality spaces* or *similarity spaces* (Clark 1993; Rosenthal 2010; Gauker 2012; Berger 2018; 2021). This literature focuses primarily on mental properties that surface in the qualitative character of conscious experience, and much of it assumes something like FIT (Clark 2000; Rosenthal 2005, 200). I've (deliberately) said nothing at all about conscious experience, and holistic icons require binding outside the constructs of FIT. But the basic idea of thinking of visual representations as plotting points in multidimensional feature spaces has long been discussed in this literature (cp. vectors and tensor products in connectionist models; Smolensky 1990; Eliasmith 2013, 387ff). It also echoes Haugeland's notion of iconic contents as "variations of values along certain dimensions with respect to locations in certain other dimensions" (1998, 192).

---

[13] This really is just for simplicity's sake. Icons can have non-pixel-like primitives (Davies 2020), and non-primitive parts of icons can bind properties; the latter is arguably required for binding of some properties, like global shape.

[14] 'Coordination' is also used, with a different meaning, in discussions of cross-modal binding (Fulkerson 2014; O'Callaghan 2017; Cohen, this volume).

[15] Some metaphysicians do hold that properties can be parts of objects in some sense, e.g., as "logical parts" (Paul 2002). My claim is only that there are *some* meaningful senses of 'part' in which properties are not parts.

The coordination model itself is very simple, and my word limit looms. I'll move directly to considering some objections, explicating the model further as I go.

One might object that the coordination model is not meaningfully distinct from the attribution model. Clark's (2000) version of the attribution model allows features to be attributed not only to objects (Clark 2004) but also to locations. For Clark, visual representations engage in *feature-placing*, placing features at locations and thereby attributing features to those locations. Why not then construe the non-locational features of an icon as attributed to locations as represented along spatial axes?

However, the apparatus of coordination provides a model of bound elements in icons that doesn't require taking any elements to be attributed to any others. A key feature of attribution is *asymmetry*: in the thought BANANAS ARE YELLOW, the property of being yellow is attributed to bananas while the property of being bananas is not attributed to yellow. But coordination does not have this asymmetry built in. A part of an icon varies along multiple dimensions at once and thereby plots a position in a multidimensional space. Nothing in the coordination model requires, or even suggests, that values along some feature dimensions are attributed to values along other coordinated dimensions. This is clear in the case of Cartesian coordinates: the x-axis and y-axis values are bound together, but neither is attributed to the other. Likewise, there is nothing in the syntactic features of an icon that compel us to construe a value along the hue dimension as attributed to a value along spatial dimensions anymore than to construe spatial values as attributed to colour values.

One might press further that there's something special about spatial properties that prevents them from being represented by means of values along dimensions that can be coordinated with values along (e.g.) colour and orientation dimensions without attribution. Clark (2000) rejects the idea of vehicle-properties that correspond to location, instead holding that features are simply placed at *external* locations, with no quasi-spatial relations holding between representations of features. Much of Clark's discussion relies on the idea that binding must be accomplished through attribution to external locations, and on his rejection of visual icons (Clark 2009).

However, phenomena like mental rotation (Shepherd & Metzler 1971), instructed scanning (Kosslyn et al. 1978), and spontaneous scanning (Finke & Pinker 1982) suggest that there *are* visual icons that encode spatial relations via functional relations between representations of features that covary with objective spatial relations—in other words, space is encoded by yet another analog dimension. It can therefore be coordinated with other analog dimensions without attribution.

One virtue of the attribution model is that it provides a straightforward account of *accuracy conditions*: the representation is accurate iff the selected individual has the properties attributed to it. I haven't said how coordination fixes accuracy conditions. But my topic here has been the *syntactic* binding problem: how is the vehicle structured such that it represents feature combinations? An answer to this question need not come prepackaged with an answer to questions about how accuracy conditions are determined. Perhaps causal relations to environmental individuals on occasions of use provide "targets" for iconically encoded feature combinations without elements of the vehicle corresponding to those individuals (Cummins 1996). But in the absence of such an answer, we're left with a familiar question: how do perceptual representations come to be accurate or inaccurate with respect to the perceiver's environment?

References

Balaban, H., Drew, T., & Luria, R. (2019). Neural evidence for an object-based pointer system underlying working memory. *Cortex* 119, 362–372.

Bayne, T. (2010). *The Unity of Consciousness*. New York: OUP.

Beck, J. (2019). Perception is analog: The argument from Weber's Law. *Journal of Philosophy* 116(6), 319–349.

Berger, J. (2018). A defense of holistic representationalism. *Mind & Language* 33(2), 161–176.

Berger, J. (2021). Quality-space functionalism about color. *Journal of Philosophy* 18(3), 138–164.

Boduroglu, A., & Yildirim, I. (2020). Statistical summary representations of bound features. *Attention, Perception, & Psychophysics* 82, 840–851.

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

Burge, T. (2010). Steps toward origins of propositional thought. *Disputatio* 4(29), 39–67.

Burns, B. (1987). Is stimulus structure in the mind's eye? An examination of dimensional structure in iconic memory. *The Quarterly Journal of Experimental Psychology* 39(A), 385–408.

Burnston, D., & Cohen, J. (2013). Perception of features and perception of objects. *Croatian Journal of Philosophy* 12(3), 283–314.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: OUP.

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

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Clark, A. (1993). *Sensory Qualities*. Oxford: OUP.

Clark, A. (2000). *A Theory of Sentience*. Oxford: OUP.

Clark, A. (2004). Feature-placing and proto-objects. *Philosophical Psychology* 17(4), 443–469.

Clarke, S. (2020). Beyond the icon: Core cognition and the bounds of perception. *Mind & Language*, doi: 10.1111/mila.12315.

Clarke, S. (Forthcoming). Mapping the visual icon. *Philosophical Quarterly*.

Cohen, J. (2004). Objects, places, and perception. *Philosophical Psychology* 17(4), 471–495.

Cohen, J. (This volume). Multimodal binding as mereological co-constituency.

Cohen, A., & Shoup, R. (2000). Response selection processes for conjunctive targets. *Journal of Experimental Psychology: Human Perception and Performance* 26(1), 391–411.

Cummins, R. (1996). *Representations, Targets, & Attitudes*. Cambridge, MA: MIT Press.

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

Davies, W. (2020). Colour relations in form. *Philosophy and Phenomenological Research*, doi: 10.1111/phpr.12679.

Dosher, B.A., Han, S., & Lu, Z-L. (2004). Parallel processing in visual search asymmetry. *Journal of Experimental Psychology: Human Perception and Performance* 30(1), 3–27.

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

Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review* 96(3), 433–458.

Eliasmith, C. (2013). *How to Build a Brain*. New York: OUP.

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

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* 16(1), 143–149.

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

Firestone, C., & Scholl, B. (2014). "Please tap the shape, anywhere you like": Shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science* 25(2), 377–386.

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

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

Fulkerson, M. (2014). *The First Sense: A Philosophical Study of Human Touch*. Cambridge, MA: MIT Press.

Gao, T., Gao, Z., Li, J., Sun, Z., & Shen, M. (2011). The perceptual root of object-based storage: An interactive model of perception and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* 37(6), 1803–1823.

Garg, A.K., Li, P., Rashid, M.S., & Callaway, E.M. (2019). Color and orientation are jointly coded and spatially organized in primate primary visual cortex. *Science* 364, 1275–1279.

Gaskin, R. (2008). *The Unity of the Proposition*. Oxford: OUP.

Gauker, C. (2012). Perception without propositions. *Philosophical Perspectives* 26, 19–50.

Goodman, N. (1968). L*anguages of Art: An Approach to a Theory of Symbols*. Indianapolis: Bobbs-Merrill.

Green, E.J. (2018). What do object files pick out? *Philosophy of Science* 85(2), 177–200.

Green, E.J. (2019). A theory of perceptual objects. *Philosophy and Phenomenological Research* 99(3), 663–693.

Green, E.J., & Quilty-Dunn, J. (2017). What is an object file? *British Journal for the Philosophy of Science*, doi: 10.1093/bjps/axx055.

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

Haberman, J., & Suresh, S. (2020). Ensemble size judgments account for size constancy. *Attention, Perception, & Psychophysics*, doi: 10.3758/s13414-020-02144-6.

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

Henke, B. (2021). A fresh look at the two visual streams. *Journal of Consciousness Studies* 28(5-6), 198–207.

Hulleman, J, & Olivers, C. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences* e132, 1–20.

Hyun, J-S., & Luck, S.J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin & Review* 14, 154–158.

Im, H.Y., & Chong, S.C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics* 71(2), 375–384.

Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge: Cambridge University Press.

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

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

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

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

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

Lande, K. (2020). Mental structures. *Noûs*, doi: 10.1111/nous.12324.

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

Leventhal, A.G., Thompson, K.G., Liu, D., Zhou, Y., & Ault, S.J. (1995). Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *The Journal of Neuroscience* 15(3), 1808–1818.

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

Liu, Y., Li, M., Zhang, X.,…, Mcloughlin, N., Tang, S., & Wang, W. (2020). Hierarchical representation for chromatic processing across macaque V1, V2, and V4. *Neuron* 108, 538 – 550.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science* 240, 740–749.

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

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

Markov, Y.A., & Tiurina, N.A. (2021). Size-distance rescaling in the ensemble representation of range: Study with binocular and monocular cues. *Acta Psychologica* 213, 103238.

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

Matthen, M. (2004). Features, places, and things: Reflections on Austen Clark's theory of sentience. *Philosophical Psychology* 17(4), 497–518.

Matthen, M. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: OUP.

Meyer, D.E. & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90, 227–234.

Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters. *Perception & Psychophysics* 41, 419–434.

Mordkoff, J.T., & Halterman, R. (2008). Feature integration without visual attention: Evidence from the correlated flankers task. *Psychonomic Bulletin & Review* 15(2), 385–389.

O'Callaghan, C. (2008). Object perception: Vision and audition. *Philosophy Compass* 3(4), 803–829.

O'Callaghan, C. (2014). Intermodal binding awareness. In Bennett, D. J. and Hill, C. S., editors, Sensory Integration and the Unity of Consciousness, pages 73–103. MIT Press, Cambridge, MA.

O'Callaghan, C. (2017). Enhancement through coordination. In Nanay, B. (Ed.), *Current Controversies in Philosophy of Perception* (New York: Routledge), 109–120.

Olson, E.T. (2017) Properties as parts of ordinary objects. In Keller, J.A. (Ed.), *Being, Freedom, and Method: Themes from the Philosophy of Peter van Inwagen* (Oxford: OUP), 62–79.

Ostertag, G. (2019). Structured propositions and the logical form of predication. *Synthese* 196, 1475–1499.

Paul, L.A. (2002). Logical parts. *Noûs* 36(4), 578–596.

Peacocke, C. (1986). Analogue content. *Proceedings of the Aristotelian Society*, Supplementary Volume 60, 1–17.

Peacocke, C. (2019). *The Primacy of Metaphysics*. Oxford: OUP.

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

Prinzmetal, W. (2012). At the core of feature integration theory: On Treisman and Schmidt (1982). In J. Wolfe & L. Robertson (Eds.), *From Perception to Consciousness: Searching with Anne Treisman* (New York: OUP), 211–216.

Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? *Journal of Experimental Psychology: Human Perception and Performance* 12, 361–369.

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

Quilty-Dunn, J., & Mandelbaum, E. (2020). Noninferential transitions: Imagery and association. In A. Nes & T. Chan (Eds.), *Inference and Consciousness* (London: Routledge), 151–171.

Quilty-Dunn, J. (2017). *Syntax and Semantics of Perceptual Representation*. PhD thesis, The Graduate Center, CUNY.

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

Quilty-Dunn, J. (2020b). Is iconic memory iconic. *Philosophy & Phenomenological Research* 101(3), 660–682.

Quilty-Dunn, J. (2020c). Perceptual pluralism. *Noûs* 54(4), 807–838.

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

Rosenthal, D. (2010). How to think about mental qualities. *Philosophical Issues* 20(1): 368–393.

Sauvan, X.M., & Peterhans, E. (1999). Orientation constancy in neurons of monkey visual cortex. *Visual Cognition* 6(1), 43–54.

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

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46, 159–216.

Sober, E. (1976). Mental representations. *Synthese* 33(1), 101–148.

Sorensen, C.E. (1956). *My Forty Years with Ford*. With S.T. Williamson. New York: Norton.

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

Spence, C., & Bayne, T. (2015). Is consciousness multisensory? In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and Its Modalities* (New York: OUP), 95–132.

Sperandio, I., Chouinard P.A., & Goodale, M.A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience* 15(4), 540–542.

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

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Tiurina, N.A., & Utochkin, I.S. (2018). Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging. *Journal of Experimental Psychology: General* 148(4), 728–738.

Toribio, J. (2011). Compositionality, iconicity, and perceptual nonconceptualism. *Philosophical Psychology* 24(2), 177–193.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology* 12, 97–136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance* 16(3), 459–478.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology* 6(2), 171–178.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions and the perception of objects. *Cognitive Psychology* 14, 107–141.

Treisman, A. (1988). Features and objects. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 40(2), 201–237.

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

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

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception & Performance* 15(3), 419–433.

Wolfe, J.M. (2021). Guided search 6.0: An updated model of visual search.

Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell.