Perceiving objects the brain does not represent

It is often assumed that neural representation, with content that is in principle detachable from the flow of natural-factive information, is necessary to perceptually experience an object. In this paper we present and discuss two cases challenging this assumption. We take them to show that it is possible to experience an object with which you are interacting through your sensory systems without those systems constructing a representation of the object. The first example is viewing nearby medium-sized groups of objects. The second is hearing objects through misbound sounds. These cases bring out two different ways object representation can fail while object experience persists, suggesting that object experience requires only that the object be revealed through sensory information, not full-blown representation. Constructing object representations is one way sensory systems reveal objects, but it is not the only way. Object representation in neural sensory systems is relatively demanding, rare, and fragile; object experience is relatively easy, pervasive, and robust. We conclude that even minimal forms of neural phenomenal internalism are false.

Keywords: perceptual experience; object representation; multimodal binding; ensemble perception; representationalism; phenomenal internalism.

1 Introduction

As Penny directs her eyes and attends to the pencil she is holding, she enjoys a visuohaptic experience of the pencil. At the same time, spiking in cortical circuits originating in her mechanoreceptors and photoreceptors represents the pencil. By ‘represents’, we mean that the spiking not only carries information about the pencil but that it serves as a *detachable stand-in*: were that spiking repeated in the absence of causal interaction with a pencil (e.g., as when dreaming), it would still function as a stand-in for a (nonexistent) pencil-looking object. We take this to be a standard use of the term ‘representation’ among both philosophers and psychologists (e.g., see Palmer (1978); Grush (1997); Martin (2002); Lee (2019)).

Most psychologists and neuroscientists, and many philosophers, (if asked) would say that object experience requires this neural object representation. On this view, to experience an object with which you’re interacting through your sensory systems, those systems must construct a representation of the object. By ‘experience’, we mean phenomenal consciousness. In experience, objects *show up* in your stream of phenomenal consciousness. We will say more about experience later (§3.2), but for now it is enough to rely on the intuitive, everyday understanding of object experience. Pick up a nearby object. Hold it. Look straight at it. You are having an experience of this object.

Sometimes scientists and philosophers explicitly commit to only the weaker view that experience requires neural information encoding, not the stronger condition that it requires detachable neural stand-ins. As we explain below (§2), however, many of these same researchers wish to endorse a minimal form of *neural phenomenal internalism* according to which reproducing neural activity that led to an experience while in causal contact with a distal stimulus will lead to or constitute the same experience, even if (during the reproduction) the subject is out of causal contact with anything like the original stimulus. This minimal neural phenomenal internalism is then explained by appeal to the *representational content* of the neural activity: the experience is reproduced, despite the lack of causal contact with anything like the original stimulus, *because* the reproduced neural activity still has the content of its original instance.[1]

This explanation only works, of course, if the neural activity is not just a mere 'factive' or 'natural' sign of the stimulus, encoding information in that sense, but is an 'intentional' or 'nonnatural' sign (Grice 1957). The neural activity must be (as we shall use the term) a *representation* that functions as a detachable stand-in for the actual, distal stimulus. Because of the widespread commitment to minimal neural phenomenal internalism, then, what we shall here call *neurorepresentationalism* is a fair reconstruction of the view held by a wide range of scientists and philosophers.[2]

Neurorepresentationalism: if there is some object *o* such that one perceptually experiences *o*, then one has some dedicated, detachable object-representation which successfully refers to *o*.

In this paper, we will argue against neurorepresentationalism. Since this is more or less entailed by minimal neural phenomenal internalism, our argument equally targets the latter view. We develop two counterexamples, each unlike the paradigm case involving Penny mentioned at the start. The first, originally suggested by Pylyshyn (2007), involves seeing objects in nearby medium-sized groups (§3). The second involves hearing an object via sounds wrongly presupposed to have been caused by some other, seen object (i.e., object perception

---

[1] More carefully: although they will deny that reproducing the neural activity that occurs when Penny looks at pencil *P*, in the absence of any corresponding distal stimulus, would produce a perceptual experience *of P*, they will insist that Penny would undergo a sensory experience as of a qualitatively identical object. To this extent, such views are opposed to disjunctivism about content in perceptual experience (or rather—since they could like Schellenberg (2018) claim that there is a level of singular content which plays some but not other explanatory roles—*at least* one level of content).

[2] Our focus is on object perception (for which the view is arguably somewhat more plausible), but one could equally use the term 'neurorepresentationalism' for the broader view that experiencing any particular—object, property instance, etc.—requires a neural representation referring to that thing.

that survives multimodal binding failure) (§4). These examples suggest that the construction of a detachable object representation by sensory neural circuits is perhaps *one way* to achieve object experience, but it is not the only way.[3] Whereas object representation in this sense is relatively demanding, rare, and fragile, object experience is relatively easy, pervasive, and robust.[4]

If experience and neural representations can come apart, are we denying physical explanations of consciousness? No. While this paper aims only at giving the case against the necessity of neural representations, there is a natural positive view which we later outline (§5), appealing to (mere) information encoding of a certain sort. To perceptually experience an object it is not always necessary to token a representation—i.e., a nonnatural or intentional sign, detachable from ongoing causal contact with its target. The object's being suitably revealed through sensory information, encoded in neural activity, can suffice.

Despite our use of the term 'neuro*representationalism*', we ask the reader not to shoehorn this paper into the debate between 'metaphysical' representationalists (e.g., Tye 1995; Dretske 2003; Matthen 2005; Burge 2010) and naïve realists or relationalists (e.g., Campbell 2002; Fish 2009). This paper is, at bottom, about the relation between what we experience and what is neurally represented. Many caveats would be needed to translate this into a paper about whether experience has 'content' or whether experiences are, at some basic level, relations or representations. Our arguments do not rule out full-fledged representationalist views that take experiences to be fundamentally representational; they only rule out versions of these views on which these representations *always* reduce to neural representations of a certain kind.

In §2 we clarify neurorepresentationalism and point to some examples of its being endorsed. §3 introduces the case of seeing objects in nearby medium-sized groups and outlines a two-premise argument for the claim that, while object representation is lacking in this case (§3.1), it is a genuine case of object experience (§3.2). Next, §4 introduces the case of hearing misbound sounds and argues that, for very different reasons, it is also a case of experiencing non-represented objects. §5 brings these two considerations together and briefly outlines the kind of information-based account that these cases support over neurorepresentationalism.

2 Neurorepresentationalism

[3] While some would say that even this concedes too much, we set broader forms of anti-representationalism aside for the purposes of this paper (*inter alia*, Gallagher (2017) and Hutto (2017)).

[4] We elucidate what we mean by this slightly rhetorical way of putting the point once our arguments have been presented, in §5.

We are specifically concerned with what neurorepresentationalists say about *perception*, i.e. *successful* (but not necessarily accurate) perceptual experience. You perceive when you experience objects, property instances, or events with which you're interacting through your sensory systems, even if you do not experience them entirely accurately. Neurorepresentationalism claims that to perceive a distal object *o* you must not only interact with *o* through your senses, but a dedicated, detachable representation of *o* (i.e., a neural representation that purports to single out a particular object and achieves reference to *o*) must be constructed.[5] Successful reference is the neurorepresentationalist's account of how mind connects to world in sense experience. A neural response which fails to achieve reference to a particular object can, of course, still cause an object experience. But it can't explain how someone manages to successfully perceive a particular object. Neurorepresentationalists (by their own lights) couldn't explain how Penny manages to perceive *the very pencil she is holding* by citing a neural representation which failed to refer it.[6]

One of the first explicit endorsements of neurorepresentationalism is Tye (1995). Tye is clear that by 'representations' he means detachable stand-ins. "Intentional mental states are states that can represent, or be about, things of a certain sort without there being any particular real things of that sort" (1995: 100). He goes on to give a neurorepresentationalist theory of perceptual experience:

> the receptor cells on the retina are [...] transducers. They have, as input, physical energy in the form of light, and they convert it immediately into symbolic representations of light intensity and wavelength. These representations are themselves made up of active nerve cells. [...] The computational procedures operating on these representations generate further symbolic representations first of intensity and wavelength changes in the light, then of lines of such changes, then of edges, ridges, and surfaces, together with representations of local surface

[5] Neurorepresentationalism is narrower than the 'one-to-one intentionalist' thesis recently criticised by Martínez & Nanay (forthcoming). According to that thesis, the phenomenal character of a perceptual experience supervenes on the purported content *of that experience*. They contend that the character of an experience instead simply depends on the content of a variety of representations potentially distributed across the brain. Albeit related, our target is the more specific claim that one's experiencing *an object* entails one's possession of a vehicle that represents *it*.

[6] It is not obvious that successful reference requires *determinate* reference, however. And so the neurorepresentationalist could claim that having a dedicated, detachable object-representation which *indeterminately* refers to *o* is nevertheless sufficient for one to perceptually experience *o*. This move would not help neurorepresentationalists deal with the kind of case discussed in §3. We discuss whether this move might help them to deal with the kind of case discussed in §4 later.

> features, for example, color, orientation, and distance away. [...] The output representations [of these procedures] track features of distal stimuli under optimal or ideal perceptual conditions. Thereby, it seems plausible to suppose, they represent those features, they become sensations *of* edges, ridges, colors, shapes, and so on (Tye 1995: 102–104)

Although the above quote only discusses the visual experience of features (properties), the aim of the book is to develop this into a broader theory of phenomenal consciousness.

Dretske (1995; 2003), diverging from earlier work (1981), famously developed a view similar to Tye's. Both Dretske and Tye agree that what we experience is what is represented by the neural representations. They reject *strong* forms of neural phenomenal internalism, holding that the content of neural representations need not be preserved across neural duplicates. What the spiking in Penny's head represents, on their view, depends in part on the normal causes of that spiking in Penny's environment, or the (biological or evolutionary) function of that spiking. Still, they accept, and are certainly motivated by, *minimal* neural phenomenal internalism: the whole point, for them, of appealing to representations is to explain how the reproduction of neural activity (assuming a fixed past environment, fixed biological function, and fixed evolutionary history) could lead to a reproduction of experience, even in the absence of the original, distal stimulus.

Some accept a stronger version of neural phenomenal internalism according to which neural duplicates are phenomenal duplicates even when those neural duplicates have different histories (e.g., Prinz (2000); (2012)). Prinz (2012) presents himself as largely just synthesizing the existing psychological literature on consciousness. Prinz is plausibly correct that the view he presents is largely in line with the standard views of scientists (even if the latter are not always sensitive to philosophically delicate distinctions). For example, in a paper on the neural correlates of consciousness, Crick and Koch write,

> For an animal to be aware of some aspect of a visual object, there has to be a group of similar neurons (to allow coarse coding) located somewhere in the cortical system [...] the firing of which is correlated with that feature in the visual scene (1995: 121).

By 'aware', it's clear from Crick and Koch's opening reference to the mind-body problem (not quoted here) that they mean experience, i.e. phenomenal consciousness. The correlated neural firing they mention is just the standard form that information encoding is assumed to take in the brain. And so what they offer is a necessary condition on perceptual experience: to experience a feature, neural encoding of that feature is required.

In that paper, Crick and Koch do not further add that this neural encoding must rise to the level of a 'representation', i.e. be a detachable stand-in. Still, it is clear from other remarks—appealing to minimal neural phenomenal internalism—that scientists like Crick and Koch have this further commitment in mind. For example, here is Koch endorsing a view of the neural correlates of consciousness (NCC) that clearly entails neurorepresentationalism:

> The content-specific NCC are the neurons (or, more generally, neuronal mechanisms), the activity of which determines a particular phenomenal distinction within an experience. For example, the NCC for experiencing the specific content of a face are the neurons that fire, on a trial-by-trial manner, whenever a person observes, imagines or dreams a face, and are silent in other circumstances. When the content-specific NCC neurons in this example are activated artificially [...] the participant should see a face even if none is present, whereas if their activity is blocked, the participant should not be able to see a face even if one is present (Koch et al. 2016: 308).

Similar remarks are made by other scientists discussing perceptual experience (e.g., Revonsuo (1995); Moutoussis (2016)). It's fair to say that this quote captures what is perhaps the standard picture among neuroscientists and psychologists who study consciousness.

As we have emphasized, if you endorse neural phenomenal internalism, you need to be concerned not only that everything experienced is encoded in the brain in the sense that neural responses are 'factive' or 'natural' signs of what prompts them, but also in the stronger sense that these encodings are 'intentional' or 'detachable'. Consider Koch's face-responding neurons. If their response is a mere natural, nonintentional sign of faces, then, when that response happens in the absence of a face, that response will not signal (or carry content as of) a face. In that case, there is no reason to suppose that 'the participant should see (*experience)* a face even if none is present'. After all, the whole explanation was supposed to be that the participant 'saw' a face because the neural response signaled a face. Clearly, those like Koch really care about nonnatural signs, or what we have called detachable stand-ins: content that is, crucially, not dependent on a concurrent flow of natural-factive information.

A more recent example of neurorepresentationalism is arguably Schellenberg's (2018) *capacitism*. According to Schellenberg, to perceive an object (property instance, event, etc.) is to employ a perceptual capacity that successfully refers to it. While perceptual capacities are person-level phenomena (attributable to subjects rather than neural systems), they have as their 'base' subpersonal, computational or functional states, events, and processes (2018: 32). Although capacities are multiply realizable and not inherently neural in basis, in typical humans they will have neural representations as their sole basis. Crucially, then, perceptual capacities are shared by neural duplicates (2018: 202); they may be employed not only successfully, when one's environment is felicitous, but also unsuccessfully, as in sensory hallucination (2018: 34). Perceptual capacities thereby constitute a level of *content type* insulated from the subject's environment (a 'common factor' in good and bad cases) and grounding phenomenal character. What is different in the case of successful perception is that these capacities afford *token (singular) content* because the employment of the perceptual capacity, on that occasion, succeeds in referring to a particular object. Here, too, then, perception is explained by the successful reference of a detachable representation, and "S seeing o entails that S represents o" (2018: 128) in this sense.

3 Perceiving objects within nearby, medium-sized groups

Suppose that, unlike Penny, Sally is looking not at a single pencil but at a jumbled cluster of 8-10 items of stationery on her desk: a couple of pens, pencils, highlighters, an eraser, and a pencil sharpener. Set aside her peripheral vision, which is in many ways a different case, and simply consider her foveal vision as she views the stationery. Sally sees the eraser but also various other objects. Or, to take another example, suppose that over the course of a few seconds Sam watches 10 balls scatter on an opening break in nine-ball pool. Across a number of saccades, Sam foveates on different areas of the pool table. The balls are not moving so quickly, nor are there so many of them, as to prevent him from discriminating them from one another via scene segmentation. In this sort of case, it would be natural to say that Sam saw each of the balls as they scattered, and it would certainly introspectively strike him that way as he watched.

Whether we do have such object experiences will be discussed further in §3.2. For now, let us clarify that we are not assuming a naïve or 'panorama' view of visual experience. We are not suggesting that Sally or Sam see *everything* in their field of view, periphery included, in rich detail. Our claim about the experience is quite limited. Their experience of these items will be more impoverished than Penny's experience of the pencil she is holding, and there is still much within foveal vision that they do not strictly speaking see. But the items we mentioned intuitively *show up* in their phenomenal consciousness.

As Sally and Sam watch these objects, do their visual systems construct representations of all the objects? Our argument that they do not runs as follows.

P1. As S experiences objects o_1–o_{10}, S's visual system maintains a dedicated, detachable representation of each object only if it's *tracking* each object through changes over time.[7]

P2. As S experiences objects o_1–o_{10}, S's visual system is not tracking each object through changes over time.

C. Therefore, as S experiences objects o_1–o_{10}, S's visual system does not maintain a dedicated, detachable representation of each object. (NEUROREPRESENTATIONALISM is false.)

The tracking involved in P1 is exemplified by reidentification: identifying *this* object near the bottom-right corner pocket seen now as the same individual as *that* object near the top-left corner pocket seen a moment ago. The evidence for P2 comes from the multiple object tracking (MOT) paradigm, developed by Pylyshyn and Storm (1988), which seems to show that the visual system is limited to tracking roughly four or five objects at once. Similar limits have been found in other modalities and go back to the start of cognitive psychology (e.g., Miller (1956)). The basic problem is that while you can only visually track four or five objects at once, you can visually experience more objects than this when they occur in slightly larger groups.[8]

That MOT studies show that we see (visually experience) more than the visual system represents is one that Pylyshyn (2007: 120–23) himself considers. Based on MOT results, he tentatively suggests that "perceptual experience [...] need not correspond to some representation that figures in an explanation of how perception or thought works in an information-processing account" (2007: 123). Pylyshyn gets to this conclusion by comparing the

[7] An anonymous reviewer raised a concern that *tracking* is a temporally extended activity, and thus that it cannot be necessary for something (such as a representation) which can be fully present in any one moment of that temporally extended activity. Even if that were so, what is truly essential to P1, as we explain in §3.1, is not tracking over time per se, but the atemporal feature of change sensitivity. We frame things here in terms of temporally extended tracking because, in the human visual system, change sensitivity is realized only by those representations involved in tracking.

[8] We note that there is some controversy concerning exactly what the MOT paradigm demonstrates. While we cannot give a full defence of P2 here, we make the following remark. MOT tasks are standardly taken to provide a behavioral test for whether the properties of an object are being tracked by the visual system, and so test whether the object itself is represented (Scholl et al. 2001: 161). If that interpretation is correct, a failure to complete a given MOT task indicates that the visual system is not representing the targets (Scholl 2001: 32; Scholl et al. 2001: 171–72). If so, the limit to tracking four or five objects in MOT tasks shows that the visual system is capped at representing only four or five objects at a time. The availability of nonstandard interpretations comes down to whether MOT performance is indicative of perceptual processing or instead is indicative of post-perceptual attention or working-memory performance (see, e.g., Block (2007: 489)). While there is more that could be said here, the question of whether MOT performance indicates limits on visual object files is to be settled by further experiments.

sparse visual (object) representations suggested by MOT results to the naïve, 'panoramic' view that visual experience presents entire scenes in rich, full detail. As noted, the naïve 'panoramic' view of visual experience is false. But it is not needed to make our point. Pylyshyn's discussion of this issue is tentative, brief, and opaque. Our aim is to provide this needed discussion.

If it strikes the reader that nobody would seriously maintain neurorepresentationalism about object experience in the face of these considerations, let us emphasize that our challenge is precisely an invitation to those who do appear to be committed to it to clarify their view. In conversations, we have found that many do unreflectively assume that object tracking is necessary for object experience. Once it becomes clear on reflection that this is a substantive assumption, and one that is neither obvious nor demanded by the vision science itself, a more liberal approach comes to seem natural: one that involves giving up on neurorepresentationalism and even minimal forms of neural phenomenal internalism (§5).[9]

3.1 The change-sensitivity requirement

The task of this sub-section is to argue for P1. While we won't attempt to give a complete account of object representation, we will argue that (dedicated, detachable) object representation requires what we'll call *change-sensitivity*. Although change-sensitivity need not take the form of a representation that tracks an object over time, *given the structure of representations in the visual system*, only those visual representations tracking objects over time satisfy change-sensitivity. Hence, the change-sensitivity requirement will establish P1.

To be clear on the dialectic, we will start with the basic notion of representation laid out in §2, at issue in neurorepresentationalism, then argue that, as applied to objects, it entails the change-sensitivity requirement. We are not giving the change-sensitivity requirement as some new, arbitrary, or 'straw man' conception of object representation we think neurorepresentationalists might adopt. Anyone who takes representations to be intentional, detachable signs must accept that object representation requires change-sensitivity.[10]

---

[9] We are not the first to make this specific point about object experience exceeding tracking via object files (or equivalent vehicles). For example, Munton (2021) writes: "object files are neither necessary nor sufficient for seeing an object: we may see objects without object files, as when they feature in the background of a scene but we fail to attend to them, or when we see a larger number of objects than we could maintain separate object files for" (351). Our point is not that everyone makes the assumption we are criticizing here, only that many do, and, specifically, neurorepresentationalists would appear to be committed to doing so given the change-sensitivity requirement we now present (§3.1).

[10] We are not the first to suggest that perceptual object representation has tracking requirements (see, *inter alia*, Burge (2010), Evans (1982: 174–75), and Matthen (2005)). While we agree with these

To set the stage, consider the approach to object representation in the empirical literature that explains it in terms of feature binding and scene segmentation. The hierarchical neural circuits of sensory systems extract features in parallel, so that (e.g.) color is processed in one area and shape in another. The brain needs some way to signal when two extracted features (properties) belong to the same object (Treisman 1998). Similarly, the brain also needs some way of parsing the spatial layout of a scene, including determining the spatial boundaries of objects (Wagemans et al. 2012). When you view nearby medium-sized groups of objects, your visual system engages in feature binding and scene segmentation without necessarily keeping track of the segmented objects through changes over time. We agree that feature binding and scene segmentation are *part* of how perceptual systems represent objects. The question is whether they are *sufficient* for object representation (see also Green (2019)). If we can establish the change-sensitivity requirement, then, because feature binding and scene segmentation do not entail change-sensitivity, we'll have established that (even by neurorepresentationalist's own lights) feature binding and scene segmentation are not sufficient for object representation.

Let's start with the basic conception of representation at issue. Representing an object involves a vehicle, such as neural activity, *standing-in* for the object in a detachable way. To borrow an example from Dretske (1988), imagine we explain a basketball play to you by using pieces of popcorn as stand-ins for the players and move the popcorn pieces around a table to indicate the players' movements. The popcorn pieces become representations of the particular objects at issue—the players—because we are using them as stand-ins for the players, a use which is not dependent on any factual or causal connection between the popcorn and players. To explain what we mean by a 'factual or causal connection', consider what happens if we misremember the play and move the popcorn pieces in ways that fail to reflect the actual movement of the players. In that case, the popcorn ceases to be a natural sign of the players' movements, but this failure to *indicate* the players' actual positions does not affect the status of the popcorn as a stand-in for the players. The popcorn remains a stand-in for the players, but is inaccurate; it *misrepresents* their movements.

Now to change-sensitivity. Consider a token representational vehicle *R*, such as a neural response, spoken word, or piece of popcorn, which purports to represent an object *O*. Further suppose that *R* is used to represent *O* as having property *P*, and *O* does in fact have *P*. So, *R* is accurate. Now suppose that *O* ceases to be *P*, but whatever feature of *R* represented *O* as *P* remains the same. We'll say that *R* is *change-sensitive* if, and only if, under these circumstances it becomes inaccurate. In other words,

---

philosophers that tracking is important for object representation in perceptual systems, we will argue that this follows from a more basic requirement: the need for change-sensitivity.

Change-sensitivity: a purported object representation *R* is change-sensitive with respect to *o* if and only if its (in)accuracy supervenes on *o*'s possessing the features that *R* attributes.[11]

We claim that change-sensitivity is necessary for purported object representations to actually *be* object representations. Consider the popcorn case. As we explained it above, we implicitly assumed that the popcorn pieces were change-sensitive. The spatial movements of the pieces relative to each other is what purports to represent the movements of the players. If the pieces' current movements stop covarying with the players' past movements, the representation, so we said, becomes inaccurate. If it didn't—if there is no change in the (in)accuracy of the representation—then it's just impossible to make sense of the idea that somehow the pieces represent *the players*. We could not make any sense of the idea that (a) the pieces represent the players, (b) the relative present movements of the pieces represent the relative past movements of the players, (c) these relative movements matched up at time *t* but stopping matching up after *t*, and (d) there is no change in the (in)accuracy of the representation at *t*. Since (b), (c), and (d) are all stipulated, it must be (a) that's the problem—a failure of change-sensitivity simply entails that a representation is not representing an object.

Now to show that *representations based on feature binding and scene segmentation need not be change-sensitive*. Suppose some neural spiking in visual circuits is engaging in feature binding: e.g., suppose it encodes the coinstantiation of *red* and *square*. Suppose it is accurate; there is, in fact, a red square appropriately causing the spiking. Now suppose that this original red square turns blue, but, at the exact moment of the color change, a different red square is swapped into its position. The spiking encoding the coinstantiation of *red* and *square* continues along without any interruption. There's no change in the representational vehicle (the spiking). There's also no reason to suppose that this vehicle has suddenly become inaccurate. After all, there is still, in fact, a red square appropriately causing and sustaining the spiking. So, the spiking is not change-sensitive.

What's happening here, we suggest, is that the spiking that performs feature binding is representing the coinstantiation of two properties without representing the object coinstantiating those properties.[12] The point is that change-sensitivity is necessary for object

[11] When we use the term 'change sensitive', we understand it as defined here, as a relation between representation (or reference) and accuracy conditions. It should be distinguished and distanced from the concept of *change blindness* in the psychological literature.

[12] A view on which representing coinstantiated properties suffices for representing the object instantiating them is suggested by Vernazzani (2022).

representation, but neural activity that performs feature binding and scene segmentation is not change-sensitive.

If mere feature binding and scene segmentation is not sufficient for change-sensitivity, how is change-sensitivity established in perceptual processing? This is where *tracking* comes back into the discussion. Consider again the hypothetical spiking responses S1 and S2. What would it take for S1 and S2 to represent not only the coinstantiation of some pair of properties, but to represent the object doing the coinstantiating? Take S1, which signals the presence of a red square and begins (suppose) as a response to object O1. If S1 is change-sensitive, it should become false when O1 transforms into a blue triangle (because S1 itself doesn't change, but continues to signal a red square). If S1 becomes inaccurate as O1 transforms into a blue triangle, that's presumably because S1 still refers to O1 and not O2, the object that started as a blue triangle but is now a red square and causally sustaining S1. Under what conditions would we be prepared to say that S1 continues to refer to O1 despite it now being causally sustained by O2? The only plausible answer is that S1 continues to refer to O1 *if S1 is functioning to track O1*. It could only be the visual system's use of S1 to track O1 which would offer grounds to interpret it as a representation of O1, even though S1 is currently causally sustained by O2.

In sum, we are arguing that the visual system's maintenance of a dedicated, detachable object representation R, referring to an object *o*, entails the change-sensitivity of R with respect to *o*, and this in turn entails that R is functioning to track *o*, as required by P1 of our argument (§3). In vision science, theorizing about this kind of tracking has been dominated by talk of *object files* (e.g., Kahneman et al. (1992); Mitroff et al. (2005); Pylyshyn (2007)). As properties are extracted from the scene, they are not only organized by coinstantiation (feature binding and scene segmentation), but also organized into 'files' that function to track whether different 'feature bundles' extracted at separate times are coinstantiated by the same object. If the above argument is correct, this type of tracking is necessary for change-sensitivity, which is in turn necessary for genuine object representation.

We take this argument to establish P1, but it's worth making one final distinction which may otherwise cause confusion. Note that representing an object must be distinguished from representing something *as an object* or *as falling under some object category* (e.g., *cat*), and also from the representation of the *presence* of a thing falling under some object category. Representations of object categories and of the presence of things falling under them happens in the ventral visual stream (e.g., Bracci et al. (2017); Seeliger et al. (2018)) and is independent of tracking. Object categorization seems to happen fast (in under 200ms), and on the first feed-forward sweep through the ventral stream, before any tracking-related processing. This work on the neural representation of object *categories* does not gainsay P1, because it's not about object representation in the relevant sense. Neural activity involved in object categorization merely *signals the presence* of something from that category; it does not necessarily stand-in for the particular thing instancing that category. Conversely, neural activity could stand-in for an object without attributing object categories like *cat* to it, or without classifying it as an object at all; the activity may only track low-level properties like color and shape.

3.2 Object experience

Earlier in this section we introduced cases of seeing objects in nearby, medium-sized groups as cases in which you experience objects not represented by your sensory systems. In §3.1 we argued that, in these cases, not all of the objects in the group are represented by your sensory systems. What is left is to argue that you *do experience* all the objects (or, at least, more than are being tracked in the technical sense).

To start, some clarity is needed on what counts as an *object experience*. Return to the example from §1. Pick up a nearby object. Hold it. Look straight at it. Listen and smell it, too. We take this to be an uncontroversial, paradigm case of an object experience in the everyday, intuitive sense. Our question now simply concerns the *range* of these experiences. What about when you catch a fleeting glimpse of a fast-moving object out of the corner of your eye? Do you experience the object? How about when an object far in the distance appears as only a dot on the horizon? Does it 'show up' in phenomenal consciousness? Suppose you look in a cluttered dish for your keys; the keys are right there on the top, in plain sight, but you still miss them. Are you experiencing your keys? What about when you look at a large tree with thousands of leaves—do you experience each leaf?

One might worry that our argument will force us to loosen the notion of 'object experience' so much that it's trivially or at least uninterestingly true that we experience objects the brain doesn't represent. If the notion of 'object experience' was so loose that fleeting, peripheral glimpses of fast-moving objects afforded experience of those objects, one should suspect that our conclusion only follows because we have questionably ratcheted down the standards for object experience. After all, what makes the claim (that we experience objects not neurally represented) interesting is the appeal to *typical* cases of object experience. What we're suggesting is that something roughly like the paradigm object experience of a pencil you're holding and attentively looking at is possible without your brain constructing a representation of it.

To alleviate the worry, the first point to make is that we are focusing on quite specific cases and making quite specific claims. Consider again the cases involving Sally and Sam. It's important to keep in mind that the claim is not that they experience each item in vivid and exhaustive detail, or that they see the entire scene falling in their field of view. Sally's experiencing each item of stationery is compatible with her experiencing relatively few of the sensible properties (e.g., shape or color) and parts of each item. So, again, the reader should not think that we are advocating for a 'panorama' view of visual experience.

The experiences had while viewing nearby medium-sized groups of objects are much more like paradigm cases of object perception than they are like the questionable cases just mentioned. Consider the missed-keys case. In this case, your keys, despite being in plain sight, are *not* available for voluntary attention. That's the sense in which you 'miss' them. In contrast, when you view nearby medium-sized groups of objects, although there will be too many objects for you to voluntarily attend to each one simultaneously, each object is immediately *available* to voluntary attention. As the pool balls scatter, Sam is not able to attend to each ball simultaneously, but he can attend to any one of them at any point. A plausible although by no means universally accepted view is that what's available for voluntary attention is what's experienced (Scholl (2001: 20); Pylyshyn (2007: 59); Levine (2010: 181); Dickie (2010: 216); Wu (2011: 109)). So, if each object in the group is immediately available for voluntary attention (in a sense that of course requires precisification), then each is experienced.

More generally, the questionable cases all have features which make them questionable. Viewing nearby medium-sized groups of objects lacks all of these questionable features. Unlike the peripheral-glimpse case, the items of stationery all fall within Sally's foveal vision (at least over a couple of seconds of scanning). Unlike the distant-object case, the pens and pencils etc. are close enough that at least some specific information from each is available. Unlike the tree case, there aren't so many items so as to obviously overload visual processing; Sally's visual system still does successful scene segmentation, discriminating the items from each other and from the background, and still binds detected features (properties) belonging to each item. This hints at what we'll propose in §5. Experience of objects depends on a certain kind of sensory activity that registers information about those objects; it does not require full-blown, detachable representation of each object.

3.3 On seeing the forest but not (all of) the trees

It might be asked whether paradigm cases of object experience (i.e., Penny's focused experience of the pencil she is holding) are the best comparison for viewing objects in nearby, medium-sized groups. Granted, perhaps experiences of nearby groups are more like paradigm object experiences than experiences of huge groups or far distant objects, but maybe they're even more like experiences of *textures*. Perhaps, as Sally views the cluster of stationery on her desk, she experiences—in addition to the subset for which she does maintain object-files—*ensemble properties* of the cluster, *not* each item.

This sort of argument to the effect that we see far fewer things than we ordinarily take ourselves to see might be motivated by a review of phenomena such as *crowding* or *ensemble perception*. We will question whether the vision science should be interpreted as showing that Sally *does not perceive* 8-10 items rather than only around 4-5. Before that, however, notice that this way of arguing is in any case tendentious. Even if it were true that some of the information-carrying vehicles Sally or her visual system has, e.g., of the eraser, are fairly sparse, why should someone who does not antecedently endorse neurorepresentationalism accept that it follows that she does not *experience* the eraser but at best only some of its features? One need not defend a naïve, 'panorama' conception of vision in order to put up some resistance to this way of arguing about the scope of object experience.

Let us first examine the phenomenon of visual crowding and consider whether it might be thought to cast doubt on our claim that Sally genuinely sees all of the items of stationery on her desk. Crowding is the ubiquitous, deleterious effect of nearby objects on visual discrimination, principally outside of foveal vision, impairing our capacity to recognize or identify (but not detect) objects within a clutter (see, e.g., Levi (2008)). But while it has sometimes been thought that perception of crowded objects is ultimately akin to texture perception of objects' low-level features rather than objects *per se*, it is now thought that object perception often survives crowding due to the preservation of high-level feature representations (e.g., faces). As Fischer and Whitney (2011) put it, "crowding does not irreversibly dismantle or destroy the 'objectness' of an object" (1397).

The related (and even more ubiquitous) phenomenon of ensemble perception refers to the perceptual representation of summary statistical information of the features of sets of similar items. 'Ensemble' here derives from the literature on texture perception, which one might think of as a kind of non-distributive plural representation, so that one perceives a plurality of things without perceiving each, or perhaps any, individually. Ariely's (2001) landmark study showed that while subjects were at chance in correctly answering whether a particular circle was present in a cluster of around 16 similar circles (observed for 500 ms), they were remarkably accurate in correctly answering whether it was smaller or larger than the mean size of the circles presented in the cluster. There is good reason to think this calculation of the mean, or other statistical properties such as variance (Haberman et al., 2015), occurs perceptually and without the need for intermediate representations of individual items (Whitney and Yamanashi Leib 2018: 115–117). With a subsequent surge of interest in ensemble perception, it has since been suggested that

> Gist-based ensemble perception is the foundation of all sensory processing […]. […] ensemble properties are the building blocks of perception, permitting the majority

> of the world to be rapidly interpreted, only later to be investigated in detail by more effortful, explicit, focused-attentional processing (Corbett et al. 2023: 53).

Upon reflecting on this work, one might wonder whether there is a natural "dichotomy" here, between "rapid, *vision-at-a-glance*" ensemble perception and "slower, *vision-with-scrutiny*" object perception via tracking with an object-file (Corbett et al. 2023: 54). If that were the case, one might have grounds to insist that Sam must only perceive 4 or 5 items, with the other items at best contributing to ensemble representations.

In reply, we first think it is worth questioning the idea that vision is 'dichotomous' in this sense. There is a more conservative and open-minded 'middle view' available on which object experience is grounded in a plurality of different neural vehicles rather than only one, or on which there are more than two perceptual 'styles' more than one of which is compatible with object experience. On this middle view, one might experience objects without rich and demanding tracking via a dedicated object file. In short, while ensemble perception has been underexplored in the philosophical literature, any inference to a robust dichotomy of the sort that might be the basis for objection here is at best speculative as things stand.[13]

Moreover, even if it were true that vision is strictly 'dichotomous' in this sense, it is simply not clear why it should follow that we do not experience objects for which we lack dedicated object-files; that is, it is not clear why ensemble representations could not, in the right conditions, with the right flow of information from the distal stimulus object in question, constitute *object experience*. Whitney and Yamanashi Leib (2018) appear to acknowledge this interpretation when they suggest it might be that, in much of ensemble perception, "individual objects are phenomenally available to consciousness but unreportable" (120; see also McLelland and Bayne (2016)). At the very least, then, we think that nothing in the vision science itself can help the neurorepresentationalist out of the argument presented in this section.

4 Object experience that survives multimodal binding failures

Multimodal feature binding occurs when one's sensory systems represent properties detected by distinct modalities as coinstantiated by the same object. For example, when Penny views the pencil she is holding, she does not have two independent experiences, a haptic experience of hardness and a visual experience of color and shape. She has an integrated, visuohaptic experience of a single object coinstantiating these properties. In the case of sounds, you often have an audiovisual experience of a seen object producing a sound. For example, when listening

[13] We are grateful to Joan Danielle Ongchoco for comments on this material.

to someone you can see speak, you experience the sounds as coming from the person you see.[14] As O'Callaghan (2019) argues, multimodal binding happens within sensory processing itself (it is not post-perceptual) and results in genuinely multimodal experiences.

Imagine watching a steel ball bearing roll across a wooden table. The bearing makes a loud, distinctive sound as it rolls. In this sort of case, we take it as uncontroversial that you *see*, i.e., *visually experience*, the bearing. We also take it as uncontroversial that you *hear*, i.e., *auditorily experience*, the sound produced by the bearing. However, as you watch the bearing in motion, you not only experience a sound as produced by a seen object, you also introspectively seem to *hear a seen object:* you have an audiovisual experience of it. The sound conveys information to you about the bearing, just as reflected light conveys information about it, and you experience the bearing through that auditory information link. Although you don't experience seen objects as instancing auditory properties like pitch, some properties are accessible via both vision and audition. For example, just as shape can be both seen and felt, temporal duration and timing can be both seen and heard (Young 2018). And when you experience a typical physical object like a rolling bearing, you might experience the rolling bearing as coinstantiating a seen shape and a heard acceleration.

The above case involves *successful* audiovisual binding. But this binding can fail. In off-track films, for example, there's a failure to bind visual and audio streams. Sometimes, though, as when watching a ventriloquist performance, you incorrectly bind visual and audio streams originating from distinct objects. Suppose you're watching what looks like a steel ball bearing roll down a slightly off-level table. The table is divided by a curtain that runs parallel to the rolling ball. The ball you see rolling is actually a well-painted foam rubber ball, completely silent as it rolls. At the moment this rubber ball was released, a real steel bearing was released behind the curtain, out of sight. You hear this bearing roll (including its acceleration) and your sensory systems misbind its sound with the seen rubber ball. You don't merely incorrectly hear the *sound* as produced by the rubber ball, you incorrectly experience the rubber ball as instantiating the *heard acceleration*. This is a case in which your sensory systems wrongly treat your auditory and visual information channels as originating from the same object.

From this setup, an argument for object experience without object representation can be run as follows.[15]

P1*. Your sensory systems have opened a single object file tracking what these systems wrongly presuppose to be a single object through both the auditory and visual channel.

[14] This integration is illustrated by the so-called 'McGurk effect' (McGurk and MacDonald 1976).

[15] For discussion of a case of unimodal binding failure that could be used to make the same point, see Openshaw & Weksler (2020).

P2*. Whatever this object file represents, it can at best represent (successfully) either the bearing or the rubber ball (not somehow both).

P3*. You experience *both* the heard steel bearing and the seen rubber ball.

C*. Your sensory systems lack a representation for either the bearing you auditorily experience, or the ball you visually experience. (NEUROREPRESENTATIONALISM is false.)

The argument is valid, so let us take the premises in turn.

P1* is supported by the deep audiovisual feature binding. It is possible to have binding without an object file. After all, color-shape binding happens when watching a nearby medium-sized group of objects, even without object files being opened (or so we argued above). But if an object file is opened, then bound extracted properties presuppose a single object file. In this case, since you're focused on tracking just one thing (the 'ball'), your sensory systems have presumably opened an object file. The audiovisual binding suggests that this object file is being fed with information from both the auditory and visual streams; hence, your sensory systems aren't treating these information streams as originating from distinct objects, and so (presumably) don't have separate auditory and visual files. There is, perhaps, a separate file for the *sound*, but not one file for the *heard ball* and another file for the *seen ball*.

One way to push back against this argument for P1* is to adapt the account of file 'linking' given by Recanati (2012: 42–47) to the multimodal case. On this suggestion, there are indeed separate auditory and visual files (for, respectively, the heard and seen balls). The multimodal binding is explained by a linking operation performed by your sensory systems which leads them to treat the two files as concerning the same object, without fully merging the files. Recanati posits linking as a post-perceptual cognitive operation, primarily as a way to explain what happens when you learn, for example, that Diana Prince is Wonder Woman.

However, if there is a meaningful distinction between *linking* auditory and visual files, and simply having *one* audiovisual file, it should somehow show up behaviorally. For example, psychologists measure *object-specific preview effects* (e.g., Zmigrod et al. (2009); cited by O'Callaghan (2019: 63)). When you first show a subject a visual property *V* paired with an auditory property *A*, that first presentation slows down responses to subsequent presentations of *V* that are paired with a different auditory property. This preview effect is standardly interpreted as a sign of a single file collecting both the visual and auditory information (see O'Callaghan (2019)). A proponent of perceptual linking would need to explain why this sort of effect is really (or sometimes) a sign of mere file linking, and offer some way empirically to distinguish linked auditory and visual files from a unitary audiovisual file.

A different set of worries instead target the success condition in P2*. First, perhaps the object file 'successfully' represents some unexpected or nonstandard object, e.g. a mereological fusion of the bearing and ball or a Meinongian nonexistent object. However, this simply doesn't help the neurorepresentationalist's case. Consider P3*. The neurorepresentationalist cannot explain how you perceptually experience *two* things by appealing to a neural representation of *one distinct* thing. By the neurorepresentationalist's lights, this wrongly implies that you experience *something other than* the bearing and the ball.

A better objection to P2* is to question whether successful reference requires referential *determinacy*. On this view, it is enough for you to perceptually experience both the bearing and the ball that you have an object file which is indeterminate in reference between the bearing and the ball. The object file determinately refers to either the bearing or to the ball, but it is indeterminate which. In response, however, if the neurorepresentationalist makes this move, they are committed to saying that P3* is determinately false. For it seems to follow that you determinately *experience either* the ball *or* bearing, despite it being indeterminate which. On either precisification, then, you fail to experience *both* the ball *and* the bearing. Given the determinate truth of P3*, then, P2* follows from the nature of object representations.

Is P3* true? We assume no one will dispute that you experience the rubber ball. You can clearly see it and, so, have a visual experience of it. What about the bearing behind the curtain? Colloquially, it's fine to say that you 'hear the bearing', but you might worry this everyday way of talking is misleading.

Clearly, you hear a *sound*, which is a distinct entity from the bearing producing it. As noted above, this sound gives access to certain properties of the bearing (e.g., its acceleration) and our sensory systems try to attribute these properties not to sounds, but to physical objects. We take this sound-mediated property attribution to be good *prima facie* evidence that we can hear more than sounds, i.e., we can hear the objects that produce sounds. Something similar goes for touch. As Penny holds the pencil, she will feel tactile sensations, but these tactile sensations afford her a further haptic experience of the pencil itself (Matthen 2021). While the case of vision is much more controversial, the same dual structure might be there as well. We experience distinctly visual 'sensations' that afford visual experience of objects (Noë 2004).

Suppose you were just watching the actual bearing roll (the curtain and rubber ball are removed). In such a case there's no good reason to deny that you hear the bearing roll.[16] You

[16] Young (2018: 2937; 2945) argues that we hear objects (e.g., the rolling bearing), not just events (e.g., the bearing's rolling), through hearing sounds. Our claim is that this still happens when there is an audiovisual binding failure.

seem to be making auditory contact with the bearing. Now, add the curtain back, so you can no longer see the bearing. If you were in auditory contact with the bearing in the first case, you don't suddenly lose that auditory contact just because you can no longer see the bearing. Add back the rubber ball. Does the resulting illusory binding of the sound with the seen rubber ball somehow break your auditory contact with the actual sound producer (the bearing)? At this point you might appeal to the misbinding, and the resulting single representation, as showing you have lost auditory contact with the bearing. (In effect, you might accept premises P1* and P2* and say that they entail P3* is false.) But this move begs the question. It presupposes that auditory contact (i.e., auditory experience) requires object representation.

To motivate the idea that misbinding on its own does not break perceptual contact with the actual source of a modality's information stream, consider the rubber hand illusion (Botvinick and Cohen 1998), which also (on standard interpretations) involves misbinding. Your (mis)experience of the seen brush strokes as the felt brush strokes does not imply you no longer feel *your real hand* being stroked. At best, it implies you misidentify the rubber hand as your hand. You experience your real hand, visually experience the rubber hand, and wrongly experience them as being the same hand. Similarly, just because you (mis)experience the seen rubber ball as having the heard acceleration does not mean you don't hear the bearing accelerate. You visually experience the rubber ball, auditorily experience the bearing, and wrongly (multimodally) experience them as one.

A final potential objection is that the bearing occluded by the curtain merely causes us to misexperience the seen ball, but *is not itself experienced*. By analogy, consider a perfectly camouflaged moth located on a tree trunk. It is tempting to say that the moth itself is unseen. Yet if the moth happens to cover the only white spot on the trunk, it will cause us to misexperience its bark as uniformly brown.[17] There are two natural replies to this objection. One line of reply is to insist on a difference between the two cases. Though we have an information link to the moth, we aren't positioned to exploit it to learn anything about it. In contrast, the auditory information link to the bearing does position us to learn about the bearing. A different line of reply is that a perfectly camouflaged moth, assuming it is sufficiently large and that it falls within foveal vision, etc., *is* seen, even if it is not recognized as a moth (see also Cohen (2019: 721–722)). Of course, there is a sense in which one fails to see a moth (as Siegel (2006: 434) notes). But there is also arguably a sense in which one sees the moth but simply fails to see it *as* a moth. On this second line of reply, it is not obvious that there *are* cases in which an object causes us to misexperience another seen object without itself being seen, and so this analysis of the multimodal binding failure case we have considered in this section is not independently plausible. The authors of this paper are inclined to take

[17] We thank an anonymous reviewer for prompting us to address this example.

different replies here, though we are in agreement that both are satisfying ways to defuse the objection and put serious dialectical pressure on the neurorepresentationalist.

5 Perception and the flow of information

The upshot, from the above cases, is that object representation in neural sensory systems is relatively demanding, rare, and fragile, while object experience is relatively easy, pervasive, and robust. Object representation is demanding. In perceptual systems it requires the construction of object files which track changes in an object's perceived properties over time. It's rare insofar as we have a limited capacity for it, especially relative to our capacity for scene segmentation and feature binding. It's fragile insofar as misbinding quickly breaks reference to objects, even when we otherwise have a usable sensory information channel to those objects.

In contrast, object experience is easy—even when considering only cases similar to unquestioned, paradigm cases of object experience. We achieve it just in virtue of rudimentary operations like feature binding and scene segmentation. It's pervasive. Much of our visual, auditory, and tactile experiences are organized around objects. It's robust, insofar as it survives processing failures like misbinding which otherwise derail object representation.

By detaching object experience from neural object representation, we are not suggesting that there is no physical explanation for how we manage to experience objects. When Sally sees each of the items of stationery in the cluster on her desk, there is still plenty of representation-mediated sensory processing happening. Her visual system detects various properties through the process of feature extraction and binds coinstantiated instances. It lacks the capacity to do the sort of tracking necessary to *represent* each item *per se*, but nonetheless still manages to establish a *usable flow of information* from each item. This flow of information is *usable* in the sense that it allows Sally to *learn about* the items. In that sense, this flow of information *reveals* the pens, pencils, and so forth to her, i.e. it allows her to acquire information concerning their properties.

Perception can be a conduit through which we learn about distal objects without itself representing those objects. Of course, to form the belief that an object you perceive has property *P*, you must have a *belief* which represents that object as having *P*. But there's no reason to think that having a post-perceptual (e.g., doxastic) object file on a perceived object requires first having a sensory file on that object. For example, while my audiovisual system is fooled by the misbound sound of the steel bearing, once one knows of the rigged setup, conceptually mediated and isotropic cognitive mechanisms are easily able to open separate files for the seen ball and heard bearing. Cognition can use its conceptual representations and variable inputs to sort through the mess of failed sensory representations.

When does a sensory system provide a 'usable flow of information'? For example, processing in the dorsal visual stream seems to provide a usable flow of information (e.g., it's usable for guiding action), but most people don't think this processing leads to visual experience of anything, including objects (see Milner and Goodale (2006)). Likewise, when you see a far-off flock of hundreds of birds, you don't seem to experience each bird. What makes that flow of information unusable?

We openly grant that we have not offered a complete, *positive* theory of perceptual experience. That would require specifying just which sensory information processing is 'usable' in the way necessary and sufficient for perceptual experience. That is the task for a separate paper—indeed a separate research programme—which we can only motivate rather than undertake here. But note that the neurorepresentationalist is no better off. Not all representations constructed within our sensory systems become conscious (become experiences). So the neurorepresentationalist needs an account of what makes the difference between the conscious and unconscious sensory representations. There are many candidate theories for what makes this difference, e.g. global workspace theory (Baars 2002; Dehaene et al. 2014), information integration theory (Tononi 2008), recurrent processing theory (Lamme 2010), PANIC theory (Tye 1995), AIR theory (Prinz 2012), and higher-order representation theories (Lycan 1996; Carruthers 2000). These theories don't give neurorepresentationalism an edge over our proposal, for two reasons. First, they all face serious difficulties and have attracted limited support. Second, many of them aren't obviously theories of what makes a *representation* conscious, as opposed to what makes *information encoding* or *processing* conscious. Hence, it would be easy enough for us to adopt many of these accounts as an explanation of what gives a flow of information the sort of 'usability' sufficient, or necessary, for experience.

Before concluding, let's return to the issue of minimal neural phenomenal internalism. It is, of course, open for neurorepresentationalists to back off commitments to *representations*, as in intentional, detachable signs, and instead reframe their position in terms of mere information encoding or (as we've put it) the flow of information. But, this move would (i) concede to us everything of importance in this debate, and (ii) force neurorepresentationalists to give up their desired and motivating commitments to minimal neural phenomenal internalism. Given their motivating commitments to minimal neural phenomenal internalism, we don't think it's plausible for a neurorepresentationalist to claim they only ever 'really' meant to talk about mere information encoding. At the very least, this claim requires admitting that they started with a confused position that failed to recognize the need for intentional, detachable signs to secure minimal neural phenomenal internalism.

6 Conclusion

We have argued that object perception does not require object representation in our neural sensory systems. Specifically, it's possible to experience an object with which you're interacting through your sensory systems even if your sensory systems have not constructed a representation of the object. Fundamentally, object perception happens when an object is revealed through the flow of sensory information. Often an object is revealed by constructing a representation of it, but this is at best only one way sensory perception can happen. When objects are revealed without being represented, the experience won't be reproduced merely by reproducing the electrochemical and functional state of the brain. While arguing for this view, we have also argued that tracking via object files is necessary for object representation within neural sensory systems.

*

Acknowledgements: This paper developed out of chapter 4 of Barkasi's 2015 dissertation. A few paragraphs may survive unchanged, but it's almost entirely a different paper. Another early version of this material was presented at the 2019 APA Central meeting. Its commentator, Brian Cutter, and the audience provided helpful discussion. This project has substantially benefited over its years of development from many discussions with Casey O'Callaghan and Mohan Matthen. They may not agree, but their ideas substantially shaped this work. Helpful feedback on this material was also received from Richard Grandy, Indrek Reiland, Dan Burnston, Alex Morgan, Rami El Ali, Thomas Raleigh, and Alfredo Vernazzani. Finally, we are very grateful to the two reviewers from this journal for their insightful comments.

Funding: James Openshaw's work on this project was in part funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101032391. Michael Barkasi has no funding to declare.

Bibliography

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science* 12(2), 157–162.

Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Science* 6(1), 47–52.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 30, 481–548.

Botvinick, M. and J. Cohen (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391, 756.

Bracci, S., Ritchie, J. B., and H. Op de Beeck (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia* 105, 153–164.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence* 47(1–3), 139–59.

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

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

Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Cohen, J. (2019). Schellenberg on perceptual capacities. *Analysis* 79(4), 720–730.

Corbett, J. E., Utochkin, I., and S. Hochstein. (2023). *The Pervasiveness of Ensemble Perception*. Cambridge: Cambridge University Press.

Crick, F., and C. Koch (1995). Are we aware of neural activity in primary visual cortex? *Nature* 375, 121–123.

Crick, F., and C. Koch (1998). Consciousness and neuroscience. *Cerebral Cortex* 8, 97–107.

Dehaene, S., L. Charles, J.-R. King, and S. Marti (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology* 25, 76–84.

Dickie, I. (2010). We are acquainted with ordinary things. In R. Jeshion (Ed.), *New Essays on Singular Thought*, pp. 213–245. Oxford: Oxford University Press.

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

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: The MIT Press.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge: The MIT Press.

Dretske, F. (2003). Experience as representation. *Philosophical Issues* 13(1), 67–82.

Evans, G. (1982). *Varieties of Reference*. Oxford: Oxford University Press.

Fischer, J., and D. Whitney. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology* 106, 1389–1398.

Fish, W. (2009). *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. Oxford: Oxford University Press.

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

Grice, H. P. (1957). Meaning. *Philosophical Review* 66(3), 377–388. Page citations to the reprint in Grice 1989, pp. 213–223.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology* 10(1), 5–23.

Haberman, J., Lee, P., and D. Whitney. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision* 15(4): 1–11.

Hutto, D. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge, MA: MIT Press.

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

Koch, C., Massimini, M., Boly, M., and G. Tononi. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews* 17, 307–322.

Lamme, V. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience* 1(3), 204–220.

Lee, J. (2019). Structural representation and the two problems of content. *Mind and Language* 34, 606–626.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: a mini-review. *Vision Research* 48(5): 635–654.

Levine, J. (2010). Demonstrative thought. *Mind and Language* 25, 169–195.

Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge: The MIT Press.

Martin, M. G. F. (2002). The transparency of experience. *Mind & Language* 17(4): 376–425.

Martínez, M., and B. Nanay. (forthcoming). Many-to-one intentionalism. *Journal of Philosophy*.

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

Matthen, M. (2021). Dual structure of touch: The body vs. peripersonal space. In F. de Vignemont, A. Serino, H. Y. Wong, and A. Farnè (Eds.), *The World at Our Fingertips: A Multidisciplinary Exploration of Peripersonal Space*. Oxford: Oxford University Press.

McGurk, H., and J. MacDonald. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.

McLelland, T., and T. Bayne. (2016). Ensemble coding and two conceptions of perceptual sparsity. *Trends in Cognitive Sciences* 20(9): 641–642.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2), 81–97.

Milner, A. D., and M. A. Goodale (2006). *The Visual Brain in Action (2 ed.).* Oxford: Oxford University Press.

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

Moutoussis, K. (2016). The machine behind the stage: A neurobiological approach toward theoretical issues of sensory perception. *Frontiers in Psychology* 7(1357), 1–9.

Munton, J. (2021). How to see invisible objects. *Noûs* 56, 343–365.

Neander, K. (2017). *A Mark of the Mental*. Cambridge: The MIT Press.

Noë, A. (2004). *Action in Perception*. Cambridge: The MIT Press.

O'Callaghan, C. (2019). *A Multisensory Philosophy of Perception*. Oxford: Oxford University Press.

Openshaw, J., and A. Weksler. (2020). A puzzle about *seeing* for representationalism. *Philosophical Studies* 177, 2625–2646.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch and B. Lloyd (Eds.), *Cognition and Categorization*, pp. 259–302. Hillsdale: Lawrence Elbaum Associates.

Prinz, J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition* 9(2), 243–259.

Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. New York: Oxford University Press.

Pylyshyn, Z. W. (2007). *Things and Places: How the Mind Connects with the World*. Cambridge: The MIT Press.

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

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

Revonsuo, A. (1995). Consciousness, dreams and virtual reality. *Philosophical Psychology* 8(1), 35–58.

Schellenberg, S. (2018). *The Unity of Perception: Consciousness, Content, Evidence*. Oxford: Oxford University Press.

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

Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa)? In D. Dedrick and L. Trick (Eds.), *Computation, Cognition, and Pylyshyn*, pp. 49–78. Cambridge: The MIT Press.

Scholl, B. J., Pylyshyn, Z. W., and J. Feldman (2001). What is a visual object? evidence from target merging in multiple object tracking. *Cognition* 80, 159–177.

Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J.-M., Bosch, S., and M. van Gerven (2018). Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage* 180, 253–266.

Siegel, S. (2006). How does visual phenomenology constrain object-seeing? *Australasian Journal of Philosophy* 84(3), 429–441.

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

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin* 215(3), 216–242.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences* 353, 1295–1306.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind.* Cambridge: MIT Press.

Vernazzani, A. (2022). Do we see facts? *Mind & Language* 37(4), 674–693.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., and R. von der Heydt (2012). A century of gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin* 138(6), 1172–1217.

Whitney, D., and A. Yamanashi Leib. (2018). Ensemble perception. *Annual Review of Psychology* 69(1), 105–129.

Wu, W. (2011). Attention as selection for action. In C. Mole, D. Smithies, and W. Wu (Eds.), *Attention: Philosophical & Psychological Essays*, pp. 97–116. Oxford: Oxford University Press.

Young, N. (2018). Hearing objects and events. *Philosophical Studies* 175, 2931–2950.

Zmigrod, S., Spapé, M., and B. Hommel (2009). Intermodal event files: Integrating features across vision, audition, taction, and action. *Psychological Research* 73(5), 674–684.