Rev.Phil.Psych. (2016) 7:721–742
DOI 10.1007/s13164-015-0294-8

Opening Up Vision: The Case Against Encapsulation

Ryan Ogilvie[1] · **Peter Carruthers**[1]

Published online: 27 November 2015

Abstract Many have argued that early visual processing is encapsulated from the influence of higher-level goals, expectations, and knowledge of the world. (Early vision is thought to result in perception of three-dimensional shapes and surfaces, prior to object recognition and categorization.) Here we confront the main arguments offered in support of such a view, showing that they are unpersuasive. We also present evidence of top–down influences on early vision, emphasizing data from cognitive neuroscience. Our conclusion is that encapsulation is not a defining feature of visual processing. But we take this conclusion to be quite modest in scope, readily incorporated into mainstream vision science.

1 On Encapsulation

According to Fodor (1983), Pylyshyn (1999), and others, early perceptual processes are incapable of receiving input from, or of being directly influenced by, cognitive processes of categorization, thought, and expectation. In short, early vision is *encapsulated* from cognition. In many ways, Fodor (1983) laid the framework for subsequent debate. He clearly articulated a variety of properties said to be possessed by perceptual modules (encapsulation being the most important), and pulled together a number of empirical and theoretical considerations in favor of the idea that perceptual (and linguistic) systems have these properties. While Fodor's claim that perceptual systems are modular in this sense offered a coherent and plausible research program at the time, and was adopted and defended by a wide range of researchers, our view is that (in the case of vision, at least) enough is now known for it to be rejected.

It is worth noting that others besides Fodor have defended the modularity of mind, and/or the usefulness of a notion of modularity for cognitive science, but have employed a weaker notion than Fodor's, one that doesn't require encapsulation (Barrett and Kurzban 2006; Carruthers 2006). This is important because some in the

✉ Peter Carruthers
pcarruth@umd.edu

[1] Department of Philosophy, University of Maryland, Skinner Building, College Park, MD 20742, USA

field seem to assume that if vision isn't modular in Fodor's sense then one cannot individuate the visual system, nor draw a distinction between vision and cognition (Clark 2013; Firestone and Scholl forthcoming). This is a mistake. One can perfectly well characterize the visual system functionally, as the set of brain-mechanisms specialized for the analysis of signals originating from the retina. The computations that the visual system performs are geared towards making sense out of the incoming light array. But one can accept that the visual system consists of a proprietary set of mechanisms while denying that it takes only bottom–up input. Indeed, this is the combination of views we proposed to defend.

Pylyshyn (1999) provides a full-scale defense of the encapsulation of vision. His central claim is that while some processes late in the processing hierarchy are influenced by cognitive states, those that constitute "early vision" (functionally defined as the processes involved in constructing a geometrical description of the scene) are not. He argues that all alleged evidence for cognitive penetrability can be explained in one of three ways, none of which constitute bona fide cognitive penetration: (1) intra-modular top–down effects (effects on lower-level processes that arise from processes later within the modular processing hierarchy itself), (2) top–down effects that target post-perceptual or post-modular systems, and (3) effects of overt and covert spatial attention on perceptual processing, where shifts of attention merely alter the *input* to visual processing, not the *manner* of that processing. In addition, most would also claim that *inter*-modular associations (such as those between vision and audition that result in the McGurk effect) are just as consistent with encapsulation as intra-modular ones.

According to Pylyshyn, genuine cases of cognitive penetration should involve semantically-relevant effects of post-perceptual cognitive states on the way in which early visual processing is conducted. As he puts it, "For present purposes it is enough to say that if a system is cognitively penetrable then the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs" (1999, 343). Since all apparent cases of cognitive penetration can be explained within his modular framework, Pylyshyn claims, there is no good reason to think—and many good reasons for *not* thinking—that early vision is cognitively penetrable.

Given the way Pylyshyn fleshes out the details of (1)–(3), we are in broad agreement that these sorts of effects don't count as instances of cognitive penetration. Our agreement rests on some subtle empirical and conceptual details, however, such as how one should precisely characterize early vision, what counts as visual as opposed to cognitive processing, and how to characterize the operations of attention. But for present purposes we propose to set these issues aside, and will lay out evidence of cognitive penetration of a sort that Pylyshyn thinks is impossible. Before we consider the evidence, however, we will briefly situate our approach with respect to previous critiques of encapsulation.

The philosophical literature contains a number of challenges to the encapsulation of visual processing. Some of these challenges comprise attacks on the theory-neutrality of observation (Hanson 1965; Churchland 1988). Others take aim at classical modularity (Churchland et al. 1994; Prinz 2006; Brogaard et al., 2014). And some have argued for the cognitive penetrability of vision more directly, by focusing on a specific set of studies purporting to show, for example, that beliefs and expectations modulate color discrimination (Macpherson, 2012), that conative states modulate spatial

perception (Stokes 2011; Wu 2013), or that perception of vague or ambiguous stimuli relies on background knowledge (Churchland 1988; Brewer and Loschky 2005). This isn't, of course, an exhaustive review, and some authors straddle these different categories.

One commonality that runs through these philosophical critiques, however, is that the evidence appealed to derives from purely-behavioral studies. There are two problems with relying exclusively on such evidence. First, interpreting behavioral data as evidence of cognitive penetration requires operationalizing what counts as *perception*. But philosophers and psychologists don't always have the same things in mind when they talk about perception (e.g. conscious experience versus functionally specified representations of a certain sort). Moreover, what *should* count as perception is notoriously up for grabs. We suspect that at least some disputes boil down to differing views about where to draw the line between perception and cognition. The second problem is that, even if there is agreement about how to draw this distinction, it is hard to construct behavioral experiments that control for it.[1] Firestone and Scholl (2014, 2015, forthcoming) have cast serious doubt on large swaths of behavioral data by arguing that the studies in question fail to control for one or another of the factors mentioned above (i.e., the factors initially identified by Pylyshyn: intra-modular effects, post-perceptual effects, and attentional effects).[2]

Our focus, in contrast, is on evidence from cognitive neuroscience, including single-cell recordings, EEG, and fMRI evidence (generally combined with behavioral data, of course). Multivariate pattern analysis of fMRI data, in particular, is an increasingly used—and perhaps transformative—tool in neuroscience and psychology. Properly employed, it can enable researchers to "read the contents" encoded in specific brain regions at a given time. The method is not, however, free of controversy and weaknesses. (For reviews of the pitfalls, and strategies for avoiding them, see Davis and Poldrack 2013; Haynes 2015.) We have tried to confine ourselves to studies that are well-regarded in the field, however, and the evidence cited is consistent (as we will see) with converging findings from single-cell recordings, EEG, and computational modelling, as well as other forms of fMRI evidence. Importantly, the brain-imaging and measurement techniques we discuss are always paired with standard behavioral paradigms. Thus, brain data does not displace behavioral data, but rather supplements it. One of our goals is to find greater consilience across a variety of experimental paradigms and disciplines.

The studies we discuss build on extensive prior knowledge of the nature and internal organization of the visual system, most of it collected using standard bottom–up paradigms. We know that signals from the retina are initially projected (primarily via the lateral geniculate nucleus) to area V1 at the back of the brain, and that these signals are fed forward to be processed for information about contour, orientation, binocular disparity, color, and motion in extrastriate areas (primarily V2, V3, V4, and V5/MT). Much is known about the distinctive properties of the neurons in these regions, their retinotopic organization, their receptive field-sizes, and their computational properties.

[1] See Masrour et al. (2015) on the methodological limitations of behavioral evidence for resolving the encapsulation debate.

[2] Firestone and Scholl (forthcoming) identify six pitfalls that befall studies purporting to show top-down effects. We think that each of the six confounds is a special case of the three discussed by Pylyshyn (1999).

We know enough, in fact, to know that this network of regions serves to realize the early visual system. Evidence that the representational properties of these regions are modulated top–down by external cognitive factors will thus be evidence that early visual processing is not encapsulated. That is what we propose to argue, beginning in Section 3. In Section 2, however, we first review and critique the main arguments that have been put forward in support of encapsulation, focusing on the work of Fodor (1983) and Pylyshyn (1999).[3]

2 Arguments for Encapsulation

Why might one *expect* perceptual systems in general (and early vision in particular) to be encapsulated? A number of arguments have been offered. Fodor (1983) emphasizes the computational benefits of encapsulation. By accessing only a limited (module-internal) body of information the system can process its input much more swiftly than if it had to consult the full extent of the agent's background knowledge. And in general, when it comes to perception, swiftness is good. As one moves through the environment, or as aspects of it change, one needs to construct a representation of its main properties in time for both planning and action. Because failure to act in a timely fashion carries mortal risk, one might expect intense selection for swift and efficient visual (and other perceptual) processing. That suggests encapsulation.

It is worth noting that even this defense of encapsulation presupposes the use of stored (module-internal) information in perceptual processing. And for good reason. This is because two-dimensional patterns of activation on the retina radically underdetermine the properties of the three-dimensional distal scene. There is simply no way for the visual system to recover the latter *without* relying on stored knowledge of the world (including such facts as that nearer objects occlude distal ones, that light normally shines from above, and so on). Much of this knowledge may be innate,[4] with the remainder being learned over the course of development. But everyone agrees that what the visual system must do is parse and organize the incoming signals by relying on background knowledge and expectations of the structure of the world.[5]

[3] We should note that there are many others besides Fodor and Pylyshyn who defend the encapsulation of vision, at least in some form. Deroy (2013), for example, argues that the kinds of color-discrimination effects discussed by Macpherson (2012) (that is, cases where objects with characteristic colors are perceived differently from color-neutral objects) can be explained with a more enriched model of perceptual processing. But Deroy's focus is on just this one strand of evidence, which we ourselves don't rely on here. More ambitiously, Firestone and Scholl (forthcoming) provide a wide-ranging critique of claims of top–down penetration of vision. But as we have already noted, most of the claims they discuss rely on purely behavioral studies, and they pay scant attention to findings from cognitive neuroscience. Finally, Raftopoulos (2009) draws on an extensive body of empirical research to defend and elaborate the Fodor–Pylyshyn view. We do not have space in this paper for an adequate critique of his account, which would require a paper on its own.

[4] Strikingly, visual illusions like the Müller-Lyer illusion are present in children functionally-blind from birth (who were previously only capable of perceiving gross motion effects such as a hand waved close in front of the face) as soon as they are enabled to see for the first time following cataract surgery and intraocular lens implant (Ghandi et al. 2015).

[5] Recall that Pylyshyn (1999) explicitly allows that intra-modular top–down effects fail to qualify as forms of cognitive penetration. So even staunch advocates of encapsulation allow that perceptual processing involves an interaction between incoming signals and stored information.

Given that early vision will comprise interactive processing, using background knowledge to reduce noise and resolve ambiguities in the input, it is natural to wonder why this should *only* be true of *early* vision. After all, without speedy visual identification of something as a predator, swift processing in early vision would be for naught. That is, from the point of view of survival it doesn't matter whether one is able to process the shape and color of an object quickly if one doesn't also recognize its behavioral significance. And just as one might predict, object recognition is fast (indeed almost as fast as object *detection*; Grill-Spector and Kanwisher 2005; Mack and Palmeri 2010). But according to Pylyshyn, the processes responsible for object recognition are *un*encapsulated. So if object recognition is fast but unencapsulated, then something besides encapsulation must explain its swiftness. Because Pylyshyn, too, must posit fast unencapsulated processes, the charge that only encapsulated systems can be quick enough for one to evade predators is empty. One can therefore ask whether one should expect the same sort of interactive processing to operate at all levels in the visual system, including those that involve conceptual knowledge and contextual beliefs. This is precisely the picture we propose to defend.

Fodor might reply that, because unencapsulated systems are, by hypothesis, sensitive to one's background theories and value structures, what one sees will depend upon what one believes and wants. But an important part of what perception is *for* is to provide some degree of confirmation or disconfirmation of one's beliefs. If what one believes or desires can somehow inform what one sees, then one would be continually at risk of being led astray by false beliefs and lofty aspirations (with serious consequences for inclusive fitness). The argument, here, is that an unencapsulated system would be implausibly unreliable.

We note that in order for this second argument to undermine the claim that visual processing is unencapsulated, one must assume an especially unconstrained notion of cognitive penetration. (Lyons 2011, spells out this point in some detail.) However, we are not saying that higher-level cognitive states *determine* perceptual content, nor that sensory input fails to strongly constrain perceptual processing. It should also be stressed that the very same point emphasized above, that the input to the visual system vastly underdetermines the properties of the distal world, applies equally at all levels of visual processing, and not just within early vision. Moreover, vision needs to be flexible in the way that it deals with variations in context. The statistical properties of one environment (within a building, say) can be quite unlike those of another (such as a forest). So one should predict that some levels of visual processing will draw on high-level statistical knowledge of these sorts, and should be influenced by one's knowledge of where one is.

Of course it remains possible that the visual system cleaves cleanly into two parts, with an early system drawing on invariant knowledge and a later system utilizing contextual knowledge, with no possibility of the later system influencing the earlier one. But such a view now seems unmotivated. For often it may be contextual knowledge that one needs to draw on to parse the low-level structural properties of what one sees. Indeed, there is a good deal of accumulating evidence of such higher-to-lower influences, as we will show in due course. Certainly there is nothing in the neuroscience to support a cleavage between low-level and high-level vision in respect of feedback connections. On the contrary, such connections are rife at all levels throughout the visual system (and elsewhere), with feedback connections often

outnumbering feed-forward ones (Rockland and Pandya 1981; Felleman and Van Essen 1991; Gilbert and Li 2013).

In fact it is possible to turn Fodor's computational and epistemological arguments on their heads. Given that swift visual recognition is often of vital importance, but given that the visual input always radically underdetermines the nature of the categories in the distal environment, one needs to swiftly (but reliably) narrow down the range of possible hypotheses for further processing. Consistent with this suggestion, there is evidence that low spatial frequency—or "gist"—representations of objects are swiftly projected to orbitofrontal cortex via the fast-acting magnocellular pathway, where they activate stored concepts that might match the gist representation (Bar et al. 2006; Kveraga et al. 2007; Chaumon et al. 2013). These concepts are prioritized in light of one's values and goals and then projected back to high-level visual areas, arriving some 50 milliseconds earlier than the high-spatial-frequency processing being conducted through the slower but more accurate parvocellular pathway. Activity in orbitofrontal cortex caused by low-spatial-frequency stimuli predicts success in recognizing those stimuli. It also increases the functional connectivity between this area and visual cortex, as well as among higher and lower levels of the visual system itself. While this account remains to be fully explored, it is at least suggestive that top–down processing might actually speed up visual recognition by rendering the computations involved more tractable.

A third argument for the encapsulation of vision appeals to the persistence of visual illusions (Fodor 1983; Pylyshyn 1999; Firestone and Scholl 2014). For example, one's knowledge that the lines in a Müller-Lyer figure are of equal length (having just measured them) is incapable of penetrating and correcting one's visual system. For one nevertheless continues to see the lines as being of *un*equal length. We have two points to make in reply to this argument. The first is this. From the fact that high-level knowledge is incapable of *dominating* bottom–up processing in these circumstances (which is what the persistence of illusions shows) it simply does not follow that high-level knowledge cannot *influence* or *contribute to* lower-level processing (which is what encapsulation requires). Indeed, neither does it follow that high-level knowledge cannot dominate low-level processing in other kinds of circumstance. The argument from illusions is simply invalid. Now, one response to this criticism might be that the argument was never meant to be deductive in character. Rather, the point is that encapsulation nicely accounts for the persistence of illusions, while on interactionist views it might seem mysterious why one's belief should fail to modify the erroneous perceptual representation. However, recent developments in computational vision-science offer different ways of accounting for the phenomenon. This leads to our second point.

We broadly support the sorts of predictive-coding theoretical frameworks that are increasingly being used to characterize specific top–down influences in perception. (We will qualify this commitment in a moment.) According to some of these views, one would expect higher-level expectations to have an impact at lower levels in cases where processing within the latter has been unable to eradicate the noise or ambiguity in the input unaided. (And perhaps also in ways that are relevant to one's current task; see Section 4.) Remember, the picture is one of *multiple* interacting levels engaged simultaneously (or nearly simultaneously) in back-and-forth processing of the input. Higher-level knowledge and expectations are used to help reduce noise and resolve

ambiguity at the levels below, and those at the levels below that, and so on. But if the input is sufficiently unambiguous then the lower levels might settle on an interpretation without the highest levels ever being called upon.

We suggest that something like this occurs when the visual system is processing depth and size information while one looks at a Müller-Lyer figure. As far as the early levels of processing are concerned, relative depth and size have been accurately calculated from unambiguous cues. Hence systems monitoring noise and error levels are being told that everything is in order: there is no need for further processing. In contrast, in cases where the input is sufficiently ambiguous or degraded one might well expect that high-level beliefs or conceptual priming would have an impact on what one experiences. As we will see shortly, there is evidence that this is so.

Now here is the promised qualification: While we are committed to predictive-coding or Bayesian frameworks generally, these come in many forms, and we do not endorse the specific sort of account that has been embraced by philosophers interested in the topic (Clark 2013; Hohwy 2014). On this view, predictions are matched against incoming signals, cancelling out when they coincide, with only error-signals being propagated upwards at each level. Something like this might make sense in connection with motor-control, since if afferent feedback from the unfolding action matches one's forward model of the likely sensory consequences of the action, then everything is proceeding as intended, and no further attention needs to be paid to the action (Jeannerod 2006). But it makes little sense in connection with perception, where expectation-matching should lead to both a sharpening and an increase in volume of the incoming signal, as well as to suppression of noise. (Unless this were so, one would never consciously experience the low-level properties of what one expects, given that neural signals need to reach a critical activation threshold in order to be globally broadcast; Dehaene 2014.) Moreover, note that the pure-error-signal version of predictive coding entails that there should be a loss of information in visual cortex when incoming signals meet one's expectations (since they cancel one another out), whereas we think that expectation-matching should result in information *gain*. In a recent fMRI study Kok et al. (2012) were able to demonstrate the latter.

In addition to the arguments for encapsulation considered so far, Pylyshyn (1999) expresses skepticism about whether a mechanism of top–down influence is even so much as possible. Consider the suggestion that perception involves "proto-hypotheses" operating as a kind of filter, increasing the system's sensitivity to particular stimuli. If one is on a beach, for example, one might expect to see sailboats on the water, and this expectation might increase one's sensitivity to sail-shaped objects on the horizon. If something like this occurs, argues Pylyshyn, "we need a mechanism that can be tuned to, or which can somehow be made to select a certain sub-region of the parameter space" (353). But Pylyshyn doesn't think this sort of mechanism is possible: "Unfortunately, regions in some parameter space do not in general specify the type of categories we are interested in—that is, categories to which the visual system is supposed to be sensitized, according to the cognitive penetrability view of vision" (353).

What Pylyshyn seems to be saying is that in order for high-level filtering to operate on low-level stimulus properties ("regions of parameter space") such as light intensities, contour detection, motion, and spectral information, then information about the typical shape of sailboats (for example) would have to be encodable directly in such terms; but

there is no way to capture such abstract information in terms of low-level stimulus properties. We agree. However, what Pylyshyn appears to overlook is that there will be multiple levels of processing between one's high-level knowledge of the typical shape of a sailboat and the low-level parameters in question. One's knowledge of sailing-boat shapes might be stored as a set of prototypical profiles, for example; and these, in turn, might be coded in terms of sets of inter-linked contours, and so on down to the lowest levels. When one's concept SAILBOAT is activated, then, one might see activations spreading all the way down to the lowest levels in the network.

Of course a full answer to Pylyshyn's impossible-mechanism argument would require that one specify the nature of the mechanism of top–down influence in detail. Given the current state of scientific knowledge, no one is in a position to do that. But as we will see, there are hints in the empirical literature as to how the mechanism in question might work. And as we will also see, there is now voluminous evidence of cognitive penetration of early vision. Since cognitive penetration happens, we can infer that there must be *some* mechanism that enables it to happen, despite Pylyshyn's skepticism to the contrary.

We conclude that the theoretical arguments that have been offered in support of the encapsulation of early vision are unconvincing. But of course the question is an empirical one. We propose to spend the remainder of this paper reviewing evidence of top–down influences of a number of different sorts. But we should emphasize that our discussion is by no means comprehensive. In each case we present just a sample of representative data. Extensive citations to other similar findings can generally be obtained from the papers we discuss.

3 Visual Imagery and Early Vision

One line of empirical support for cognitive penetration of early vision derives from the study of visual imagery. When one visually imagines something—say a four-headed red dragon on the roof—concept-involving goals can be used to construct a visual or quasi-visual representation in a top–down manner (issuing in a state somewhat like *seeing* a four-headed red dragon on the roof, including the sorts of low-level properties that are normally processed within the early visual system). Moreover, there is ample evidence that visual imagery and vision share the same cortical regions, including those involved in early visual processing (Kosslyn 1994; Mechelli et al. 2004; Kosslyn et al. 2006; Reddy et al. 2010). So we can conclude that these top–down signals have caused activity within the early visual system of a semantically relevant sort. And if top–down signals can be used endogenously to *create* representations within the visual system in this way, then it is hard to understand why those same signals should be incapable of influencing bottom–up processing in cases of exogenously-caused perception.

It is not sufficient for our purposes to show merely that high-level goals can impact the visual system at *some* level, of course. It needs to be shown that the impact is specifically on early visual areas. In addition to the points already made above (that imagery often includes the sorts of low-level properties that are processed by early vision, and that activity in early visual areas shows up using fMRI in imagery tasks), another source of evidence derives from the study of people who have undergone localized brain damage. It seems that primary visual cortex (V1) is not strictly

necessary to support visual imagery, since someone with complete V1 damage can have normal imagery (Goldenberg et al. 1995; Bridge et al. 2012). But it does seem that surrounding areas V2, V3, and MT are necessary for visual imagery, since damage to these regions causes corresponding damage to capacities for visual imagery (Moro et al. 2008). Note that these areas, too, are known to engage in the sorts of processing characteristic of early vision. So it appears that endogenously-caused activity in these early visual areas is a necessary condition for visual imagery to occur.

Another source of evidence that visual imagery activates content in early visual areas comes from multivariate pattern-analysis of fMRI data. Thus Vetter et al. (2014) trained pattern-classifiers to identify what people were hearing or seeing from patterns of neural activity in the early visual system (V1, V2, and V3). The trained classifier was then able to discriminate whether people were imagining a forest, or traffic, or a group of people talking. It seems that the high-level goal of forming such an image is capable of causing category-specific patterns of activity in early visual areas. For how could a pattern-classifier trained on perceptually-caused activity in the early visual system generalize thereafter to identify what people are imagining, unless many of the same voxels in early visual cortex were differentially active in each case? And then since the patterns of activity in the latter cases are caused top–down by semantically rich intentions ("Visualize a group of people talking"), we can conclude that these early visual areas are cognitively penetrable.

Albers et al. (2013) also used pattern classifiers to investigate neural activity in early vision in cases where people saw a grating, or held a representation of a perceived grating in working memory, or followed instructions to generate a mental image of a grating with a particular orientation. In each case the fMRI classifier was able to identify the orientation of the grating from patterns of neural activity in early visual cortex (either from V1, V2, and V3 collapsed together, or within each individually), and it did so with highly significant degrees of reliability. Notably, when participants were visually imagining the gratings, patterns of activity in early visual cortex closely resembled the observed patterns in cases where people perceived a grating of the same orientation, suggesting that the same mechanisms are implicated in each. Moreover, this resemblance was greater in people who are better at visual imagining generally.

We can envisage two possible lines of reply to this argument. The first would be to say that although vision and imagination depend on the same early-visual brain regions, the relevant neural populations within these regions are interwoven but disjoint. (fMRI pattern-classification paradigms would then fail to distinguish this possibility from the one we are advancing, since the patterns are constructed on a voxel-by-voxel basis, where each voxel houses thousands of individual neurons.) That is, it might be said that one set of neurons can be activated in top–down manner for purposes of visual imagery, whereas a distinct set is involved in bottom–up visual processing; and the former cannot influence the latter. This is possible in principle, of course. (And if true, it might be capable of explaining away some of the data to be discussed in later sections, as we will see.) But it runs into trouble in accounting for why visual imagery and visual perception should interfere with one another when targets are located concurrently in the same position in the visual field (Craver-Lemley and Reeves 1992), as well as why the content of visual imagery should bias subsequent perception (Pearson et al. 2008).

Moreover, notice that this parallel-systems idea makes specific commitments about the neuronal architecture of the visual system. It requires that there should be one set of

neurons that receives only bottom–up input from earlier in the visual processing stream and passes output only to higher levels. And there should be a distinct set of neurons that only receives input from higher levels while also only passing output back to higher levels. In addition, there should be no sideways connections among neurons from the two sets. Although a great deal is now known about patterns of neural connectivity in visual cortex, we are aware of no evidence for any such arrangement.

Furthermore, the parallel-systems account carries significant explanatory costs. It requires us to postulate that much of the functionality of the visual system is replicated in a separate imagery system. For the top–down-caused patterns of activity in early visual areas would need to be bound together and integrated into a coherent quasi-visual percept, much as happens in vision itself. Those patterns would also have to be capable of becoming targets of attention, resulting in the global broadcast (and conscious status) of the images in question. A one-system view is simpler, since it appeals to the same set of binding and broadcasting mechanisms in both cases. Moreover, we have no evidence of the existence of two parallel mechanisms. So there are no known mechanisms in terms of which we can explain how visual imagery is processed. In contrast, if we assume that the same mechanisms are used for each, then we can explain how imagery becomes integrated and conscious by appealing to known properties of the visual system. Overall, then, it is much more plausible to think that imagery involves top–down-caused patterns of activity in the very same neuronal populations in early visual areas that can be stimulated bottom–up during perception.

A second possible reply to the argument from visual imagery would appeal to the alleged phenomenon of *neural re-use* (Anderson 2010). It might be said that the very same neural assemblies in early visual areas that are activated bottom–up during perception are also activated top–down during imagination, but without top–down activation ever having an influence on perception itself. Notice, however, that this response concedes the cognitive *penetrability* of the visual system; it merely claims that it never happens in the case of online perception of the external world. Such a claim is possible in principle, no doubt; but it seems arbitrary and unmotivated, particularly in light of the data reviewed below, which demonstrate top–down modulation of low-level visual processing.

We conclude that there is good reason to think, not only that visual imagery involves top–down-caused activity in early visual brain areas, but that these effects involve modulations of the early visual system itself. If so, then that system is not encapsulated. We now turn to consider direct evidence of top–down effects on visual processing. We begin (in Section 4) with the effects of task-goals on individual neurons in V1. Thereafter (in Sections 5 and 6) we consider top–down effects of a cognitive sort.

4 Task-Related Effects in V1

There have been a number of demonstrations of task-related effects in V1, showing that neurons in this region encode information differently depending on the nature of the task. For example, Gutteling et al. (2015) used pattern analysis on activity in V1 obtained through fMRI to decode with 70% accuracy whether participants were preparing to point to, or to grasp, an oblong shape. They cite this and other evidence to argue that action-preparation causes task-related changes in the processing of early

visual areas, designed to increase the accuracy of visual representations of task-related features (such as orientation and width to guide grasping).

Among a wealth of similar evidence we propose to focus on a study by Li et al. (2004), which seems to us to be particularly instructive.[6] The investigators designed a set of stimuli that would allow them to manipulate the task while holding fixed the visual input. Early phases of the study involved training macaque monkeys to perform two different tasks with two distinct sets of stimuli. The stimuli for the *end-flank task* consisted of central bar flanked at each end by two identically sized bars. The two end-flanking bars were always collinear, but could be offset on either side of the middle bar. After presentation of the bars, left and right fixation dots appeared: the macaques' task was to fixate their gaze on (saccade to) the dot located on the side that the end-flanking bars were offset. (During the response phase, the main display was extinguished, and only the two fixation dots on either side of the display were presented.) The stimuli for the *side-flank task* consisted of identical bars, but this time the central bar was flanked on each side by bars located at varying distances. The task, here, involved saccading to the side that was closest to the center bar.

Following training the two sets of stimuli were combined into 25 distinct stimuli. (See Fig. 1). For each trial, a visual cue indicated which task (whether end-flank or side-flank) was to be performed. As the macaques performed the visual tasks, Li and colleagues took single-cell recordings from V1 neurons, each with receptive fields corresponding to a particular location on the stimulus. The experimenters reasoned that if visual processing is encapsulated from task-related effects, then we should expect there to be no difference between the two task conditions.

This is not what they found, however. Rather, neural tuning-curves were more sensitive to changes in locations of the side- and end-flanking bars, *when those changes were relevant to the visual task being performed*.[7] For example, when the experimenters manipulated the side flanks while the macaques were performing the end-flank task, neural response curves tended to remain flat—i.e., neural responses across the different side-flank positions remained (more or less) constant. However, when the experimenters manipulated the end flanks during the end-flank task neural responses varied considerably across the different end-flank positions. They found mirror-image results when looking at the effects of side- and end-flank manipulation during the side-flank task. The general finding is that neural response-curves are more sensitive to changes in the display when those changes are relevant to the task.[8]

[6] For a recent replication and elaboration of Li et al.'s (2004) results, see Ramalingam et al. (2013). For related data, see Ghose and Bearl (2010) and McManus et al. (2011). For a recent review of the literature, see Gilbert and Li (2013).

[7] A tuning-curve represents a neuron's response profile to changes in a stimulus along a particular dimension. (Roughly, it tells us what, along that dimension, the neuron represents.) A flat tuning-curve in response to some change in the stimulus indicates that the neuron is not displaying sensitivity to the manipulated stimulus property.

[8] Note that it is quite possible that the macaques acquired some perceptual expertise over the course of the training periods. Perceptual learning is thought to be a long-term change to how visual information is processed. But recall that the effect here is context sensitive. Any expertise gained from, for example, performing the side-flank task must be flexibly deployed in accordance with the current goal. So even if perceptual learning occurs in this experiment, there must be some sort of contentful relation between the high-level goal and visual processing.

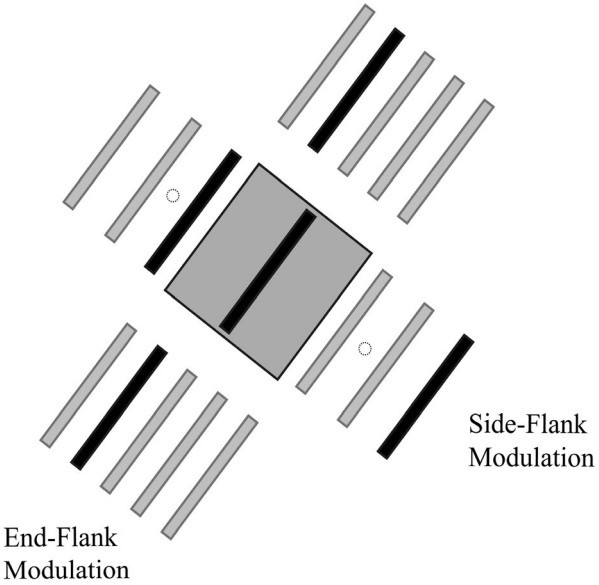

Fig. 1 A schematic diagram of Li et al.'s stimuli. The *black bars* indicate one of the possible displays; *gray bars* indicate the 24 alternative locations. The *circles* represent the two fixation *dots* presented following the main stimulus. To cue each task, the experimenters used colored task-relevant bars (middle and end bars for the end-flank task, and middle and side bars for the side-flank task). The *colored bars*, alone, had no effect on the activity of the measured neurons. The *center gray box* is included here only to provide a spatial frame of reference, and did not appear in the actual display

As a way to quantify these differences between task-relevant and task-irrelevant conditions, Li and colleagues utilized a mutual-information metric, which found that task-relevant neural responses better predicted the stimulus than did task-irrelevant ones. This suggests that the differential neural responses in the task-relevant conditions carried information that could be useful to the monkeys when performing the task. Moreover, although differential neural responses were observed from the start of the experiment, the amount of task-relevant information carried by neuronal tuning-curves increased as the monkeys' behavioral performance improved, suggesting that the former was an underlying cause of the improvement.[9]

Someone might object that these results could be explained in terms of an influence of ocular motor planning on neurons in V1. And if this is the case, then we haven't shown that visual processing is sensitive to genuinely cognitive states. To be sure, the data from this experiment are consistent with the hypothesis that signals from motor plans or intentions are what modulate these V1 neurons. However, even if this is so, we

[9] How is it possible for goals to modulate visual processing in this way? Li et al. (2004) explain that in addition to multiple feed-forward and feed-back connections between V1 neurons and higher areas, there are also strong lateral connections among neurons in V1 itself. These can have either excitatory or inhibitory roles, making the activity of one neuron partly dependent upon the activities of many of its neighbors. So a plausible explanation is that a change in task modulates how a given V1 neuron is influenced by some of these neighbors rather than others. That is, the effect of the side-flank task on a V1 neuron amounts to telling it something like, "be more influenced in your response by *these* nearby neurons [which code for the distances between the center line and the side-flanks] than by *those* ones [which code for the alignment of the center line with the end-flanks]."

still have good reason to think that the upshot is a case of cognitive penetration. The first thing to note is that there is independent reason to believe that the measured neurons are low-level *visual* neurons, responsible for encoding structural properties of the scene (as opposed to motor neurons). Hence, the effects discovered in this experiment are effects on how visual information is encoded. Secondly, the effects are task specific: change the task being performed, and there is a change in the visual information encoded. So even if these neurons are modulated by motor plans, the motor plans must (somehow) be sensitive to the (a) task being performed, and (b) the configuration of the bars relevant to the task. Thus if it is signals from ocular motor areas that are responsible for modulating these neurons in V1, then they mediate a semantic relation between high-level states and low-level visual processing (which is inconsistent with the encapsulated status of the latter).

Alternatively, can these findings be explained as effects of attention on V1? In a sense, perhaps. This is because some scientists have operationalized the distinction between attention and expectation in terms of the distinction between *task-relevance* and *probability* (Summerfield and Egner 2009). They have devised experiments in which one can either vary the task that participants perform, while holding constant the probabilities of various task conditions, or vary the probabilities while holding the task constant. (The rationale is that attention varies depending on what is relevant to one's task, whereas probabilities can be known independent of task, and hence independent of current attention.) And in the experiments of Li et al. (2004) it was the task that varied. So using this heuristic, the single-cell work of Li and colleagues comes out as a form of *attentional* modulation of the coding-properties of V1 neurons.

Considered very abstractly, therefore, this sort of work is consistent with Pylyshyn's account of encapsulation, which allows that attention can have an impact on early vision. Note, however, that Li and colleagues' results cannot be explained as an effect of *spatial* attention. This is because they also ran an experiment that looked at the effects of attending versus not-attending to the location, and found that attending merely boosted neural responses overall. Attending didn't affect *how* the neurons responded to changes in the stimulus (as the task did). Yet even if these results reflect a kind of feature-based or object-based attention, they nonetheless demonstrate the subtlety and content-dependence of these forms of attention. Attention is not always a simple spotlight, boosting the responses of whatever falls within its spatial focus. Indeed, what changes in the two conditions is not just the *strength* of each neuron's response, but the *information* that is carried by its pattern of responding. The effect of attention on V1 in this experiment is thus a *content related one*, which is inconsistent with Pylyshyn's account of encapsulation. For there is a semantic relationship between the high-level intentions created by the monkey's task on a given trial ("Look for end-flank offset" versus "Look for side-flank offset") and the information about the stimulus that is carried by a given V1 neuron.

5 Implicit Expectations Bias Processing in Early Vision

We now turn to consider the effects of higher-level cognitive states on visual processing. There have been a number of careful demonstrations of the way in which people's expectations (especially their implicit beliefs about the statistical structure of the

environment) can influence processing in early vision independently of the effects of attention (Kok et al. 2012, 2013, 2014; Wyart et al. 2012). For example, using multivariate pattern analysis on fMRI data Kok et al. (2014) show that when people explicitly expect a grating to appear that is actually omitted, the pattern of anticipatory neural activation in V1 is similar to the pattern that occurs when such a grating is perceived. The finding suggests that expectation sets up a sensory template in V1 that can then be matched against the incoming information.

Here we will focus on one expectation-study in particular: Kok et al. (2013). Participants in this experiment were set the task of judging the direction of motion of a random-dot-motion pattern. (In such stimuli most of the dots move randomly, but a subset move coherently in a single direction.) They were told that the direction of motion could be anywhere within a 90° arc, and used left/right buttons to control the orientation of a line to report the perceived direction of motion. In fact, however, the directions of motion were restricted to five equally-spaced directions within the 90° arc. During both the experimental and prior learning trials, auditory cues (a high tone or a low tone) were probabilistically-paired with specific orientations (27.5° for the high tone and 62.5° for the low tone). Participants were told to ignore the tones, and for the most part learning was implicit.[10] The task was performed in both a conventional behavioral setting and in an fMRI scanner.

Behaviorally, Kok and colleagues found that auditory cues biased participants' judgments of orientation in the cued direction. For example, when hearing a tone that predicted motion oriented at 27.5°, a participant might perceive motion objectively orientated at 45° as oriented at 35°. So we know that participants had learned the statistical properties of the experimental environment, and that their implicit expectations were biasing their judgments. The question is whether this was merely an effect on post-perceptual judgment, or whether these expectations were influencing motion-processing early in the visual system itself.

The fMRI data confirm the latter. The investigators used a forward-modelling approach to estimate the perceived direction of motion of the stimulus on each trial. This essentially involved collecting fMRI data from motion-selective voxels in areas V1, V2, and V3 on each trial, and using that as input to a direction-sensitive artificial neural network. The fMRI models matched the participants' reports of the perceived direction better than they did the actual directions of motion, which suggests that the model accurately represents direction-sensitive processing in early vision. Further support for the validity of the model comes from the fact that there was a positive correlation between participants' behavioral and modeled responses. For example, if someone showed a stronger bias than others in the behavioral condition, then so did her fMRI forward model. It seems that people's implicit expectations were biasing motion processing in early visual areas in the expected direction.

It is possible to ask, however, where such expectations are encoded. For as noted above, most subjects were unaware of the predictive nature of the cues. Perhaps the predictions were stored as inter-modular associations between auditory cortex and early

[10] Post-experiment interviews revealed that 80 % of participants suspected no relationship between the auditory cue and orientation of motion. Of the remaining 20 %, one participant was aware of the true significance of the cues, one was aware of *a* relationship, but had their predictive character reversed; and the remaining three participants suspected a relationship between just one of the auditory cues and presented orientation.

visual cortex. If so, this need not be a problem for Pylyshyn. The evidence suggests, however, that associative connections of this sort are not stored as direct links between otherwise-encapsulated sensory systems. Rather, neuropsychological data suggests an essential role for medial temporal cortex, or parahippocampal cortex, or both, which are areas not generally thought to be part of the visual system (nor the auditory system), but rather form crucial components of the long-term memory system. Thus Murray et al. (1993) show by removing various regions from the brains of monkeys that the medial temporal lobe is necessary for learning new statistical associations. And Schapiro et al. (2014) studied a human patient known as LSJ who has suffered complete bilateral loss of the hippocampal formation and surrounding medial temporal lobe. It was found that not only is LSJ incapable of forming new episodic memories (which was already known), but that she is also incapable of implicitly learning new statistical associations between events in her environment. So it seems that the findings described earlier in this section demonstrate the penetration of early vision by information stored outside the visual system itself.[11]

As with the data from Li et al. (2004) discussed in Section 4, it might be possible to explain these results as effects of attention. For expecting the overall pattern of motion on a given trial to be oriented at 27.5° (say) might lead one to attend more to dots whose motion is most consistent with that expectation, thereby according them greater weight in the process that calculates overall direction. But here too (as in Section 4), this cannot be an effect of mere spatial attention. Attending to a particular location within the display couldn't cause any such effect. And again, even if the effect *is* an attentional one of some sort, it seems to violate Pylyshyn's no-semantic-influence constraint on encapsulation. Rather, it seems that an expectation of motion oriented at 27.5° causes modulations of attention with the content, *motion at 27.5°* or something similar, thereby altering the perceived direction of motion.

It will also be instructive, here, to re-visit the two alternative proposals discussed in Section 3. One was that visual imagery *re-uses* neurons in early vision for a different purpose. That cannot, of course, apply here. For trials with and without any prior expectation about direction of motion are each perceptual in nature. So if the same neurons are involved in both, prior expectations must be modulating the behavior of perceptual neurons in a top–down manner, in a way that alters the content of subsequent perception. Notice, too, that what we know for sure is that there is expectation-driven *information* about direction of motion encoded in patterns of activity in early visual cortex; otherwise the pattern analyzer couldn't discriminate between one expected orientation and another. And since the presence of that information has an impact on people's reports of what they see, it is reasonable to think that top–down modulations of activity in early vision are altering the representational contents encoded in early visual areas, and subsequently conscious perception.

The other alternative account we considered in Section 3 is that while neurons in early visual cortex can be modulated by visual imagery and other top–down influences, they form a disjoint set with those involved in bottom–up perceptual processing. That

[11] This kind of implicit biasing of visual processing doesn't involve conscious beliefs or desires penetrating vision, of course. As a result, many supporters of visual modularity may not regard such cases as particularly interesting counterexamples. Given that much of cognition operates at an implicit level, however, we fail to see why this should make the effects any less important.

idea could also apply here, provided that imagery and perception can simultaneously influence perceptual judgment. Thus it might be said that implicit expectations of direction of motion cause activity in image-system neurons in early visual cortex of the appropriate sort. Since these are interleaved with visual-system neurons, both sorts of activity are picked up by the fMRI pattern-classifier. And both imagistic and veridical perceptual representations are received as input by the decision-making process, where they each influence the resulting judgment. While this remains a possibility, we think it is quite unlikely, for the reasons given in Section 3.

6 Semantic Knowledge Alters Early Processing

Many researchers have used so-called "Mooney" images to investigate the effects of semantic knowledge on perception. (These are two-tone images in which shading, shadows, and color are converted to black and white. A famous example that may be familiar to many is the Dalmatian dog walking in dappled light, hidden in a two-tone figure consisting of black splotches on a white background. See Fig. 2). For example, in a suggestive early experiment Moore and Cavanagh (1998) found evidence that people's perception of three-dimensional volume in Mooney images of known objects is not constructed bottom–up, but rather depends on higher-level semantic knowledge of the identity of the object. For Mooney images of unfamiliar objects are not perceived volumetrically, and nor are images formed by recombining the parts of a familiar image.

More recently, Hsieh et al. (2010) used fMRI and pattern analysis to show that perceiving the meaning in a Mooney image alters processing in early visual cortex. They first scanned participants while they viewed a series of Mooney images without recognition; then participants were scanned while they saw the full grey-scale images from which the two-tone figures were derived; and then they viewed the Mooney images again, this time with recognition. The pattern of activity in early visual cortex during the final phase (when the Mooney image was meaningful) was more similar to activity when perceiving the related grey-scale image than it was to the pattern of

Fig. 2 The hidden Dalmatian

activity when perceiving the Mooney image in the first phase. Since the external stimuli in the first and final phases were identical, bottom–up processing should likewise have been the same. So the shift in the pattern of activity in the final phase can only result from the top–down influence of participants' knowledge of the meaning of the stimulus.

Teufel et al. (forthcoming) also used Mooney images, contrasting localized "edge detection" performance before and after experience with the full grey-scale image had rendered the two-tone image meaningful. Participants' only task was to detect the orientation of a faint line in a cued location, of a size that would fall within the receptive fields of V1 neurons. The line in question was either aligned with, or orthogonal to, the unseen "imaginary" edge of the object represented in the Mooney image. The finding was that detection was better (for aligned but not for orthogonal lines) when the target could help complete a recognized object, as opposed to the same figure in the absence of recognition. This suggests that high-level knowledge of the identity of the object had altered the sensitivity of V1 neurons, and Teufel and colleagues argue that the effect they obtained was independent of focal attention.

Neri (2014) also required participants to identify a small target line-orientation, but embedded in briefly presented natural scenes. Perception of the meaning of the scenes was disrupted on half of the trials via upside-down presentation. The finding was that participants were better at probe identification when the probe aligned with a component in a meaningful scene. Neri argues that this result is best explained in terms of a top–down predictive strategy in which global meaning properties are used to guide and refine local image construction. Computational modeling confirms this interpretation, and is able to show that the effect is independent of the effects of attention.

It might be argued in connection with all these findings that the real work in biasing early visual processing is done by a structural representation of an object or scene, not by its more abstract semantic–conceptual meaning. Although a Mooney image might be recognized as a woman kissing a horse on the nose, for example, it may be that the concepts KISS and HORSE play no direct role in biasing the processing of the image and its imaginary boundaries in V1. Rather, it may be the structural–spatial description of this particular horse that does the work. And this, recall, is regarded by Pylyshyn (1999, 2003) as the highest level of description produced by the early visual system. So the effect might be said to be an *intra*-modular one, which would be consistent with encapsulation.

This is a fair point. But Hegde and Kersten (2010) show that distinct brain regions are involved in storing the meaning of Mooney images and in processing those images to recognize them online. Using fMRI, Hegde & Kersten provide evidence that the storage function depends on regions of medial parietal cortex, whereas recognition implicates a second set of regions including the superior temporal sulcus. Moreover, functional connectivity between the two sets of regions is greatest in cases where recognition of the Mooney image is strongest. These findings suggest that memory for Mooney images is stored and activated outside the early visual system, since neither of these regions is normally reckoned to be part of the latter.

We know, moreover, that perception of meaning in Mooney figures can be secured not only by showing the original grey-scale picture, but also by conceptual priming of various sorts (such as being told that the hidden-Dalmatian figure contains an image of a dog, or by hearing a dog barking). And there are also demonstrations that conceptual

priming can bias the perception of ambiguous figures such as the duck–rabbit (Balcetis and Dale 2007). We are not aware, however, that anyone has succeeded in separating such effects from the influence of attention. Even so, such findings are instructive. For there is nothing about the concept DOG that would lead one to attend to one portion of the visual field rather than another. In order to work, the conceptual prime must guide attention toward low-level features or shapes that might constitute parts of the relevant object (such as a dog's nose or ear). At the very least, attention and real-world knowledge must work in concert with one another in such cases.

There are, moreover, other demonstrations of the impact of concepts on experience that seem unlikely to reduce to the effects of attention. Costello et al. (2009) show, for example, that consciously experienced semantic primes influence the speed with which a word can break through continuous flash suppression.[12] When the suppressed image is a word like "Pepper" it emerges from suppression more quickly if participants are previously primed with a semantically-related word like "Salt" than by an unrelated word like "Tree". Likewise, Lupyan and Ward (2013) show that images of objects hidden by continuous flash suppression emerge from suppression more swiftly and reliably when preceded by a valid word cue (e.g. "pumpkin", when the suppressed image is a pumpkin) in contrast to either no cue or an invalid cue. Similar results were obtained by Pinto et al. (2015), who were also able to show that valid primes had no effect on the speed with which a stimulus could be identified that gradually increased in intensity from zero in the *absence* of flash suppression. They take this to show that the effects of cuing on emergence from suppression is not merely attentional, but result rather from the use of top–down information to help resolve processing of ambiguous stimuli (such as are present during flash suppression).

The question for us is whether conceptual primes alter processing of the suppressed stimulus at early stages in the visual system, or whether they merely have an effect on high-level semantic representations. In fact many in the field think that continuous flash suppression results in neural dominance early in the visual-processing stream, interfering especially with early representations of orientation and contrast (Tsuchiya and Koch 2005; Lin and He 2009; Yang and Blake 2012). For example, Krieman et al. (2002) took single-neuron recordings of neurons in the medial temporal lobes of human subjects during continuous flash suppression. They found that none of the 51 neurons that responded to categories or to specific objects in normal viewing conditions responded above baseline when images of those categories or objects were suppressed. This suggests that continuous flash suppression blocks all processing beyond the early visual system. If this is so, then the results of the conceptual priming studies described above cannot be explained in terms of influences on high-level vision. Rather, it would seem that activating the concepts in question sensitized processing within the early visual system to help boost the suppressed signals to the point of visibility.

In contrast with this apparent consensus, Sklar et al. (2012) provide evidence of high-level semantic processing in continuous flash suppression. They show that semantically-incongruous sentences such as, "I ironed coffee", emerge from flash

[12] Continuous flash suppression is a form of binocular rivalry, in which different stimuli are presented to each eye simultaneously. But in continuous flash suppression the stimuli presented to one eye consist of high-contrast dynamically changing Mondrian-like colored patterns. These tend to dominate conscious experience, with the stimulus presented to the other eye taking considerable time to become visible.

suppression more swiftly than do semantically coherent ones ("I made coffee"). They also show that presentation of two-step subtraction problems under flash suppression (such as "9 – 3 – 4") prime correct ("2") but not incorrect answers, suggesting that the problems can be represented and solved unconsciously. We do not know how to resolve the conflict between these bodies of evidence. Notice, however, that even if high-level semantic properties are processed and represented in continuous flash suppression, the visibility-boost provided by conceptual primes is unlikely to be merely a bias in post-perceptual judgment. This is because it is object *configurations* that pop out from flash suppression. Hence concepts would seem to be capable of influencing the final stages of early visual processing, at any rate. And even if a concept like PUMPKIN only directly primes pumpkin-like shapes and configurations, these may in turn have primed lower-level pumpkin-relevant representations, with expectations cascading down the hierarchy to early visual processing in the manner suggested by our discussion of Mooney images.

Consistent with the latter suggestion, a number of studies have used EEG measurements (which have high temporal resolution, unlike fMRI) to show effects of concepts on early visual processing, and prior to the first impact of attention which occurs at around 200 milliseconds (Thierry et al. 2009; Mo et al. 2011). For example, Boutonnet and Lupyan (2015) used a cued category-recognition task combined with EEG measurements of electrical activity in early visual cortex. Specifically, they measured the amplitude of the P100 which occurs over early visual cortex about 100 milliseconds after stimulus onset, and which is thought to be a pre-attentive signal of the initial processing and integration of low-level features of the stimulus. Participants were presented with an auditory cue on each trial, then shortly afterwards a picture was displayed and they had to indicate via a button-press whether or not the picture matched the cue. For example, they might hear the word "dog", or hear the sound of a dog barking, before a picture, either of a dog or another category of object, was displayed. In general, valid cues speeded people's reaction times, and the verbal cue had a bigger effect than did the category-distinctive sound (e.g. barking). These are findings that a modularity theorist might also predict. But they were mirrored in the amplitude of the P100, suggesting that the concept activated by the previously presented word had sensitized category-specific representations in early visual cortex.

7 Conclusion

We believe that the notion of encapsulation has outlived its usefulness in the study of vision (and also in the study of perceptual systems generally, although we have not argued for this here). The theoretical arguments for expecting there to be a level of visual processing that is encapsulated from the rest of cognition are unconvincing. And as we have shown, there is substantial evidence that higher-level goals, expectations, and concepts can interact with the visual system at the earliest stages of cortical processing, helping to speed up and improve the reliability of our perceptions.

We should stress, however, that the interactive account of visual processing that we have defended here is not scientifically radical. (We make no claims for its consequences for philosophy.) We are not calling for a revolution of the perceptual sciences, but a modest amendment to standard models of visual processing. Indeed, all of the

studies we have discussed presuppose models of the visual system gained from standard bottom–up investigations; and many of the top–down factors we have cited involve biasing variables employed in standard, bottom–up, models of visual processing. Nor do we think that anything we have said need eradicate the distinction between perception and cognition. But if we are right, then informational encapsulation isn't what demarcates that boundary. Rather, the distinction is between cortical regions that are specialized for processing the incoming stream of information from the retina (while also receiving top–down signals), and those whose functions are more general.

Acknowledgments The authors are grateful to Tom Carlson, Chaz Firestone, Zoe Jenkins, Peter Kok, Eric Mandelbaum, Jake Quilty-Dunn, and an anonymous referee for their comments on earlier drafts of this paper.

References

Albers, A., P. Kok, I. Toni, C. Dijkerman, and F. de Lange. 2013. Shared representations for working memory and mental imagery in early visual cortex. *Current Biology* 23: 1427–1431.

Anderson, M.L. 2010. Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 33: 245–266.

Balcetis, E., and R. Dale. 2007. Conceptual set as a top-down constraint on visual object identification. *Perception* 36: 581–595.

Bar, M., K. Kassam, A. Ghuman, J. Boshyan, A. Schmid, A. Dale, M. Hämäläinen, K. Marinkovic, D. Schacter, B. Rosen, and E. Halgren. 2006. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences* 103: 449–454.

Barrett, H., and R. Kurzban. 2006. Modularity in cognition. *Psychological Review* 113: 628–647.

Boutonnet, B., and Lupyan, G. 2015. Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*.

Brewer, W.F., and L. Loschky. 2005. Top-down and bottom-up influences on observation: Evidence from cognitive psychology and the history of science. In *Cognitive penetrability of perception*, ed. A. Raftopoulos, 31–47. New York: Nova Science.

Bridge, H., S. Harrold, E. Holmes, M. Stokes, and C. Kennard. 2012. Vivid visual mental imagery in the absence of the primary visual cortex. *Journal of Neurology* 259: 1062–1070.

Brogaard, B., K. Marlow, and K. Rice. 2014. The long-term potentiation model for grapheme-color binding in synesthesia. In *Sensory integration and the unity of consciousness*, ed. D. Bennett and C. Hill, 37–72. Cambridge, MA: MIT Press.

Carruthers, P. 2006. *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford: Oxford University Press.

Chaumon, M., K. Kveraga, L. Feldman Barrett, and M. Bar. 2014. Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral Cortex* 24: 2899–2907.

Churchland, P.M. 1988. Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science* 55: 167–187.

Churchland, P.S., V. Ramachandran, and T. Sejnowski. 1994. A critique of pure vision. In *Large-scale neuronal theories of the brain*, ed. T. Sejnowski, C. Koch, and J. Davis, 23–60. Cambridge: Bradford.

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

Costello, P., Y. Jiang, B. Baartman, K. McGlennen, and S. He. 2009. Semantic and subword priming during binocular suppression. *Consciousness and Cognition* 18: 375–382.

Craver-Lemley, C., and A. Reeves. 1992. How visual imagery interferes with vision. *Psychological Review* 99: 633–649.

Davis, T., and R. Poldrack. 2013. Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences* 1296(1): 108–134.

Dehaene, S. 2014. *Consciousness and the brain*. New York: Viking.

Deroy, O. 2013. Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies* 162: 87–107.

Felleman, D., and D. Van Essen. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1: 1–47.
Firestone, C., and B. Scholl. 2014. "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science* 25: 38–46.
Firestone, C., and B.J. Scholl. 2015. Can you experience "top-down" effects on perception?: The case of race categories and perceived lightness. *Psychonomic Bulletin & Review* 22: 694–700.
Firestone, C. and Scholl, B. forthcoming. Cognition does not affect perception: Evaluating the evidence for "top–down" effects. *Behavioral and Brain Sciences*.
Fodor, J. 1983. *The modularity of mind*. Cambridge: MIT Press.
Ghandi, T., A. Kalia, S. Ganesh, and P. Sinha. 2015. Immediate susceptibility to visual illusions after sight onset. *Current Biology* 25: R359.
Ghose, G., and D. Bearl. 2010. Attention directed by expectations enhances receptive fields in area MT. *Vision Research* 50: 441–451.
Gilbert, D., and W. Li. 2013. Top–down influences on visual processing. *Nature Reviews Neuroscience* 14: 350–363.
Goldenberg, G., W. Müllbacher, and A. Nowak. 1995. Imagery without perception: A case study of anosognosia for cortical blindness. *Neuropsychologia* 33: 1373–1382.
Grill-Spector, K., and N. Kanwisher. 2005. As soon as you know it is there, you know what it is. *Psychological Science* 16: 152–160.
Gutteling, T., N. Petridou, S. Dumoulin, B. Harvey, E. Aarnoutse, J. Kenemans, and S. Neggers. 2015. Action preparation shapes processing in early visual cortex. *Journal of Neuroscience* 35: 6472–6480.
Hanson, N. 1965. *Patterns of discovery*. Cambridge: Cambridge University Press.
Haynes, J.D. 2015. A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron* 87: 257–270.
Hegde, J., and D. Kersten. 2010. A link between visual disambiguation and visual memory. *Journal of Neuroscience* 30: 15124–15133.
Hohwy, J. 2014. *The predictive mind*. Oxford: Oxford University Press.
Hsieh, P., E. Vul, and N. Kanwisher. 2010. Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *Journal of Neurophysiology* 103: 1501–1507.
Jeannerod, M. 2006. *Motor cognition*. Oxford: Oxford University Press.
Kok, P., J. Jehee, and F. de Lange. 2012. Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron* 75: 265–270.
Kok, P., G. Brouwer, M. van Gerven, and F. de Lange. 2013. Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience* 33: 16275–16284.
Kok, P., M. Failing, and F. de Lange. 2014. Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience* 26: 1546–1554.
Kosslyn, S. 1994. *Image and brain*. Cambridge: MIT Press.
Kosslyn, S., W. Thompson, and G. Ganis. 2006. *The case for mental imagery*. Oxford: Oxford University Press.
Krieman, G., I. Fried, and C. Koch. 2002. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences* 99: 8378–8383.
Kveraga, K., J. Boshyan, and M. Bar. 2007. Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience* 27: 13232–13240.
Li, W., V. Piëch, and C. Gilbert. 2004. Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience* 7: 651–657.
Lin, Z., and S. He. 2009. Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Progress in Neurobiology* 87: 195–211.
Lupyan, G., and E. Ward. 2013. Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences* 110: 14196–14201.
Lyons, J. 2011. Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues* 21: 289–311.
Mack, M., and T. Palmeri. 2010. Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception and Performance* 36: 1067–1079.
Macpherson, F. 2012. Cognitive penetration of color experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research* 84: 24–62.
Masrour, F., G. Nirshberg, M. Schon, J. Leardi, and E. Barrett. 2015. Revisiting the empirical case against perceptual modularity. *Frontiers in Psychology* 6, doi: 10.3389/fpsyg.2015.01676.
McManus, J., W. Li, and C. Gilbert. 2011. Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences* 108: 9739–9746.

Mechelli, A., C. Price, K. Friston, and A. Ishai. 2004. Where bottom–up meets top–down: Neuronal interactions during perception and imagery. *Cerebral Cortex* 14: 1256–1265.

Mo, L., G. Xu, P. Kay, and L.H. Tan. 2011. Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences* 108: 14026–14030.

Moore, C., and P. Cavanagh. 1998. Recovery of 3D volume from 2-tone images of novel objects. *Cognition* 67: 45–71.

Moro, V., G. Berlucchi, J. Lerch, F. Tomaiuolo, and S. Aglioti. 2008. Selective deficit in mental visual imagery with intact primary visual cortex and visual perception. *Cortex* 44: 109–118.

Murray, E., D. Gaffan, and M. Mishkin. 1993. Neural substrates of visual stimulus–stimulus association in Rhesus monkeys. *Journal of Neuroscience* 13: 4549–4561.

Neri, P. 2014. Semantic control of feature extraction from natural scenes. *Journal of Neuroscience* 34: 2374–2388.

Pearson, J., C. Clifford, and F. Tong. 2008. The functional impact of mental imagery on conscious perception. *Current Biology* 18: 982–986.

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

Prinz, J. 2006. Is the mind really modular? In *Contemporary debates in cognitive science*, ed. J. Stainton, 22–36. Malden: Blackwell.

Pylyshyn, Z. 1999. Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22: 341–365.

Pylyshyn, Z. 2003. *Seeing and visualizing*. Cambridge: MIT Press.

Raftopoulos, A. 2009. *Cognition and perception: How do psychology and neural science inform philosophy?* Cambridge MA: Bradford Books.

Ramalingam, N., J. Mcmanus, W. Li, and C. Gilbert. 2013. Top-down modulation of lateral interactions in visual cortex. *Journal of Neuroscience* 33: 1773–1789.

Reddy, L., N. Tsuchiya, and T. Serre. 2010. Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage* 50: 818–825.

Rockland, K., and D. Pandya. 1981. Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain Research* 212: 249–270.

Schapiro, A., E. Gregory, B. Landau, M. McCloskey, and N. Turk-Browne. 2014. The necessity of medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience* 26: 1736–1747.

Sklar, A., N. Levy, A. Goldstein, R. Mandel, A. Maril, and R. Hassin. 2012. Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences* 109: 19614–19619.

Stokes, D. 2011. Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philosophical Studies* 158: 477–492.

Summerfield, C., and T. Egner. 2009. Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences* 13: 403–409.

Teufel, C., et al. forthcoming. Talk delivered at the cognitive penetration workshop, Bergen, June 2015.

Thierry, G., P. Athanasopoulos, A. Wiggett, B. Dering, and J.R. Kuipers. 2009. Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences* 106: 4567–4570.

Tsuchiya, N., and C. Koch. 2005. Continuous flash suppression reduces negative afterimages. *Nature Neuroscience* 8: 1096–1101.

Vetter, P., F. Smith, and L. Muckli. 2014. Decoding sound and imagery content in early visual cortex. *Current Biology* 24: 1256–1262.

Wu, W. 2013. Visual spatial constancy and modularity: Does intention penetrate vision? *Philosophical Studies* 165: 647–669.

Wyart, V., A. Nobre, and C. Summerfield. 2012. Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences* 109: 3593–3598.

Yang, E., and R. Blake. 2012. Deconstructing continuous flash suppression. *Journal of Vision* 12: 1–14.