

ATTENTION AND ENCAPSULATION

Jake Quilty-Dunn

Faculty of Philosophy & Brasenose College, University of Oxford
Department of Philosophy & PNP Program, WUSTL

Forthcoming in *Mind & Language*
(Penultimate version)

Abstract: The question of whether perception is encapsulated from cognition has been a major topic in the study of perception in the past decade. One locus of debate concerns the role of attention. Some theorists argue that attention is a vehicle for widespread violations of encapsulation; others argue that certain forms of cognitively driven attention are compatible with encapsulation, especially if attention only modulates inputs. This paper argues for an extreme thesis: no effect of attention, whether on the inputs to perception or on perceptual processing itself, constitutes a violation of the encapsulation of perception.

§1—Introduction

The question of whether perception is ‘informationally encapsulated’ from cognition (Fodor, 1983) has loomed large in the study of perception in the past decade. The issue is debated amongst psychologists (e.g., Lupyan, 2015a, 2015b; Firestone & Scholl, 2016) as well as philosophers (e.g., Prinz 2006, unpublished; Macpherson, 2012; Siegel 2012; Clark, 2013; Mole, 2015; Block, unpublished). Much of the disagreement involves differing interpretations of specific studies, often focused on nitty-gritty details of experimental design (e.g., Durgin et al., 2009; Proffitt, 2009; Gantman & van Bavel, 2014; Firestone & Scholl, 2014). A significant amount of disagreement, however, concerns relatively abstract disputes about which mental processes constitute violations of encapsulation.

A major point of contention concerns the role of attention. Some theorists argue that effects of cognition on perception that are mediated by certain forms of attention are compatible with encapsulation (Deroy, 2013; Firestone & Scholl, 2016). Others write that attention is a vehicle for widespread, systematic violations of encapsulation (Mole, 2015;

Lupyan, 2015a, 2016; Wu, 2017; Prinz, unpublished).¹ Still others occupy intermediate positions (e.g., Macpherson, 2012). I'll argue for an extreme thesis: no effect of attention, either on the inputs to perceptual processing or on perceptual processing itself, constitutes a violation of the encapsulation of perception. Encapsulated processes have proprietary stores of information; a cognitive state can alter a perceptual process via attention without being thereby accessed and computed over by the perceptual process, thus leaving its proprietary store of information intact. Therefore, cognitively driven attention does not violate the encapsulation of the processes it modulates. The upshot of this thesis is not only that all effects mediated by attention are compatible with encapsulation, but also that, *contra* current practice, debates about the encapsulation of perceptual processing should be separated from debates about the theory-ladenness and normative force of perception.

§2—Encapsulation

Encapsulation is a matter of information access. Jerry Fodor writes that an encapsulated system 'does not have access to all of the information that the organism internally represents; there are restrictions upon the allocation of internally represented information to input processes' (1983, p. 69; see also Scholl, 1997). Saying that System A is encapsulated from System B means that A is constrained (in some way to be spelled out later) from accessing information in B. The notion of information access is foundational to the computational theory of mind. Every computational system must have some information available to it. Computational systems transform inputs into outputs, and what determines the relevant input-output mappings is the information available to the system together with the algorithms it implements. Part of what distinguishes computational processes from mere transductions is that the former do, and the latter don't, bring stored information to bear on their transformations such that the output is not merely a slavish registration of the input (Pylyshyn, 1984, Chapter 6).

As an illustrative example, take visual 'shape from shading,' a process that uses information about the shading of an object's surface to derive its three-dimensional shape (Ramachandran, 1988). The system that implements this process seems to 'assume' that light comes from above and slightly to the left (Sun & Perona, 1998). Somehow, the information that light comes from above and to the left guides the mapping from shaded two-dimensional surfaces to three-dimensional shapes performed by the shape-from-shading system. Talk of the

¹ Some of these authors frame the issues in terms of cognitive penetration rather than encapsulation, typically assuming that an encapsulated system is impenetrable and *vice versa*. I will argue below that the relation between these ideas is more complicated.

system making an ‘assumption’ is a metaphorical way of stating that a certain piece of information structures the relevant computations.

This structuring can take place either explicitly or implicitly. In the explicit case, a representation with the content <light comes from above and slightly to the left>—or, more likely, some much more precise specification—is stored in a manner that makes it accessible to the shape-from-shading system. When the system is fed its inputs, it accesses this representation (and other stored information) and derives its outputs in accordance with whatever algorithms it implements. The notion of access involves the activation of a token representation and its functioning directly and without independent intervention in the relevant processes.

In the implicit case, on the other hand, there’s no such representation stored anywhere. Instead, the algorithms implemented by the system accord with the principle that light comes from above and to the left. That information isn’t available for use by any process; it’s merely implicit (or ‘embodied’—e.g., Devitt, 2005, p. 45) in the operations of the system (Shea, 2014). For this piece of information to be ‘accessed’ is simply for it accurately to describe the transformations demanded by the relevant algorithms.

Note that ‘explicit’ in this context does not mean ‘conscious’. Some (e.g., Fodor, 1981) argue that rules of grammar are explicitly stored in the language faculty and are accessed in deriving the syntactic properties of linguistic expressions, but would nonetheless deny that these representations are ever conscious. The present use of ‘explicit’ thus differs from the use of that term in discussions of implicit vs. explicit attitudes, such as the distinction between implicit biases and conscious attitudes. On a view like Mandelbaum’s (2016), for instance, implicit biases are explicit in the sense that they are realized by mental representations that are directly available for use in various mental processes, and are implicit only in the (very different) sense that they are unconscious.²

We can loosely use the term ‘store of information’ to describe both the explicitly and implicitly stored information used by a system to structure its input-output mappings. This term is loose in that implicitly stored information, unlike explicitly stored information, is not literally stored anywhere. Similarly, implicitly stored information, unlike explicitly stored information, is not literally accessed while running computations within the relevant system.

² Though ‘implicit’ in the sense of being unconscious does not entail ‘implicit’ in the sense of being unavailable for use in any mental process, the latter will entail the former given the reasonable assumption that only token representations can be conscious. Any token representation fails to be implicit in the latter sense, and thus the scope of explicit (in the sense of conscious) representation is limited to explicit (in the sense of available) representation.

Nonetheless, it's part of the set of information that shapes the processing of a given system, and can thus be counted as part of its store of information.

It's useful to have a neutral way of talking given how routinely difficult it is to determine whether a piece of information is stored explicitly or implicitly. The case of light coming from above and to the left demonstrates this methodological problem. Is that information an explicit or implicit part of shape-from-shading processing? This is not a rhetorical question, but it's difficult to know how to go about answering it (though the malleability of the light-from-above assumption in response to feedback from other modalities might suggest explicit storage—see Adams et al., 2004). Another illustrative case is generative grammar (Chomsky, 1986; Devitt, 2005; Pietroski, 2008). Critics notwithstanding, generative linguists believe that the rules of grammar are part of the language faculty's proprietary store of information. There is, however, no consensus on whether this information is explicitly stored.

With all this in the background, we can give a more precise characterization of encapsulation: System A is encapsulated from System B when A has a proprietary store of information that excludes information stored in B. While operating on its inputs to produce its outputs, A can only access a certain limited domain of information, and a piece of information held in B—even if it's semantically highly relevant to the processing in A—simply fails to fall within that domain, and thus cannot be accessed by A.

A passage from Fodor usefully illustrates this point:

One can conceptualize a module as a special-purpose computer with a proprietary database, under the conditions that: (a) the operations that it performs have access only to the information in its database (together, of course, with specifications of currently impinging proximal stimulations); and (b) at least some information that is available to at least some cognitive process is not available to the module.

(Fodor, 1985, p. 3)

Encapsulation in this sense underwrites the modularist explanation of recalcitrant perceptual illusions: despite the fact that a piece of information in cognition is semantically relevant to the illusory percept (e.g., one knows that the stimulus is illusory), that information falls outside the proprietary store of information available to perception, and thus the illusion persists.

A primary reason encapsulation matters for cognitive science is that it provides a good reason for positing distinct systems. If two processes occur in the same system, then semantically relevant information that is accessible to one process should typically be accessible to the other as well. For example, guessing where a person is from and deciding whether to ask them for directions are distinct cognitive processes, but a single piece of information (e.g., that they only speak Finnish) can be relevant to both processes.³ In this example, when told that the person speaks Finnish you might form only a single representation, THAT PERSON ONLY SPEAKS FINNISH, which is stored in central cognition and is accessible to both processes (e.g., it might lead you to guess that they are from Finland and, if you don't speak Finnish, to decide not to ask them for directions). This sort of case would provide a compelling reason to think that the processes of guessing where a person is from and deciding whether to ask them for directions both take place in the same central-cognitive system.

If, on the other hand, two processes cannot access the same store of information, then there is good reason to think they don't take place in the same system.⁴ It's for this reason that Firestone and Scholl, for example, take the idea of widespread violations of the encapsulation of perception from cognition to threaten the idea of 'a salient "joint" between perception and cognition' (2016, p. 3). Some theorists on the other side of the debate agree. Clark, for example, writes that the putative prevalence of cognitive penetration 'makes the lines between perception and cognition fuzzy, perhaps even vanishing' (2013, p. 190). Others, such as Block (unpublished), argue that cognitive penetration is widespread but that perception is nonetheless distinct from cognition, appealing to factors like differences in representational format (cf. Quilty-Dunn, 2016; forthcoming; Mandelbaum, forthcoming). And even independently of the perception–cognition border, developing a complete account of the

³ An anonymous referee questions whether these are really distinct processes, since they both seem to be instances of decision-theoretic reasoning. I take them to be distinct processes only in the minimal sense that they are different computational events that could occur independently of one another and could, in principle, occur in different systems. They may involve the same sort of underlying processing at some level of description.

⁴ The usefulness of encapsulation for individuating systems is dependent on other factors. For example, encapsulation may fail to be sufficient for dividing systems if systems share many other features. There may be encapsulated information stores within central cognition, for example, such as information used for processing rules of social exchange (Cosmides et al., 2010). The independence of this processing from stimulation, its relatively abstract content, and the use of its inputs and outputs for other forms of reasoning may suffice to place it within central cognition rather than as constituting a wholly distinct system. The individuation of systems is determined by larger theoretical concerns as well. For example, consider hierarchical predictive coding models wherein perceptual processing consists of layers each of which can only receive information from the layers immediately above and below it. Technically speaking, each layer is therefore informationally encapsulated. However, the wider picture of mental architecture presupposed by such accounts essentially obviates the explanatory usefulness of encapsulation as a means by which to individuate mental systems (indeed, the very idea of distinguishing mental systems is often questioned by proponents of predictive coding—see Lupyan, 2015b).

computational structure of perception requires taking a stand on whether perceptual processes are encapsulated.

The hypothesis that perceptual processes are encapsulated from cognition thus offers to solve at least two open questions about the nature of perception. The first question is: What grounds the perception–cognition border? If perceptual processes have proprietary stores of information that exclude information stored in cognition, then that may provide a basis for distinguishing perceptual processes from cognitive ones (most likely in conjunction with other distinguishing features). This encapsulation-based account of the perception–cognition border allows for perception and cognition to share other features (e.g., representational format), and is consistent with effects of cognition on perception that allow perceptual processes to retain proprietary stores of information. The second question to which the encapsulation of perception can provide an answer is: How are perceptual processes computationally tractable? Perception requires tortuous computational problems (such as the underdetermination of 3D surface layout by 2D retinal stimulation) to be solved at breakneck speed. If the processes that solved these problems had to sift through all information stored in central cognition, they would face an unwieldy computational burden. Indeed, models of centralized reasoning face the notorious ‘frame problem’ because of the supposedly global (i.e., unencapsulated) reach of information access in central cognition (Fodor, 1983, p. 112ff, 2000; cf. Pinker, 2005). If instead perceptual processes are encapsulated, then they need only check input against their proprietary stores of information and thereby avoid the problems of global computational availability that plague theories of central cognition. Encapsulation can therefore provide a unified account of perceptual processes as computationally tractable operations that occur outside of central cognition.

In what follows, I’ll consider the significance of cognitively driven attention for the encapsulation of perceptual processing. Before doing so, I note that the question of whether a cognitive state violates the encapsulation of perception (e.g., by means of attention) is a question about *explicit* information access; talk of information storage/access below will thus be restricted to explicit storage/access unless otherwise noted. As noted above, the notion of implicit information is figurative—it does not involve literal storage or access. What justifies this metaphorical talk is that we often have to characterize a process (e.g., shape-from-shading) without knowing whether the relevant information is explicitly stored or not. In the case of top-down effects of cognition on perception, however, the relevant information is an explicit representation stored in cognition such as a belief or an intention. There is in that case no explanatory need to invoke the notions of implicit storage and access. Instead, the substantive outstanding question is whether the explicit representation in cognition is accessed by perception.

§3—Attention

3.1 Attention and perception. Attention is the selection or prioritization of information for further processing (though how this should be understood is a matter of debate—see, e.g., Nanay, 2010; Mole, 2011; Prinz, 2012; Wu, 2014; Fazekas & Nanay, forthcoming). Attention can consist in some information being processed rather than other available information, or in constrained forms of modulation, such as increasing ‘gain,’ i.e., signal strength, and ‘tuning,’ i.e., noise reduction (Ling et al., 2009). Attentional modulation can increase spatial resolution (Carrasco & Barbot, 2014), boost perceived contrast (Carrasco et al., 2004), and warp perceived distance between items (Vickery & Chun, 2010).

In the visual system, attention comes in three broad forms. *Spatial* attention selects information from certain spatial locations and inhibits other locations. Spatial attention can be overt, as when you move your eyes to change your gaze, or covert, as when you maintain fixation on a single point but attend to something in your periphery. *Feature-based* attention selects representations of certain features and inhibits others; for example, you might search for your favorite red shirt in your drawer and attend to red over the other available colors. Finally, *object-based* attention selects representations of certain objects; for example, you might watch a flock of birds and attend to a particular bird, tracking it as it moves.

Cognition can direct all three varieties of visual attention. For instance, you might attend to a certain location because you expect something to appear there; you might attend to a certain feature because you want to find your shirt; and you might attend to a certain object because it’s a bird you intend to track. One might be tempted to conclude that these are all instances of cognition violating the encapsulation of perception.⁵

⁵ It is important to distinguish attentional top-down effects from effects of expectation. For example, if an association is learned between two stimuli, then presenting one may activate an expectation of the other, such that processing of that stimulus will be facilitated and/or failure of that stimulus to appear will disrupt processing. Such expectations can plausibly occur within perceptual processing and cannot obviously be reduced to attentional modulation (see Summerfield & Egner, 2009 for an overview). In principle, the expectation could take the form of a cognitive state such as a belief that two stimuli A and B tend to co-occur. Upon perceiving A, some perceptual process might access this belief and thereby more readily process B, or less readily process stimuli that are incongruent with B. This would be a failure of encapsulation, but one that was explained in terms of cognitive expectation affecting perception rather than cognitively driven attention. Bar et al. (2006) provide evidence that low-spatial-frequency stimuli cause top-down processes originating in orbitofrontal cortex to facilitate processing in visual cortex. Summerfield and Egner (2009) take this sort of effect to be an expectation-driven (as opposed to attentional) effect on visual processing. They argue that top-down attentional modulation originates primarily in the dorsolateral prefrontal cortex and posterior parietal cortex, and that the distinct neural origins of top-down expectation effects suggest a distinct mechanism from attention. The arguments below that attentional effects are not failures of encapsulation are compatible with the claim that this effect shows that

Firestone and Scholl (2016) argue instead that at least some forms of attention merely constitute changing of inputs to otherwise encapsulated perceptual systems. An example would be when you choose to change where your eyes are fixated; though this is a cognition-driven change in perception, it operates not by violating encapsulation but instead by moving the eyes to change the inputs to perception. Though covert spatial attention doesn't literally involve moving the eyes, it's intuitively the same sort of effect and thus changes perception only by changing inputs (see also Deroy, 2013).

Fiona Macpherson (2012) raises the possibility that feature-based attention directed by cognition seems to violate encapsulation (see also Mole, 2015; Block, unpublished). Unlike spatial attention, it's not simply a change in where we are looking, and it seems moreover to require antecedent processing (e.g., it seems to require processing of red in order to attend to the red things in the room). A modularist may reply that attention does not only select *sensory* inputs, but also selects *representations* as inputs for downstream processing, even if those representations were the result of earlier processes.⁶ Others argue that attention has widespread effects on perceptual processing that are not limited to prioritization and modulation of inputs (e.g., Mole, 2015; Lupyan, 2016; Clark, 2016). Even Firestone and Scholl only take their argument to save encapsulation from 'peripheral' (i.e., input-modulating) forms of attention, and say they are 'sympathetic' (2016, p. 61) to an attack on encapsulation that focuses on non-peripheral forms of attention. They object only that such proposals have at present remained too abstract relative to the experimental evidence, thus leaving encapsulation-based theories of perception vulnerable to detailed accounts of non-peripheral forms of top-down attention.

3.2 Attention and encapsulation. A methodological problem arises here: how do we determine which attentional effects might, and which might not, violate encapsulation? The debate, so stated, threatens to become about warring intuitions. The idea that covert spatial attention fails to violate encapsulation seems to rest on an intuitive parallel with overt spatial attention (e.g., Macpherson, 2012, p. 28). Theorists like Macpherson (2012), Prinz (unpublished), and Block (unpublished) invite us to have the opposite intuitions about feature-based and (in Prinz's case) object-based attention, since these forms of attention seem to involve a tighter connection between the contents of our intentions and the contents of our resultant perceptual

cognitive expectations violate the encapsulation of vision (Ogilvie & Carruthers, 2016). More recently, however, Bang and Rahnev (2017) and Rungratsameetaweemana et al. (2018) have provided evidence that expectation only directly effects post-perceptual processing and that top-down modulation is really best understood in terms of attention rather than perceptual computation over cognitive expectations.

⁶ This reply requires the assumption that the visual system consists of a hierarchy of modules, with encapsulated subsystems whose outputs function as inputs to other encapsulated subsystems.

states. Indeed, Pylyshyn's classic formulation of cognitive penetration is of a 'semantically coherent' causal relation between cognition and perception (1999, p. 342), and the notion of semantic coherence is hard to explicate without intuitions. For example, feature-based attention to red things that alters perception of redness in response to thoughts about redness seems to instantiate a causal, semantically coherent relation (MacPherson, 2012; Block, unpublished).

Pylyshyn's influential formulation of cognitive penetration in terms of semantically coherent causal relations between thought and perception thus threatens to make mundane forms of attention into examples of cognitive penetration. Note, however, that the presence of a semantically coherent causal relation between a thought and the output of some perceptual process says nothing about stores of information. For all that Pylyshyn's informal definition of cognitive penetration says, there could be semantically coherent causal relations between thoughts and percepts without those thoughts being part of the store of information available to the relevant perceptual process.⁷ Typical formulations of cognitive penetration in the literature therefore fail to track the notion of violation of encapsulation.

Encapsulation is a formal, architectural notion, entailing that the only token representations accessible to a process fall within a proprietary store. Cognitive penetration is a semantic notion, entailing that the content of some state in cognition influences some process such that the content of the output of the process bears some interesting semantic relation to

⁷ Raftopoulos and Zeimbekis (2015, p. 27ff) argue that Pylyshyn (1999, 2003) endorses a 'vehicle criterion,' according to which 'perception has to draw directly on the informational resources of a cognitive system' (Raftopoulos & Zeimbekis, 2015, p. 27) to constitute cognitive penetration. However, the passages they cite merely specify that attention cannot operate '*within* the early-vision system itself' (Pylyshyn, 2003, p. 90; see also Pylyshyn, 1999, p. 344), a specification that is sometimes cast in terms of 'directness' (Gross, 2017; Block, unpublished). Cognitively driven attentional modulation that alters 'the early-vision system itself' in a semantically coherent way would therefore seem to suffice for cognitive penetration by Pylyshyn's lights. Like Firestone and Scholl, Pylyshyn's argument against violations of encapsulation by attention rests on the empirical assumption that attention operates only on the inputs and outputs of perceptual systems rather than within the operations of those systems.

Pylyshyn does sometimes cast impenetrability in terms of information access (e.g., 1999, p. 360), and Gross (2017) stipulates that cognitive penetration involves 'directness' in terms of operation within perceptual processing as well as 'directness' in terms of information access. A primary goal of this paper is to pull these various ideas apart and argue both that encapsulation should be understood solely in terms of information access and that attentional effects do not violate encapsulation so understood. The notion of cognitive penetration in terms of semantically coherent effects (even building in 'directness' in terms of operation within perceptual processing rather than merely on inputs) is distinct from encapsulation and has normatively significant upshots. *Pace* Gross, this notion of cognitive penetration should therefore be articulated independently of questions about information access and encapsulation. The hypothesis that the content of thought can exert some degree of control over the content of perception may be normatively significant even if it is compatible with the encapsulation of perception—see Section 4 for discussion.

the content of the cognitive state. The content of a token representation can influence some process without that representation being computed over. Perhaps, as in cases mediated by top-down imagery, the cognitive state causes the formation of a separate imagistic representation that shares elements of its content and enters into a perceptual process; the cognitive state is not itself computed over by the perceptual process despite the fact that its content influences that process via imagery (Macpherson, 2012; cf. Wu, 2013). Cognitive penetration (*qua* semantic notion) thus doesn't entail a violation of encapsulation (*qua* formal notion).

One may worry that distinguishing the semantic notion of cognitive penetration and the formal notion of encapsulation may render the defense of encapsulation offered here irrelevant to the concerns of theorists who tend to frame the impact of attention in terms of cognitive penetration. However, in distinguishing violations of encapsulation from cognitive penetration, I intend to disentangle a conflation made by virtually all theorists working on the topic of cognitive effects on perception. Even Pylyshyn, for example, takes 'encapsulation' and 'impenetrability' to be interchangeable terms (1999, p. 344). Moreover, theorists who *explicitly* take the idea of attentionally mediated cognitive penetration to entail a failure of encapsulation include Block (unpublished), Lupyan (2015, p. 549), Prinz (unpublished), Stokes (2014, p. 3) and Wu (2017, p. 6ff). As far as I can tell, not a single theorist who argues for attention as a mechanism of cognitive penetration is in favor of anything like Fodorian modularity in perception.⁸

If a representation is held in a system's proprietary store, it is directly available to be accessed in running computations. No additional process is necessary for it to affect processing. Every computational system can immediately access its proprietary information store either implicitly—i.e., running algorithms that accord with it—or explicitly—i.e., activating the representation and having it enter into the online computational process. Access of proprietary

⁸ A possible exception may be Macpherson, who argues for attentionally mediated cognitive penetration of perceptual experience and points out that this is compatible with the existence of encapsulated early perceptual systems (2012, pp. 26–29). However, Macpherson separates these hypotheses only because one concerns conscious experience and the other concerns early perceptual processing which may fail to determine experience; she doesn't allow for a single process (either the generation of experience or early perceptual processing) to be simultaneously cognitively penetrable and informationally encapsulated. Mole (2015) does not explicitly mention Fodor or encapsulation as such, but given the ambient conflation of cognitive penetration and violation of encapsulation in the literature, one might naturally read him as arguing against the encapsulation of perception as well as its impenetrability. His emphasis on the continuity between perception and cognition, and his characterization of perceptual processes as merely 'a subset of the system that enables working memory to do its work' (Mole, 2015, p. 235), suggest that he takes his arguments for cognitive penetrability to undermine a robust perception–cognition distinction of the sort licensed by the encapsulation of perception from cognition.

information is direct or immediate in the sense that there's no intervening process except (in the explicit case) the activation of the stored representation.

Consider, for example, the access of a belief stored in long-term memory in central cognition. Barring some sort of malfunction or independent intervention, once the belief is accessed (which, since it's explicitly stored, amounts to activating the representation), it will enter into central cognitive processes like inference (Braine & O'Brien, 1998; Quilty-Dunn & Mandelbaum, 2018) and facilitating lexical recognition (Meyer & Schvaneveldt, 1971). The fact that representations stored in a system currently running some computation are immediately entered into that computation *via* access explains why, for example, subliminally presenting the minor premise in a *modus ponens* argument after consciously presenting the major premise results in acquisition of the conclusion (Reverberi et al., 2012). Assuming that *modus ponens* describes an algorithm implemented in central cognition, then once the major premise (e.g., 'If there's an X then there's a Y') is represented in central cognition, activating the minor premise (e.g., 'There's an X') is causally sufficient to yield the conclusion as output (e.g., 'There's a Y'). The representation, once accessed, is integrated with currently running processes to yield an output.

It may be clear at this point why an effect of cognition on perception mediated by attention cannot violate encapsulation: it is not a form of information access. A state in cognition may cause an attentional effect on perceptual processing and thereby alter the output of perception. Moreover, it may do so in a semantically coherent way, such that the output of perception is semantically related to the state in cognition. But the state in cognition isn't thereby just another representation housed in the perceptual system's store of information. Instead, it can only exert an influence on the perceptual system by way of attentional prioritization and modulation.⁹ If a representation in cognition can only affect perceptual processing *via* attention, that immediately separates it from representations in the relevant proprietary information store; the latter are accessed directly, while the former isn't. States in cognition that can only affect perception *via* attention are thus never part of the store of

⁹ An anonymous referee points out that this line of reasoning may apply to some other putative cases of encapsulation failure. One example might be an effect of cognition on perception mediated by imagery (Macpherson, 2012). A cognitive state that affects perception by forming a mental image which then affects perceptual processing alone is not necessarily accessed by the perceptual process. Whether top-down effects involving mental imagery are always mediated in this way is an empirical question, and not one that can be addressed here (e.g., one might argue that mental images can include cognitive states as constituents, and thus could suffice for a cognitive state's being computed over by central cognition).

information proprietary to a perceptual system, and so don't violate the encapsulation of perception from cognition.¹⁰

Take the famous image by R.C. James in Figure 1, which to an innocent eye may appear to contain a chaos of dots. However, upon learning that the image depicts a Dalmatian, the dog may become visible. Suppose that this effect works through activating your concept of Dalmatians and using conceptual knowledge of what they look like to guide which locations and features you attend to, thereby affecting shape perception in a semantically coherent way. So described, there is no reason to say additionally that this cognitive information is *accessed* by the shape-perception process. The cognitive states instead guide spatial, feature-based, and, ultimately, object-based attention, and these forms of attention modulate perceptual processing. The store of information proprietary to the relevant shape-perception processes need not be expanded to incorporate those cognitive states.

¹⁰ One can also think of information access as defined by the array of parameters that can be set in a computation. For example, distance perception may contain a parameter for whether the perceived object is desired (Balctis & Dunning, 2010)—e.g., if the object is desired, then the parameter is set to YES. Effects mediated by attention don't involve a parameter being set directly by the presence of a state in cognition, so such effects don't violate encapsulation.



FIGURE 1 Top-down attention at work (taken from Gregory, 2005, by permission of Royal Society)

Christopher Mole (2015) cites Kravitz and Behrmann (2011) as providing experimental evidence for top-down attention operating within (rather than before) perceptual processing. Kravitz and Behrmann found effects of attentional cues on a letter-discrimination task that were driven by the shape, color, and location of previewed objects. A cue would improve letter discrimination not only if the letter appeared at the same location or in the same object as the cue, but also if it appeared in the location of another object that shared features

(e.g., shape or color) with the object in which the cue appeared. Mole (as well as Kravitz and Behrmann) argues that spatial, feature-based, and object-based attention conspire to organize perceived scenes and guide perceptual discriminations. Mole argues further that this interaction can be sensitive to learned categories such as letter identities (2015, pp. 230–231).¹¹

As mentioned above, a modularist could reply that these interactions of attentional effects take place at the output-input interfaces of hierarchically nested modules. But the point presently at issue is that encapsulation isn't violated *even if* attention operates within modular processing. Attention is the prioritization and modulation of perceptual signals, not a form of information access. Prioritization and modulation within modules may, for example, affect which stored information is accessed in online perceptual processing by biasing competition between signals for 'representation, analysis, or control' (Desimone & Duncan, 1995, p. 194; see Mole, 2015, p. 231ff). The cognitive states that drive such attentional effects, however, are not thereby among the information accessed and computed over in the relevant perceptual processes.

One might object that attention simply *is* perceptual processing. Mole comes close to this position, writing that attention is 'inherent to the basic structure of the neural architecture' of perception and thus cannot be viewed as a separate mechanism (2015, p. 233). If attention is a perceptual process, one might argue, then cognitively driven attention is *ipso facto* a case where a perceptual process accesses cognition (see also Wu, 2017). But this claim, depending on how it's meant, is either trivial or false. If the claim is merely that attention (*qua* perceptual process) accesses information in cognition, then it's trivial; nobody on any side of the debate about encapsulation would deny that attention can be driven by cognitive states. If, on the other hand, the claim is that the independent perceptual processes modulated by attention thereby access information in cognition, then it's false. There is a clear difference between (i) a process that accesses representation R and (ii) a process that's affected by a separate process that accesses R. The latter process does not access R, and therefore does not have R within its store of information.

The argument in this paper thus does not require arguing that attention is a separate system from perception, but only that attention is not identical to the processes (e.g., distance

¹¹ It's worth pointing out here that perceptual sensitivity to learned categories is compatible with encapsulation. The core idea of encapsulation is that a process has a proprietary store of information at any given moment. Perceptual learning allows a transformation of stored perceptual information over time (Goldstone, 1998). This learning can be bottom-up or even top-down, allowing concepts to make their way into perceptual information stores (Quilty-Dunn, 2017; Mandelbaum, forthcoming). The notion of encapsulation explored in this paper is *synchronic* (at a single time) rather than *diachronic* (across time), and is thus compatible with diachronic top-down influences on perceptual information stores.

perception) it modulates. Since attention *modulates* such processes, it cannot be *identical to* such processes, even if it is genuinely part of the perceptual system and is deeply intertwined with the processes it modulates.¹² Perception of contrast, for example, can be modulated by attention (Carrasco et al., 2004). Suppose for the sake of argument that this attentional modulation dramatically alters the perception of contrast to such an extent and with such regularity that contrast is never computed independently of attentional modulation. Attention may therefore be ‘inherent to the basic structure of the neural architecture’ of the perception of contrast. Nonetheless, the perceptual computation of contrast is not identical to the attentional modulation of that computation. If they were identical, it would not make sense to say that the one modulates the other. The fact that attention can modulate perceptual processing and can be driven by cognitive states therefore does not entail that the modulated perceptual processing accesses and computes over those cognitive states.

Consider a putative case where a cognitive state is genuinely accessed by perception. For example, suppose that a visual subsystem that implements perception of distance can access a desire for the perceived object and thereby make it appear closer (Balcetis & Dunning, 2010). In that case, the store of information accessible to the distance-perception system includes desires stored in central cognition, and these desires can be computed over by the distance-perception system. Suppose instead, however, that the desire guides attention such that you attend more to things you want, and the perceptual signal is thus boosted (even, we can stipulate, at a point within the relevant perceptual process rather than at the input stage). Boosting the signal then causes the object to appear larger and closer. In the second case, there’s no explanatory reason to say the desire is part of the store of information accessible to the distance-perception system. The only ‘perceptual process’ to which the desire is accessible is the deployment of attention.

It turns out that Firestone and Scholl’s question of whether attention is merely a changing of inputs is largely beside the point. Even if top-down attention operates *within* the processing of a module, it doesn’t violate the encapsulation of that module. The module can still retain a proprietary store of information that affects its outputs without being mediated by attention or any other mechanism (aside from information access). There can thus, in

¹² Mole, as well as Wu (2013, 2017), may balk at the claim that attention modulates perception, instead holding that attention is an emergent feature of integrated processes oriented toward some task rather than a mechanism of modulation (or indeed a mechanism at all). However, Mole and Wu would still allow that top-down biasing of competition modulates perception, in which case the matter can be reformulated as being about top-down biasing rather than attention *per se*. While the question of whether perception is modulated by attention itself or something else (e.g., top-down biasing) is centrally important for understanding the metaphysics of attention, it is orthogonal to the question of encapsulation since neither way of talking entails that encapsulation is violated. For ease of exposition, I assume here that attention does indeed modulate perception.

principle, be widespread effects of top-down attention on ‘every stage and level of processing’ (Clark, 2016, p.23) without violating the encapsulation of perception.

§4—Larger Issues

A reader familiar with the philosophical literature on cognitive penetration might come away with the following question: Who cares about informational encapsulation? The idea that cognition affects the outputs of perception has serious consequences for epistemology (e.g., Siegel, 2012; Deroy, 2013), aesthetics (Stokes, 2014), the theory-observation distinction (Churchland, 1988), and even the role of implicit bias in police shootings (e.g., Correll et al., 2002). The technical notion of encapsulation, on the other hand, is purely a construct of theories of mental architecture. The thesis that perceptual systems are distinguished from cognitive systems because the former are encapsulated from the latter is important for answering foundational questions in cognitive science about the basic structure of the mind (Quilty-Dunn, 2017; forthcoming; Mandelbaum forthcoming). It’s far from obvious that this thesis should be expected to do important work in epistemology, ethics, or the theory-observation distinction (*contra* Fodor, 1984).

Some theorists (e.g., Stokes, 2014) advocate for a merging of all these various interests in adjudicating putative cases of cognitive penetration. One of the upshots of this paper is that these interests should not be so merged in evaluating the informational encapsulation of perception. Widespread effects of top-down attention may not threaten the informational encapsulation of perceptual systems, but they may nonetheless have serious consequences for the epistemic status, theory-ladenness, and moral import of perception. A percept distorted by top-down attention guided by the belief that *p* may be used to justify the belief that *p*, for example, thereby generating the kind of circular patterns of justification that worry Susanna Siegel (2012). Theorists working on these topics have often found Pylyshyn’s notion of cognitive penetration as a semantically coherent impact of the content of thought on the content of perception to be useful in characterizing top-down effects that have this sort of normative significance. It may turn out, however, that the epistemic significance of such effects does not require cognitive penetration in Pylyshyn’s sense (Siegel 2013, p. 244), in which case the notion of semantically coherent causal effects may fail to play any philosophically important role.

There are surely constraints on the ways in which cognition can affect perception *via* attention. For example, cognitively driven attention can enable you to see the Dalmatian in Figure 1; but while you may associate Dalmatians with firehouses, attention alone cannot make you see a firehouse in the image. This sort of limitation means attention does not completely

erase the epistemic role of perception in constraining what it is reasonable to believe about your environment. But cognitively driven attention can still wreak epistemic havoc. For example, a fearful perceiver might boost signals pertaining to snake-like features of ambiguous stimuli, resulting in a percept as of a snake (Prinz & Seidel, 2012). This instantiates a circular pattern whereby a fear-state influences perception, and the percept then purports to justify the fear-state (Siegel, 2012, 2013).

The debate about top-down effects has proceeded in a way that ties together questions about information access with questions about whether perceptual processing is influenced by cognition in theoretically and normatively significant ways (cf. Mandelbaum, 2018). Dustin Stokes's (2014) 'consequentialist' methodology that ties together epistemically significant effects with violations of encapsulation is a particularly explicit example, but the tendency runs subliminally through much of the literature (including, at times, Fodor and Pylyshyn). Questions about whether cognition penetrates perception—i.e., whether cognition has a semantically coherent effect on how the perceptual process operates—are substantively different from questions about whether cognition violates the encapsulation of perception—i.e., whether cognitive states are accessed and computed over by perceptual processes. Running the two together creates unnecessary confusion.

Effects mediated by attention need not violate encapsulation to 'demonstrate an epistemologically significant way in which the mechanisms that are responsible for our thinking and judging are entwined with the mechanisms that are responsible for our perception of the world' (Mole, 2015, p. 219). And in cases where top-down attention is shaped by our theories and biases, the consequences may be of moral and theoretical concern as well. Research on these topics shouldn't be beholden to questions about whether perceptual systems can be architecturally distinguished from cognition by appeal to informational encapsulation.

By the same token, questions about encapsulation and mental architecture should not be beholden to intuitions about the normative significance of interactions between cognition and perception. What matters instead is the structure of those interactions—specifically, whether they involve a genuine expansion of the computational reach of perception to include what we think and feel, or whether they are simply cases of cognitively driven modulation of encapsulated perceptual processes. Encapsulation remains a topic of foundational interest to cognitive science since it can explain how perceptual processes avoid problems of computational intractability and can be neatly distinguished from cognition by virtue of their proprietary information stores. It is thus of paramount importance for theories of mental architecture that the various forms of top-down attention, despite their enabling the content

of perception to be sensitive to the content of thought, are compatible with the encapsulation of the perceptual processes they modulate.¹³

¹³ Thanks to the audience at the European Society for Philosophy and Psychology and to the organizers of the Minds Online Conference, especially E.J. Green and Wayne Wu for their insightful and challenging comments. Thanks also to Ned Block, Dan Harris, Zoe Jenkin, Eric Mandelbaum, Jesse Prinz, and Nick Shea for comments on an earlier draft, and to three anonymous reviewers for *Mind & Language*. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 681422.

References

- Adams, W.J., Graf, E.W., & Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, 7(10), 1057–1058.
- Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as closer. *Psychological Science*, 21(1), 147–152.
- Bang, J.W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, 7(17072), doi:10.1038/s41598-017-16885-2.
- Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hämäläinen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *PNAS*, 103(2), 449–454.
- Bhalla, M., & Proffitt, D.R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 1076–1096.
- Block, N. (Unpublished). *The border between seeing and thinking*.
- Braine, M.D.S., & O'Brien, D.P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Erlbaum.
- Carrasco, M., & Barbot, A. 2014: How attention affects spatial resolution. *Cold Spring Harbor Symposia on Quantitative Biology*, 79, 149–160.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7(3), 308–313.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Westport, CT: Praeger.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.
- Correll, J., Park, B., Judd, C.M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314–1329.
- Cosmides, L., Barrett, H.C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *PNAS*, 107(2), 9007–9014.
- Deroy, O. (2013). Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies*, 162(1), 87–107.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
- Devitt, M. (2005). *Ignorance of language*. New York: OUP.

- Durgin, F.H., Baird, J.A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, 16(5), 964–969.
- Fazekas, P., & Nanay, B. (Forthcoming). Attention is amplification, not selection. *The British Journal for the Philosophy of Science*.
- Firestone, C. & Scholl, B.J. (2014). Enhanced visual awareness for morality and pajamas? Perception vs. memory in ‘top-down’ effects. *Cognition*, 136, 409–416.
- Firestone, C. & Scholl, B.J. (2016). Cognition does not affect perception: Evaluating the evidence for ‘top-down’ effects. *Behavioral and Brain Sciences*, 39, 1–72.
- Fodor, J. (1981). Introduction: Some notes on what linguistics is talking about. In N. Block (Ed.), *Readings in the philosophy of psychology*, vol. 2. Cambridge, MA: Harvard University Press, 197–207.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, 51(1), 23–43.
- Fodor, J. (1985). Précis of The Modularity of Mind. *Behavioral and Brain Sciences*, 8(1), 1–42.
- Fodor, J. (2000). *The mind doesn't work that way*. Cambridge, MA: MIT Press.
- Gantman, A.P., & Van Bavel, J.J. (2014). The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. *Cognition*, 132, 22–29.
- Goldstone, R. (1998). Perceptual learning. *Annual Review of Psychology* 49, 585–612.
- Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology* 8(221), 1–12.
- Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, 49(10), 1194–1204.
- Lupyan, G. (2015a). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, 6(4), 547–569.
- Lupyan, G. (2015b). Reply to Macpherson: Further illustrations of the cognitive penetrability of perception. *Review of Philosophy and Psychology*, 6(4), 585–589.
- Lupyan, G. (2016). Not even wrong: The ‘it’s just X’ fallacy. *Behavioral and Brain Sciences*, 39, 40–41.
- Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24–62.
- Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3), 629–658.

- Mandelbaum, E. (Forthcoming). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*.
- Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.
- Mole, C. (2011). *Attention is cognitive unison*. Oxford: OUP.
- Mole, C. (2015). Attention and cognitive penetration. In J. Zeimbekis and A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives*. Oxford: OUP, 218–238.
- Nanay, B. (2010). Attention and perceptual content. *Analysis*, 70(2), 263–270.
- Ogilvie, R., & Carruthers, P. (2016). Opening up vision: The case against encapsulation. *Review of Philosophy and Psychology*, 7(4), 721–742.
- Pietroski, P. (2008). Think of the children. *Australasian Journal of Philosophy*, 86(4), 657–669.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Prinz, J. (2006). Is the mind really modular? In R. Stainton (Ed.), *Contemporary debates in cognitive science*. Malden: Blackwell, 22–36.
- Prinz, J. (2012). *The conscious brain: How attention engenders experience*. Oxford: OUP.
- Prinz, J. (Unpublished). Faculty psychology without modularity.
- Prinz, J., & Seidel, A. (2012). Alligator or squirrel: Musically induced fear reveals threat in ambiguous figures. *Perception*, 41(12), 1535–1539.
- Proffitt, D.R. (2009). Affordances matter in geographical slant perception: Reply to Durgin, Baird, Greenburg, Russell, Shaughnessy, and Waymouth. *Psychonomic Bulletin & Review*, 16(5), 970–972.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341–365.
- Quilty-Dunn, J. (2016). Iconicity and the format of perception. *Journal of Consciousness Studies*, 23(3–4), 255–263.
- Quilty-Dunn, J. (2017). *Syntax and semantics of perceptual representation*. Ph.D. dissertation, The Graduate Center, CUNY.
- Quilty-Dunn, J. (Forthcoming). Perceptual pluralism. *Noûs*.
- Ramachandran, V.S. (1988). Perceiving shape from shading. *Scientific American*, 256(8), 76–83.
- Reverberi, C., Pishedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta Psychologica*, 139, 244–253.

- Rungratsameetaweemana, N., Itthipuripat, S., Salazar, A., & Serences, J.T. (2018). Expectations do not alter early sensory processing during perceptual decision-making. *The Journal of Neuroscience*, 38(24), 5632–5648.
- Scholl, B. (1997). Reasoning, rationality, and architectural resolution. *Philosophical Psychology*, 10(4), 451–470.
- Shea, N. (2014). Distinguishing top-down from bottom-up effects. In D. Stokes, M. Matthen, and D. Biggs (Eds.), *Perception and its modalities*. Oxford: OUP, 73–91.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs*, 46(2), 201–222.
- Siegel, S. (2013). Can selection effects on experience influence its rational role? In T. Gendler (Ed.), *Oxford studies in epistemology, vol. 4*. Oxford: OUP, 240–270.
- Stokes, D. (2014). Cognitive penetration and the perception of art. *Dialectica*, 68(1), 1–34.
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409.
- Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, 1(3), 183–184.
- Vickery, T.J., & Chun, M.M. (2010). Object-based warping: An illusory distortion of space within objects. *Psychological Science*, 21(12), 1759–1764.
- Wu, W. (2013). Visual spatial constancy and modularity: Does intention penetrate vision? *Philosophical Studies*, 165, 647–669.
- Wu, W. (2014). *Attention*. New York: Routledge.
- Wu, W. (2017). Shaking up the mind's ground floor: The cognitive penetration of visual processing. *The Journal of Philosophy*, 114(1), 5–32.