Cognitive Penetration and the Cognition-Perception Interface

Daniel C. Burnston
Philosophy Department, Tulane University

PENULTIMATE DRAFT: please cite final version, forthcoming in *Synthese*:
http://www.springer.com/-/1/AVSxX1QenYabzt6nW6b2

1. Introduction

There is considerable debate in both philosophical and empirical settings about whether cognition "penetrates" perception. Recently, philosophers have offered cases of supposed penetration ranging from color (Macpherson, 2012) and shape (Stokes, 2012) perception to attention guided eye-movements (Wu, 2013), and have proceeded to consider the epistemic upshot of penetration (Siegel, 2012). I argue that there is currently insufficient clarity in these debates about (i) what distinguishes perception and cognition in the first place, and (ii) what *kind* of relationship is taken to be evidenced by empirical studies cited as instances of penetration. There is potential variability on both sides, but how we construe the answer to (i) constrains what we can say about (ii). In regards to (ii), many characterizations of cognitive penetration suggest a relationship that I will argue is highly implausible, given a somewhat more plausible view about (i).

Consider the following quotes:

"Jill believes that Jack is angry at her, and this makes her experience his face as expressing anger." (Siegel, 2012 p. 202)

"Cognitive contents play a role in perception and partly determine the contents of perception." (Deroy, 2013, p. 94)

"There are two experiences with different phenomenal characters and the reason for this difference is best explained by a difference in the propositional attitudes of two subjects who each have one of the experiences" (Macpherson, 2012, pp. 23-24).

"Perception changes the way it operates in direct response to goings-on elsewhere in the mind." (Firestone & Scholl, Forthcoming, p. 8)

These quotes come from both proponents and deniers of cognitive penetration—proponents argue that they are true, deniers that they are false. It is not my purpose here to say whether these claims are right or wrong as such, but instead to argue that they are ambiguous, and that alleviating this ambiguity is important for understanding the relationship between perception and cognition. On a stronger reading, which is compatible with what much of what proponents of cognitive penetration hold, cognition implements *highly specific*, *directed*, and *isolated* computational effects on perceptual processes—cognitive states provide informational input to specific perceptual processes, which change their computations in specific ways based on this input. Since on this stronger reading, cognitive input modifies the computations performed by individual perceptual processes, I call it the "internal effect view (IEV). IEV captures the kind of idea presented by any metaphor according to which cognition "makes" perception come out a certain way, "determines" the content of a perceptual representation, "enforces its content" on perception, etc. I propose an alternative, weaker reading of the above claims, the "external effect view" (EEV), which contends that cognition can and does exert diverse causal influences on perception without affecting the computations performed by perceptual processes, but instead by biasing *which* perceptual processes will occur when a stimulus is present. This view, rather than positing a computational influence of cognition on perception, posits an associational one. Cognitive states are associated with a wide range of possible perceptual outcomes, but do not determine which of those will be perceived in a given case.

I claim that perceptual and cognitive representations should be differentiated by their internal structure and the type of relationships they have to their referents. I will argue that, on this kind of distinction, IEV is false, because it posits a computational role for cognitive states in perceptual processing that they lack the needed content to perform. Since IEV and EEV have not been clearly distinguished, I will not attempt to place each discussant of cognitive penetration as a supporter of one view or the other. Those with the stronger view in mind will find here an argument against their intended views. Those who favor a weaker interpretation will, I hope, find here a detailed way of thinking about their intended position. I also will not weigh in on whether the kind of relationship I describe genuinely counts as "penetration," since the meaning of that term is multiply interpretable depending on one's intuitions about the strong and weak views. I think EEV definitely holds; I think IEV definitely doesn't. This has the advantage of taking a unified approach to a range of possible interactions between perception and cognition, rather than needing to look at each of them and argue which one genuinely counts as "penetration." Whether the weaker relationship described by EEV is worth calling "penetration," I think, is of less importance than getting the nature of the relationship right.

Before beginning, let me offer a few clarifications. First, while I am interested primarily in cognitive architecture, I am not here proposing a specific mechanism by which cognition affects perception.[1] I am attempting to cash out the logical space of possible relationships and argue that one type is more plausible given the current range of evidence; the details of the mechanisms by which this kind of relationship is implemented are obviously matters for empirical investigation. Second, while I will adopt computational language in this paper, I am only loosely committed to "computationalism" about perception. There is ample evidence that perception performs "intelligent" abductive processes over particular inputs, which can abstract from and classify inputs (for an early formulation of this idea, see Rock, 1983). I will describe perceptual processes as individual functions that that perform these types of operations—this is as much as is needed to describe a process as "computational" on my view.

The paper proceeds as follows. In section 2, I introduce my construal of the distinction between perceptual and cognitive states. In section 3 I define IEV and show that it is strongly consistent with the conclusions that proponents of cognitive penetration have pursued. Section 4 presents

[1] Even though philosophers generally claim to be interested in the relationship between cognition and perceptual *experiences*, all arguments surrounding cognitive penetration involve characterizing and attempting to establish (or deny) particular causal relationships between the two. As such, a typology of possible causal relationships is highly pertinent.

the argument against IEV. Section 5 introduces EEV, and explains how it can account for the data often taken as evidence for IEV, and section 6 presents further evidence in EEV's favor. Section 7 concludes.

2. A representational distinction

I contend that stereotypical cognitive states such as beliefs, desires, and intentions should be distinguished from perceptual states based on the *types* of representation that they are: the way that the representations are structured, and how they relate to their referents. Cognitive representations, on my view, are *discrete*, meaning they have no *referentially relevant* internal structure. The primary exemplars of discrete representations are lexical and atomic concepts. Lexical concepts have structure (think of the phonemic structure of 'mongooses'), but this structure is arbitrary with respect to the referent of the concept. Atomic concepts, according to philosophical wisdom (Fodor, 1998), have no internal structure, meaning that they cannot be divided into parts at all, and a content that does not describe or specify any properties of their referents (Millikan, 2000). Perceptual representations, alternately, do have a representationally relevant internal structure, in that they have parts that carry distinct information about the distinct properties of their referents. I will call this the "form" distinction.

An old idea in philosophy is that visual states are *images*, which are distinct from *sentences* or propositional representations more generally. This is an easy first gloss of the form distinction, though few take the idea that there are literal pictures in the mind seriously. More sophisticated versions of the form distinction can be found in Dretske (1981), and subsequently Kulvicki (2005). For Dretske, the difference between perception and cognition is "fundamentally a coding difference" (1981, p. 143), that is, a difference in how information is carried by perceptual versus cognitive states. For Dretske, an *analog* representation carries information, say that s is F, only along with other information, perhaps a lot of it (p is H, q is I, etc.). For instance, an image may carry the information that s is an old lady, but does so only in conjunction with information about how tall she is, what she is wearing, etc. The purpose of cognition, on Dretske's view, is to extract particular information that is carried in analog form in perceptual states, and to represent it *digitally*—that is, without any accompanying information.[2]

Kulvicki uses the notion of isomorphism to cash out the form distinction. The idea is that perceptual representations have *parts*, and that the parts of the representation are vital to their carrying information in analog form. Kulvicki claims that the parts of the perceptual representation correspond one-to-one with the properties of their referents, about which they carry information. They can do so even if they don't *resemble* what they carry information about. Unsurprisingly, both Dretske and Kulvicki posit that extracted, cognitive states are language-like. Seeing a square involves having a representation with parts that correspond to the properties of the square (the sides, corners, angles, etc.), which carry information about, e.g., the length of the sides. The word 'square', however, carries only one piece of information—that s is a square, and no information at all about, e.g., the length of the sides. For Dretske, the

[2] This is a slight oversimplification, since Dretske argues that, even in a digital representation, information can be logically nested—for instance, the information that s is a square nests the information that it is a rectangle. I will ignore this complication here.

distinction also corresponds with propositional structure—cognitive states are propositional, perceptual ones aren't.

While it is difficult to nail down the details of the digital/analog distinction and isomorphism (for instance, on any view in which perception can implement abduction, the idea of one-to-one mappings is certainly false), the core idea behind the form distinction is easy enough to cash out: the internal structure of a perceptual representation implements a mapping from its parts to the structure of what it represents. If a perceptual state represents a square, it has a structure that maps to the metrical structure of the square. If a perceptual state represents a particular shade of red, it represents the particular structural properties that place that shade at a particular location in color space—that is, as taking up specific values along the dimensions that define the color space. If a perceptual state represents a banana, it represents both the metrical and color space properties of the banana, as well as its other features.

Perception, on the form distinction, involves a large number of ways of taking transduced perceptual signals as input and outputting structured representations whose parts map to the structure of some aspect (or aspects) of the perceptual scene. Throughout, when I refer to a "perceptual process," I mean a computational process that implements some such function. I mean this rather liberally. For instance, the form distinction itself is compatible with the idea that perceptual processes can being categorical, can relate information about multiple features of the perceived world, can embody assumptions about the nature of perceptual stimuli, and can be multi-modal.[3] For convenience, I will often speak both of processes and "outcomes," the latter being the outputs of the former. Intuitively, perceiving an item as a face involves a distinct computational process from seeing it as a dog, and perceiving an object as red involves a distinct process from seeing it as blue.

It is worth pausing on the notion of "categorical." Some authors presuppose that any perception corresponding to a category must itself be the result of a concept, and hence cognitively penetrated (e.g., Macpherson, 2012, Siegel, 2012). But this is a mistake. There are many processes that are categorical—resulting, for instance, in outcomes of perceived chasing and perception of biological motion, and perhaps certain forms of social perception (Toribia, 2015)—that, despite corresponding to a high-level category, bear all the hallmarks of perceptual processes, such as being fast, relatively mandatory, and closely tied to particular arrangements of perceptual features (see Burnston & Cohen, 2015, Burnston, in submission for full arguments to this effect). So, it is at least possible for perceptual processes to correspond to categories or kinds, without being penetrated. Later I will discuss categorical perception of this sort in more detail.

Two final clarifications regarding what the form distinction is *not* committed to. While it is often assumed that perceptions as a whole are more *specific* than cognitive states (e.g., that they are more fine-grained), and while this is often true, this point must be distinguished from the ideas that (i) every perceptual state is, *eo ipso*, more specific than every cognitive state, and (ii) that perceptual states are *maximally* specific. As Dretske (1981) notes, we could, at least theoretically, write an extremely long sentence that captured every piece of information in a

[3] See (Burnston & Cohen, 2013, 2015) for a view of perceptual architecture that explicitly builds in these features.

perceptual state in digital form. We can also bitmap images, which involves representing their analog information in digital form. More mundanely, we can of course cognitively represent complex scene structures via syntactic combination—for example, 'There are three green mongooses in the corner atop the bookshelf'. This does not undermine the distinction. A perceptual representation of these rather unlikely mongooses will have a different structure, which itself maps to their actual (e.g.) metric and color-space properties.

Similarly, the form distinction is not committed to the view that we perceive as much structural detail as there is in the scene. In Julesz's (1981) famous studies of pre-attentive texture perception, for instance, he suggested that texture is perceived, not in maximal detail, but in terms of four or five prototypical texture features or 'textons'—small shapes which, in different combinations, provide a good statistical summary of a wide range of textures. If true, the fact that we only really perceive four or five textons pre-attentively does not mean that we don't perceive the structure of textures *as defined* by the textons. Texture perception, on this view, consists of a range of processes that take luminance inputs, interpret their statistics in terms of textons, and output representations of texture structure as defined by the textons. To take another example: I am color blind. I therefore make considerably fewer color discriminations than people with normal color vision. But the perceptual representations I *do* form of color have a dimensional structure corresponding to the areas of color space I can represent. As mentioned, perceptual processes often abduct, abstract, etc., but so long as some mapping is implemented by the structure of the output these are compatible with the form distinction.

The form distinction is compatible with many ways of articulating the perception/cognition distinction in philosophy and in cognitive science more generally. Philosophers are in the habit of describing cognitive states as attitudes taken towards propositions, which are syntactically structured semantic units composed of something like lexical or atomic concepts. A variety of approaches to perceptual states in cognitive science count representations as perceptual if they are "imagistic," (Barsalou, 1999) or "depictive" (Kosslyn, 2005), where this is not understood to mean a literal picture, but instead the embodiment of relations that capture perceptual structure. Theorists within this perspective often distinguish perceptual from *lexical* representations, which have a different, syntactic structure (Barsalou, Santos, Simmons, & Wilson, 2008).[4] On *dimensional* views of perception within philosophy, perceptual qualities are defined structurally. Rosenthal (2005), for instance, argues that it is *definitive* of shades of perceived color that they exhibit the particular relational structure they do vis-à-vis the dimensions of color space. The form distinction as I've drawn it is highly compatible with this approach to perception. Finally, the form distinction has been important for those interested in debates about non-conceptual content. Heck (2007) has argued that non-conceptual representations are "iconic," meaning that they embody relations between their components, and even Fodor (2007), otherwise a staunch supporter of a purely lexical view of representation, has considered this possibility (although he has gone back on this in more recent work; Fodor & Pylyshyn, 2015).

In addition to capturing a lot of what people mean by the perception/cognition distinction, the form distinction also has pragmatic advantages. Given, for instance, that it is not obvious that perceptual processes can't correspond to categories, when analyzing a particular categorical

[4] I will often discuss lexical concepts hereafter, but it should be noted that I don't take these as predominant over or exclusive of atomic representations.

representation we need a way of saying whether it's cognitive or perceptual. The form perception is neutral as to, ultimately, what properties are represented by cognition and perception, and hence is helpful in this regard.[5] I fully admit, however, that the form distinction is only one way of distinguishing perception and cognition. Pylyshyn, for one, strongly rejects it, arguing that perceptual representations "are in every respect like conceptual representations except the categories are sub-personal and are not available outside some modular process" (2007, p. 72). We can see here the negative pragmatic consequences of drawing the distinction this way, since this definition both prescribes a set of represented properties inherent to perception, and interdefines perception with impenetrability, which is precisely what's at stake. Moreover, Pylyshyn (2007, ch. 4) seems at least partially motivated by the idea that "depictive" representations *really are* pictures after all, and hence must represent the whole of the perceptual scene to some arbitrary level of detail. But this is certainly not entailed by the version of the form distinction I sketched above. Rather than attempting to establish the form distinction over all possible comers, however, I am interested in exploring its ramifications. If the form distinction is right, I claim, then IEV is wrong. I turn to that argument now.

3. IEV

Some conditions on cognitive penetration are widely agreed upon. In order to penetrate a perceptual process, a cognitive state must interact with it causally (either directly or mediated by something else). Moreover, there must be a semantic "coherence" relationship between the cognitive state and the modification of the perception. If the penetrating belief is about apples, the modified output of the perceptual process has to be intelligibly related to categorical facts about apples. This second condition is needed to rule out causal influence on perception that is architecturally uninteresting. If I've been hypnotized into closing my eyes whenever I believe something in front of me is an apple, then I will go from seeing the apple on the table to not seeing it—a perceptual modification. Though the output of the process leading to apple-perception is modified (to a null output), the modification is not interestingly related to any beliefs I have about apples.

The combined causal and semantic coherence conditions, however, are not sufficient to capture the kind of relationship that some proponents of cognitive penetration seek to establish. Another condition, the *computation condition*, is often advanced, either explicitly or implicitly:

> *Computation condition:* The modification of a perceptual process P by a causal influence from a cognitive state S is due to a computation over S's content.

Different theorists phrase the computation condition in different ways. Wu claims that cognitive states penetrate perception in virtue of providing an "informational resource" for perceptual computations (2013, p. 649). Pylyshyn (1999) suggests that there needs to be a logical as well as semantic relationship between the content of S and the change in P, and explicitly cashes this out in computational terms: "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, that is, it can be altered in a way that bears some logical relation to what the person knows" (1999, p. 343). That

[5] For a full discussion of these topics; see Burnston (in submission).

is, there is not only an intelligible relationship between the cognitive state and the perceptual output, but the change is the logical change to make, given S—P gets modified *according to* the semantics of S. While Pylyshyn does not clearly distinguish the semantic coherence and computation conditions, he is thus committed to both. Macpherson (2012) cites Pylyshyn's condition as constitutive of cognitive penetration.[6]

The computation condition informs several aspects of the cognitive penetration debate, for instance the nearly universal identification of cognitive impenetrability and modularity. According to Fodor (2001), a process X is modular if it has a "proprietary database" of inputs, and only those inputs count when it performs its function. Penetration of X by Y means that Y's content counts as an input to X (i.e., that the output of X can be different depending on what Y says), and therefore that X is not encapsulated, and not modular, with respect to Y.

The computation condition is partially constitutive of IEV, which can be defined as follows:

> IEV: A perceptual process P is penetrated if, over a specific input, it would perform a certain computation C leading to content R1 in the absence of a cognitive state, S, but performs a different computation C2, yielding content R2, when S is present, where the causal, semantic coherence, and computation conditions are met.

IEV posits an "'internal" modification because the function that P computes is modified: P *would* take the input and map to R1, but given that it also takes the cognitive input under consideration, it maps to R2. Cognition in this sense "determines" the function computed.

A vital thing to note about IEV is that it is proposed as an *explanatory* principle. We notice a perceptual effect that seems to be related to the presence of a cognitive state. Whether the effect is a case of cognitive penetration then depends on whether it is best explained by an IEV-type relationship. The view places a large computational burden on the cognitive state—it must, in virtue of its content, determine the specifics of the computation performed by a perceptual process. In the next section, I argue that, on the form distinction, IEV is false.

4. Against IEV

The standard arguments for cognitive penetration rely on empirical cases in which the perceived category membership of a stimulus modifies perception of its features. That is, when an object is perceived to belong to a category, its features are perceived as having values closer to "stereotypical" features for the category—e.g., a redder shade for an object shaped like a heart. Aside from arguing that the effects are not perceptual, because they occur either before or after perception,two other strategies have emerged (or re-emerged) to combat these arguments. The first is to argue that the effects do not in fact occur, by citing methodological problems in the studies under discussion (Firestone & Scholl, Forthcoming). The second is to argue that even if

[6] Stokes' (2012) definition of cognitive penetration does not explicitly include either the semantic or computation conditions, but he has a weaker view than many in this respect. Wu (2013), in some places the most explicit proponent of the computation condition, also occasionally describes cognitive influence on perception in terms of "structuring causes," which could potentially be interpretable as a weaker relationship. I offer such a weaker characterization in the next section.

the effects occur, they are not best explained by cognitive penetration, but instead can occur due to associative processes *within* perception. To these arguments I will add a novel one, an *architectural* argument, which focuses solely against IEV as a way of characterizing these effects. Much care is required here. Obviously, if the effects do not occur, then they do not need to be explained by any view. Despite these concerns, it is still important to pursue the architectural argument, for two reasons. First, my point is a conceptual one—if one buys the form distinction, then IEV cannot explain these types of effects, to whatever extent they occur, which at least must still be considered an open question. Second, there are effects, discussed below, which are not clearly targeted by the methodological critiques, which seem to involve an interaction between perception and cognition, and which cannot be explained purely intraperceptually. But the architectural argument against IEV applies in these cases as well. Finally, there are a range of *other* relevant effects that have not been discussed in the cognitive penetration literature (section 6)—IEV's failures point the way to a view that can also explain these effects.

Given all this complexity, let's start with a non-controversial case of perceptual modification in non-categorical contexts. Consider the following different presentations of the Cornsweet illusion. In the basic stimulus on the left, two equiluminant patches are presented with an edge between them. On the left side of the edge is a thin light-to-dark gradient. On the right is a dark-to-light gradient. Viewers generally perceive the patch on the left side as brighter, and the patch on the right side as darker, despite the fact that they have identical luminance values (aside from the gradient). The standard explanation for the effect is that the combined gradients provide a depth cue (i.e., the edge is closer to the viewer than the background), such that the patch perceived as darker is likely facing away from the light source, and therefore likely to be shaded, while the patch perceived as lighter is more likely to be facing the light source. Interestingly, the effect is modifiable depending on the amount of evidence for the depth interpretation. If the lines distinguishing the stimulus from the background are removed, the effect disappears entirely (center panel). If the depth cues are accentuated by bowing the two panels, the effect is heightened (right panel), and the effect can be made extremely pronounced by further manipulation (Purves, Shimpi, & Lotto, 1999). There is also a color version of the effect.

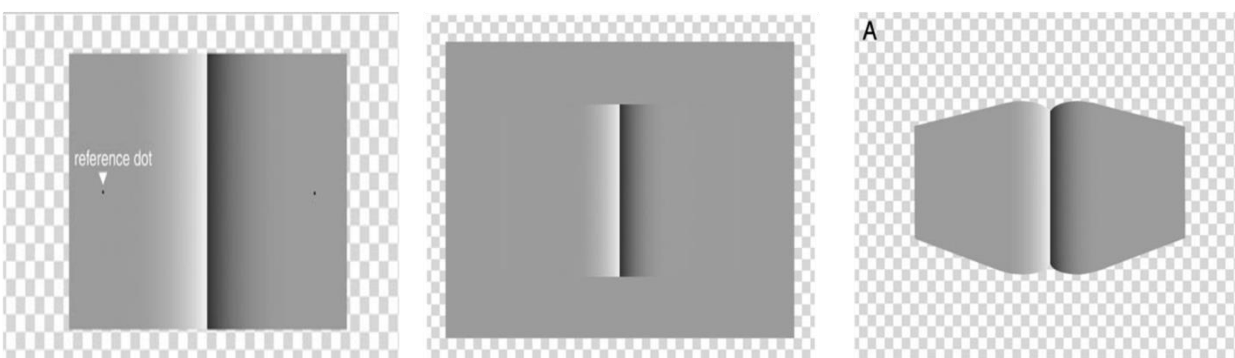

Figure 1. Permutations of the Cornsweet illusion. From Purves et al. (1999).

The important features to note are that the Cornsweet effect (i) depends on multiple cues, and (ii) is *graded*. What determines the effect on luminance is the overall evidence for a depth interpretation, which is in turn defined by a range of figure/ground, luminance, and discontinuity (edge) cues. Moreover, the effect is graded depending on the extent to which these features provide evidence for the depth interpretation. This is a clear case of perception abducting—using multiple cues to come to an interpretation of the stimulus that goes beyond the strictly available stimulus values, and map them to the structured representation that determines the percept. There is thus strong evidence that perceptual modification occurs in non-categorical settings. The question is then whether these kinds of effects occur in categorical settings, and if so, how to explain them.

Consider one of the examples Macpherson (2012) takes as evidence for cognitive penetration. In Delk and Fillenbaum's (1965) study, subjects were shown objects presented in an orange-red color, and asked to match the background color to that of the presented object. Subjects tended to make the background a more saturated red for stereotypically red objects (e.g., hearts, apples) than for stereotypically non-red objects. Macpherson concludes that the belief that the object is a heart modifies the color perception process, producing the distinct outcome. Thus, it seems Macpherson takes the case to exhibit IEV.

Assume for the sake of argument that these studies are methodologically impeccable, and the veracity of the results universally accepted. Even so, if the form distinction is granted then IEV cannot explain the effect. This is due to the fact that IEV is intended to explain specific perceptual modifications. Normally, P (color perception) would take the current input and output R1, but given the content of S (heart belief) as an additional input, it takes the same perceptual input and outputs R2 (more saturated red). Recall that, per the form distinction, a cognitive state doesn't contain any content that maps to the structure of perceptual stimuli. However, per the same distinction, perceptual representations *must* contain such structure, and the function that perceptual processes compute involves mapping inputs to structured outputs. Since the change from R1 to R2 is precisely a change in what perceptual structure (e.g., metric properties or location in color space) is represented, the content of cognitive representations cannot determine the change. That is, the cognitive state doesn't have the right kind of content to provide the "informational resource" that tells perception to modify its function in a particular way—viz., to end up at a particular R2 instead of R1. For a Dretskean, the argument can be put in terms of the notion of extractability.[7] When cognitive states extract specific properties out of a structured perceptual representation to represent in a distinct format, the other information embodied in the structured perceptual representation is *lost*. But since perceptual representations with particular structure are the outcome of a perceptual modification, and since cognitive states contain no such content, they fail to explain perceptual modification.

In the Delk and Fillenbaum case,what is needed to explain the effect is to show what causes the perceptual output to move from one specific spot in color space to another. Why should the color-representing process move from representing R1 (a less saturated shade of red) to this specific R2 (a more saturated shade)? Per the form distinction, there is nothing in the content of the concept 'heart', or the concomitant belief, 'That is a heart', that represents *anything* about the specific perceptual structures inherent in perceiving hearts. So there are no content resources in the cognitive representation to guide the perceptual one to a specifically structured endpoint. But

[7] Thanks to Matt Fulkerson for encouraging me to stress extractability here.

since, on the computation condition, the content of the concept/belief is what is supposed to explain the shift in the computation, IEV fails to capture what is going on in the case.

It will be helpful to flesh out the argument further by considering some likely initial responses.

4.1. The Blanket Command Objection

The blanket command objection contends that I've over exaggerated the claim of IEV, by positing a more sophisticated computational influence than intended. Perhaps the influence of the cognitive state needn't represent in detail the structure of the stimulus, but merely convey a kind of command to perception—e.g., 'If x is a heart, modify saturation by 5% to produce a redder shade'. Perhaps such a command could brute-force cognitive influence on perception in a way that could determine perceptual content, without falling prey to the argument against IEV.

Two aspects of perceptual modifications are incompatible with the blanket command objection. The first is the possibility that the result of feature interactions can be graded depending on the *amount* of evidence for the perceptual modification. We saw in the case of the Cornsweet illusion that such cases occur in non-categorical cases. Importantly, Deroy (2013) cites similar graded effects for categorical perception—categorical effects on color perception are more pronounced when there are corroborating depth and texture cues for the category (Olkkonen, Hansen, & Gegenfurtner, 2008). There are now two points to be made. First, given that perception can be independently shown to implement graded effects in non-categorical situations (as in the Cornsweet case above), and that the effects in categorical cases are also graded, what justifies positing a difference between the two cases? It seems that the only reason would be the flat assumption that all categorical effects must be cognitive, which begs the question against intra-perceptual readings. Second, and more importantly for current purposes, the command objection posits a *specific* command—5% adjustment, or whatever—as the way in which cognition modifies the perceptual function. But the effects are *not* specific in this way; instead they are graded depending on the evidence.

The second aspect incompatible with the command view is the *diversity* of potential categorical effects. The fact that a perceived object belongs to a category possibly implies a large number of perceptual consequences, both within and across modalities. The fact that something is dog-shaped increases the likelihood that it will bark, and vice versa. It also increases the likelihood that it will be furry, move in a tail-wagging manner, etc. While the possibilities are fewer with hearts, something's being heart shaped not only increases the likelihood that it is red, but also that it will make lub-dub noises. Suppose the belief 'That is a heart' is tokened. Which perceptual process should be modified, in which way? Given that the cognitive representation, ex hypothesi, lacks perceptually structured content, there is no way for the cognitive state to determine which out of the possible potential perceptual effects should be implemented.[8]

[8] Why not suggest that a single cognitive concept can convey *multiple* distinct commands to distinct perceptual processes? All that this would require would be that the cognitive content was *interpreted* differently by distinct perceptual processes to which it was connected (e.g., perceptual processes in distinct modalities would interpret the same signal as a command to modify their mappings of arguments in their distinct domains), and this would support distinct internal effects coming from the same process. While this argument, appropriately supplemented, might show how distinct processes could be modified, it doesn't account for evidence-sensitivity or graded effects. So

I will discuss diversity further in section 6. For now, the foregoing can be summarized by pointing out that IEV cannot hope to account for both the specificity and diversity of categorical effects on perception. The argument against IEV claimed that cognitive representations could not account for specific changes in perceptual outputs. If one tries to modify IEV to account for these specifics—e.g., via the command view—one loses the ability to account for diverse, gradedeffects stemming from the same category. On the other hand, since perception is able to make feature associations in an evidence-dependent way, the command claim seems superfluous.

4.2. The 'Expand the Cognitive Representation' Objection

Another objection contends that I have oversimplified the cognitive representations involved in modifying a perceptual state. Beliefs generally do not come singly. Perhaps it is not one propositional attitude ('That is a heart') but instead multiple beliefs ('That is a heart' and 'Hearts are red') that in conjunction modify the perceptual process in the requisite way. Initially, this strategy has relatively little plausibility. 'Red' is a determinable, and the concern over IEV is based on the inability of the content of cognitive representations to account for changes between representations of determinates—e.g., from a specific shade to an output with a more saturated red. So adding the cognitive representation 'red' to the cognitive representation 'heart' does not seem to add enough content to explain what needs to be explained.

One might suggest that further discrete representations could be added to specify the needed content. One might, for instance, add 'That is a heart colored Oxblood red' to the store of occurrent beliefs, where 'Oxblood' picks out a specific shade of red. After all, as per section 2, the form distinction itself makes no claims as such on how many cognitive representations there are, and how specific they might be. But alas, this strategy won't work either. Why is the belief that the heart is colored Oxblood red tokened? If only one particular shade belief can co-occur with the heart belief (e.g., 'All hearts are Oxblood red'), then this objection collapses into the command objection, which has already been dismissed. On the other hand, if the belief that the heart is a particular shade needs to rely on some evidence, then it needs to rely on *perceptual evidence*. But if perception already has some way of determining the shade to which the output should be adjusted, such that it could serve as evidence for the belief that the heart is colored Oxblood red (as opposed to some other shade), then there is no reason to posit a computational influence of cognition on the perceptual process. IEV, in this case, would get the causal story the wrong way around.

4.3. The Translation Objection

Some proponents of cognitive penetration suggest that the cognitive influence on perceptual processing is indirect (Macpherson, 2012; Wu, 2013). This implies that some mediating process conveys the content of the cognitive state to the perceptual system for use in modifying its function, and it could be contended that at some point during this process the content of the cognitive state is translated into the kind of content that perception can use to modify its functioning. Wu (2013) suggests a scheme for translation, on which it involves a process that, at

---

long as the appropriate amount of modification depends on the evidence for the category, simply conveying a different command to each perceptual process won't solve the problem. Thanks to an anonymous reviewer for raising this objection.

each stage, includes at least some of the semantic content from the last stage. If, at the end of the process, the content (or some of it) of the original message is included in the final stage, the translation has been (perhaps partially) successful.

Positing translation, however, will not solve the problems for IEV. Recall that on IEV the effect on perception has to not just reflect the content of the cognitive state, but the content has to cause perception to modify its function in a particular way. To avoid the argument against IEV, the translation process would need to not just convey the content, but in fact change it into the type of content that could cause a specific perceptual effect, viz., perceptually structured content. While it could be proposed that such a mechanism exists, the argument against IEV would then arise *within* the translation mechanism. Suppose that some stage in the translation mechanism is proposed to implement the change in representational type. One way for it to do so would be to represent both kinds of content, such that it can map the discrete to the perceptual. To which perceptually structured content should it map the cognitive representation? Without some theory of why the mechanism should "choose" one specific perceptual content as opposed to another, no explanatory gain has been made by positing translation. If one posits that the translator just always maps a particular cognitive content to one particular perceptual one, then the translation objection collapses into the command objection, with its attendant failings.[9]

In the next section, I present a positive view of the cognition/perception relationship that starts from this state of affairs. The "external effect view" (EEV) posits that the semantic coherence condition is fulfilled in many diverse interactions between perception and cognition, but eschews the computation condition entirely.

5. EEV

I start from an intuitive distinction between affecting a process internally and affecting it externally. Suppose a company CEO wants a vote amongst the board members to go a particular way. She is worried, however, about the outcome. One way she could affect the outcome is by talking to the individual board members, trying to change their minds. Her effect on the board's deliberations is thus *internal*, because she is attempting to modify the deliberation processes of the individual components. Now suppose, alternately, that some of the board members are similarly minded to the CEO, and vote with her consistently. The CEO, rather than talking to anyone individually, surreptitiously rewrites the by-laws of the company to give greater weight to the members who tend to vote with her. Importantly, none of the individual board members do anything different than they would otherwise do. The effect on the outcome is thus *external* to their deliberations.

In both cases, the CEO gets her way. This corresponds to the semantic coherence condition being fulfilled, since the eventual outcome of the vote in either case matches the CEO's intention. The important question is how to describe the causal influence: IEV approximates the internal influence, EEV the external.

---

[9] Similarly, it won't work to propose that the translation mechanism doesn't represent anything at all, but is a pure format converter, similar to a digital-to-analog converter. Without going too far into it, digitial-to-analog converters work by mapping each digital signal, via a filtering or interpolation function, to *one* specific analog waveform or value. So, positing a format converter is simply a sophisticated version of the command objection.

> EEV: Tokening of a lexical/atomic concept as part of a cognitive state provides a bias towards any perceptual processes associated with the concept, raising the probability that those processes will be applied to a perceptual stimulus.

Suppose we have a probability distribution over possible perceptual outcomes. IEV posits that, at least in certain perceptual situations, the presence of a cognitive state collapses this distribution to a unity. On EEV, however, the tokening of a concept simply shifts the distribution to a new shape, one that does not determine any particular perceptual outcome, but instead re-sets the initial probabilities, now skewed in favor of a certain *set* of possible outcomes. Read this way, IEV would entail EEV, since both shift the distribution. However, one should not be tempted to conclude based on this idea that IEV and EEV are on a continuum. EEV is compatible, where IEV is not, with the possibility of several (possibly many) *distinct* outcomes each being potentiated by tokening a single concept. Recall from the last section that accounting for diversity was one of the major problems for IEV. Given that IEV posits a change to the function itself, namely how a given function maps given inputs to given outputs, it struggles to account for diversity in effects. EEV, however, in shifting the distribution without determining the specifics, *potentiates* a diverse set of *already existent* functional mappings. This is why I call it "external"—where IEV posits a change to the *identity* of a function, in virtue of changing how it maps arguments to solutions, EEV leaves perceptual functions intact. Another way to put the difference: according to EEV, the relationship is associational, not computational. EEV suggests that cognition can affect what outcomes perception arrives at (thus fulfilling the semantic coherence condition) without perception computing over cognitive contents. As I will discuss below, there is ample evidence that causal relationships between perception and cognition in fact work something like this.[10]

Biasing has been discussed before in the literature, mostly surrounding the notion of "biased-competition" models of attention (e.g., Mole, 2015; Raftopoulos, 2009). While I will discuss attention below, I think the notion of biasing is much more fundamental, and applies equally well to cases that are not obviously instances of attention. Moreover, the import of biasing, I contend, is only best appreciated within an EEV framework. I will discuss a range of cases below to illustrate these points. To flesh out EEV in more detail, however, and to inspire discussion of those cases, it is necessary to take a short detour into the nature of categorical perception. In section 2, I stressed the possibility that categorical perception can be genuinely *perceptual*. Perceptual categories, according to one well-established set of views (see, e.g., Palmeri & Gauthier, 2004), are representations of a category's location in in a "space" defined by perceptual dimensions. Take the example of facial categories. Perceivers have the ability to distinguish both individual faces and types of faces. These accounts claim that each face category has place in a "face space," defined along the particular dimensions of faces (e.g., the

---

[10] To acknowledge some debts: EEV is inspired partially by Barsalou et al.'s (2008) theory of language-perception interaction, which treats linguistic representations as a kind of indexing system for perceptual representations, and by Millikan's (2000) underappreciated theory of atomic "substance" concepts. However, EEV neither shares all of the assumptions, nor recapitulates all of the claims of these views, and supplies a more detailed view of the interface than either. Like these other views, EEV presupposes some learning theory that explains how particular lexical or atomic concepts get hooked up to particular processes. I won't speculate as to that here; see Millikan (2000) for one proposal.

shapes of the features, their configurations, the distance between them etc.), and recognizing a face consists in perceiving it as occupying a certain place in the space. One can think of a perceptual category as an exemplar representation in this space, and occurrent object recognition as perceiving an object as "close" to the exemplar for the category.

Importantly, this view also shows, at least conceptually, how perception of individual objects can be *modified* in the course of categorical perception. Several studies have shown that forming a perceptual category consists in developing "morphspaces"—objects grouped into the same perceptual category are perceived as *more alike* along the relevant dimensions than those that are not. As Goldstone and Hendrickson put it, in categorical perception "our perceptions are warped such that differences between objects that belong in different categories are accentuated, and differences between objects that fall into the same category are deemphasized" (2010, p. 69). These modifications have the potential to explain how intra-perceptual feature modification could occur.[11] It is also clearly an *abductive* process: it consists in perception representing an occurrent stimulus as perceptually closer to a categorical exemplar than is strictly entailed by the stimulus values.

Perceptual categories, then, are supraordinate abductive processes that hijack lower-order processes, including ones for the particular features that are stereotypical of the category, and produce representations representing the configurations and relations between features as occupying a particular place in perceptual space. It is theoretically possible, and in fact very likely, that these categories could extend to multimodal features. Goldstone et al. (2013), for instance, showed in a priming study that the presence of a stereotypical category feature in one modality primed recognition of stereotypical features in other modalities. If this is the case, then we should expect categorical effects on perception to range across modalities. This kind of architecture has been established both for faces and for non-face objects such as cars and animals, and even for arbitrary objects (Palmeri & Gauthier, 2004).

The proposal of EEV is that tokening of a concept potentiates any of the perceptual processes associated with the category, including those that represent particular object features and the processes that integrate them, but does not determine any of their specific outcomes. Put simply, there are a range of ways of perceiving dogs, corresponding to the particular dog categories one recognizes (as well as to the general perceptual properties of dogs), their various shapes, barks, etc., and the particular postures in which and views from which one can recognize them. Tokening 'dog' potentiates the functions resulting in this particular range of perceptual outcomes, construed as representations in "dog-space."

---

[11] There are a number of remaining questions about this kind of view, for instance how viewpoint-dependent these representations are, whether categories consist in abstract prototypes or a range of exemplars, etc. I won't go into these here (but see the end of section 6). Also, I should note some different potential readings of these claims—Goldstone and Hellickson (2010) suggest that these categories reflect "permeability" between cognition and perception, since they take the reading that categorical effects must be cognitive. I think this is a mistake, for the reasons given in section 2 (cf. Burnston, in submission]). Raftopolous (2009, p. 70) takes the fact that categorical representations like these are representations in *memory*, and must be matched to incoming stimuli, to show that they are cognitive. Others have focused on the fact that they are *learned* to make a similar point (Stokes, 2014). But given that it is possible to store representations with perceptual form in long term memory (Barsalou, 1999), these inferences don't follow. I can't argue for this convincingly here; I say much more on these topics in Burnston (in submission).

This has a variety of important consequences for thinking about the cognition/perception relationship. First, it shifts the computational burden in determining perceptual detail. IEV, given the difference in format between cognitive and perceptual representations, can't determine perceptual structure. But, on the perspectives I've been outlining, this is the kind of thing that perception *already* does—we have seen this in non-categorical contexts such as the Cornsweet illusion, and I have given the conceptual background for it in categorical contexts here.

Second, EEV has the potential to account for the diversity of perceptual modulations discussed in previous sections. Not only are there a range of features associated with each perceptual category, but each of these features can take a range of values depending on the viewing conditions. EEV, in simply positing a bias towards or potentiation of the range of perceptual processes associated with the category, doesn't say which of any of these outcomes should *actually* come about. It just makes it more likely that *one of them* will. So, the eventual percept should be an interaction effect between whatever processes are potentiated and whatever stimulus actually occurs.

Finally, this view opens up a number of interesting properties that, as shown below, are important for characterizing the perception-cognition interface. First, potentiation can occur through tokening a concept corresponding to the overall category (e.g., 'dog') or through tokening one corresponding to one of its subordinate features (e.g., 'furry'). Second, to the extent that two objects share similar perceptual outcomes—e.g., 'dog' and 'cat' both involving 'furry' perceptions as a subordinate process—we can expect that tokening a concept of one category might involve effects on *other* perceptual categories, to the extent that those categories share subordinate features.

In what follows, I will show how this perspective accounts for a range of data about the perception-cognition interface that is hard to square with IEV. Here is what is important about the range of cases discussed below. First, the range of cases comprises interactions that are taken by almost everyone to be cognitive penetration (lexical influence on perceived categories), *and* cases historically seen as more marginal instances of cognitive penetration. These include task and lexical effects on eye movements and priming effects, which only some proponents of cognitive penetration think should count. But since the *same* type of interaction occurs in each case, there is no deep question here about which type of case *genuinely* counts as penetration. There is simply one type of relationship (an EEV type) being enacted in distinct circumstances. First, in this section, I will show how EEV accounts for straightforward interactions between lexical and perceptual representations, then move on in section 6 to discuss the broader range of attentional and priming results.

Consider another case offered by Macpherson. In Levin and Banaji's (2006) study, the experimenters constructed faces that were racially ambiguous—that is, their features did not suggest that they belonged either to a black or a white person, but instead were somewhere in between. Aside from external influence, perceivers see these faces *as* ambiguous, as not being either black or white. However, if the faces are presented along with a visually presented word—i.e., 'black' or 'white'—subjects see the ambiguous face as either lighter or darker, depending on which word is presented. Macpherson interprets the case as evidence for cognitive penetration, and indeed it is hard to interpret it along *purely* intra-perceptual lines, since it is due to the tokening of a lexical concept and not to varying perceptual evidence. Moreover, this particular study is not directly criticized by the major methodological opponents of cognitive penetration, and is thus proposed by some authors as avoiding these criticisms (Nanay, 2014).

A potential IEV explanation faces the argument outlined in section 4. EEV, however, has a better set of resources for making sense of the effect. On the EEV view, "ambiguous" stimuli can be interpreted as cases for which the inputs provide insufficient evidence to determine which of two distinct processes—abducting from the input to a representation closer to a stereotypically black face, or towards a stereotypically white face—should be implemented, leading perception to suspend judgment and represent an ambiguous face (Burnston & Cohen, 2013). However, when the lexical label 'black' is presented (and thus the corresponding concept tokened), this shifts the baseline probability that processes corresponding to 'black' outcomes will be perceived, amongst which is the process that (partially) eventuates in perceiving black faces, thus disambiguating which process is to be employed, and resulting in the darker perception.

EEV does not have the same problems as IEV, because it shifts the computational burden of the explanation. Perception, on EEV, is already in the business of modifying perceptual outcomes abductively based on the available evidence to arrive at an interpretation of the stimulus. Unlike on IEV, the influence of the cognitive state does not determine the perceptual outcome, or in fact modify the process at all. The actual computing is still left entirely to perception; it is only the baseline probability that a particular function will be run that is changed. Another thing to note is that the effect can occur even if it is the entire set of perceptual processes involved in perceiving stereotypically black (or white) things that get a biasing influence.[12] The specific effect is due to one particular one of these processes (the one for black or white *faces*) being confluent with the rest of the perceptual input. As it turns out, this perspective can be extended to wide range of other interactions.

6. Further Evidence for EEV

x

If EEV is true, as I outlined it above, there are a variety of properties we should expect of perception-cognition interactions that are generally out of keeping with IEV. Here is a (probably not complete) list:

> *Diversity*: Since multiple processes are biased in virtue of a single cognitive representation, we should expect facilitative effects in diverse perceptual circumstances, at distinct "levels" of perception, and multi-modally (including motor effects), from tokening a single concept.

[12] For simplicity's I will talk about a uniform biasing of any process associated with the concept, but EEV is not in principle committed to this. Some processes, in virtue of being more typical of the category, could be more heavily biased than others.

Inducibility: Since the biasing effect is inherently *probabilistic*, whether an effect occurs should not be mandatory, but instead sensitive to whether corresponding perceptual evidence is present.
Generality: Since biasing does not determine specific outcomes, specific values for perceptual outputs should not be biased.
Categorical overflow: If perceptual processes can be shared between categories (as discussed in the previous section), then we should expect effects that cross-cut categorical boundaries.

In what follows, I suggest that evidence exists for all of these properties.

First, take diversity. It would be hard to list all of the effects that have been found between linguistic, perceptual, and motor representations (see Mishra & Marmolejo-Ramos, 2010; Zwaan & Madden, 2005 for reviews). I'll settle for a few examples. Reading language about actions and objects causes activation of motor plans (Glenberg & Kaschak, 2002), and number of imaging studies have shown overlap of activation in motor areas when processing action-words and actually performing those actions (see, e.g., Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005). Within perception, multiple different modalities are affected by lexical processing, as well as, within vision, multiple different feature and object-levels. Stanfield and Zwaan (2001) give evidence for the priming of orientation perception by linguistic stimuli, and a variety of priming effects are caused in color and object perception by color words (Heurley et al., 2012). Zwaan and colleagues (Zwaan, Madden, Yaxley, & Aveyard, 2004) show facilitation of direction-of-motion perception due to linguistic stimuli. Moreover, there is at least some evidence that olfactory experiences are affected by lexical labels (Herz & von Clef, 2001).

Discussion of two specific methodologies will highlight the other properties, as well as the fact that effects occur at different levels of perception (further supporting diversity). The first is the "visual world" methodology. In the visual world setup, subjects are shown a visual stimulus as they process an auditory lexical one; their eye movements are then tracked to see how the lexical stimulus has affected the perceptual salience of items in the visual display. If eye-movements are directed to particular objects more due to semantically related lexical stimuli than to controls, the lexical stimulus is taken to have affected perception in a semantically relevant way.

Now, this is an attentional effect, and attentional effects are not traditionally considered good cases for cognitive penetration. One reason for this is that the eventual outcome of this process is a change the input to perception (i.e., the change in what is foveated post-saccade), and most traditional definitions of penetration rule out changes-of-input as cases of penetration. However, a number of theorists have recently questioned this assumption, by questioning the notion that attention only involves effects that occur *before* perception. On biased competition models of attention, for instance, attention helps shape what perception will compute by biasing a competition for perceptual outcomes. On this view, "the processes responsible for the allocation of attention [are] inextricable from the processes that are responsible for the perception of the things to which we attend" (Mole, 2015, p. 225). That is, attention helps determine which processes are the victors in a perceptual competition. There is some support for this perspective in the visual world paradigm, since the influence of cognition occurs *in the course* of perceptual processing, rather than before it—the influence is what *causes* eye movements to be directed in a certain way to already-presented stimuli. In support of this interpretation, Huettig and Altmann take these effects to show "the time-course of the access of various types of lexical knowledge from spoken words and its integration with concurrent visual input" (2011, p. 139).

Ultimately, as I've stated, I am not concerned with what counts as "penetration," but instead with the properties of the cognition-perception interface in general. At the very least, these processes entail that cognition is interacting, in a semantic way, with perceptual processes—otherwise the effects would not privilege attentional shifts towards one object rather than another. So, it is undoubtable that these effects involve an interaction between perception and cognition; the important question is whether the interaction exhibits the properties suggested by EEV.[13]

In these studies, categorical overflow is shown by the prevalence of *competitor effects*. A range of results follow the general pattern that if a category is presented lexically, then both visual objects within the category and other objects that have similar perceptual properties will attract eye movements, even if they are otherwise semantically unrelated. So, if subjects are presented with a sentence containing the word 'snake', they will move their eyes towards a picture of a cable, since shape properties are shared between those items (Huettig & Altmann, 2007). Some categorical effects also exhibit generality. If, during a 'snake' sentence, subjects are shown *both* an uncoiled snake and a coiled rope, they will alternatively fixate on both. So, *specific* snake shapes are not biased—instead, percepts eventuating in *any* stereotypical snake shape will be facilitated (no matter what the objects actually are).

Categorical overflow is deeply out of keeping with the computation condition. Since, on IEV it is supposed to be the *content* of the belief that explains the effect, we should expect the belief to *rule in* snakes and *rule out* cables. So, the effect on vision seems to operate in a more brute, less sophisticated fashion than IEV posits. Generality is also incompatible with IEV, since it suggests that, cognitive representations do not modify perceptual processes in order to produce specific outputs, but instead potentiate a range of possible outcomes.

Huettig and Altmann's (2011) study in the visual world paradigm illustrates inducibility by comparing biasing towards occurrent and non-occurrent perceptual properties. In one example, subjects process a sentence involving the word 'spinach', while a variety of black and white drawings are shown to them. In one of the photographs, a semantically unrelated object, a frog, is prototypically the same color as spinach, but in this case no preferential fixation occurs. However, if a different semantically unrelated object *without* that paradigm color, for instance a shirt, *is* shown in green, preferential looking occurs. Why should this be? The results suggest that cognitive representations, individually and on their own, are not sufficient to produce specific perceptual effects, but instead require the presence of corroborative perceptual information. Inducibility, like the other properties, is incompatible with the computation-determining role for cognitive representations posited by IEV.

Another important methodology involves recognition studies, in which the subject is asked whether a presented perceptual stimulus matches or doesn't match the categories involved in a previously presented lexical stimulus. The reasoning here is that reaction times should be faster

[13] Thanks to an anonymous reviewer for encouraging me to get clearer about the role that the visual-world studies are playing in the argument here.

when the stimulus matches the lexical stimulus semantically, and EEV accounts for this kind of result in a similar way as for the visual world paradigm, via the biasing of processes that are then quickly applied to the input. While there are a variety of effects falling under this rubric, consider just one case. Zwaan and colleagues (Zwaan, Stanfield, & Yaxley, 2002) presented subjects with one of two sentences—'The ranger saw the eagle in the sky', or 'The ranger saw the eagle in its nest'. They then presented drawings of eagles with their wings either folded or outstretched. In match conditions, where the folded wings were paired with the 'nest' sentence, and the stretched wings with the 'sky' sentence, subjects were significantly faster to respond correctly than in mismatch conditions, where the opposite was the case.

The results highlight generality and inducibility. Clearly, "eagle" doesn't potentiate particular effects, since either match condition is successful at facilitating responses. An IEV reading of the sentence-effect is ruled out by the response to the "add more cognitive representations" objection given in section 4.2. Just adding the discrete 'nest' doesn't say anything about the specific shape that should be computed. Again, EEV offers better resources: while 'eagle' doesn't facilitate any specific eagle shape, and 'nest' doesn't facilitate any specific organism (e.g., eagle's nest versus hornet's nest), there is some overlap between the perceptual processes that are biased, since some eagle perceptions represent eagles in their nests, and some nest perceptions involve birds. Since some processes for perceiving eagle-information (i.e., the ones with wings folded) will get double bias from being *both* 'eagle'- and 'nest'-related processes, they will receive a greater bias and thus be the ones facilitated—so long as the perceptual input is consistent, as inducibility suggests. This explanation is admittedly vague, but it is consistent with the views of the scientists working on these topics, (see Altmann & Mirković, 2009; Zwaan & Madden, 2005 for roughly similar interpretations of a variety of cases), and offers a general approach for thinking about these effects, which I've argued are conceptually incompatible with IEV.

A final type of cognitive effect on perception is through perceptual imagery. MacPherson (2012), for one, takes cognitive penetration to work through the medium of perceptual imagery, and it might be thought that EEV cannot explain imagery, particularly due to inducibility. Inducibility requires the interaction of a potentiation or bias from cognition with actual perceptual input, and in many cases of imagery there is none. However, imagery does seem to be at least *highly* inducible. Consider Macpherson's (2012) description of the Perky effect, where subjects are asked to imagine a particular item, and then an image of that item is shown on a screen at sub-visual threshold. When a banana was shown in a vertical position at sub-threshold levels, "one subject reported [that] they tried to imagine a banana in a horizontal position and … that they had ended up imagining it in a vertical position" (2012, p. 29). While Macpherson takes the case to be one of cognitive penetration of imagery, this kind of effect seems clearly to evidence inducibility, which I argued above was what we should not expect to occur if the cognitive state in fact determined perceptual structure.

Now, neither IEV nor EEV, as I've currently phrased it, seems to fully explain imagery effects when there is no perceptual input whatsoever—IEV because of the problems raised in section 4, EEV because it relies on inducibility. Clearly, more will need to be said here. Perhaps there is a particular exemplar of a perceptual category that is particularly central to one's concept (e.g., my standard image of 'cat' looks a lot like my family's cat), and hence likely to receive a greater bias. Or maybe there are a variety of exemplars and which exemplar is tokened depends on broader contextual factors. Or maybe both. I've already speculated enough in this paper, so I won't pretend to know which of these is right. But they are options, and IEV seems to have relatively few of those. In every case I've discussed, full explanations will require diving in detail how to the associative processes work, how the biasing signal is enacted, etc. My arguments thus far have only been meant to suggest that these investigations should be undertaken under the umbrella of EEV, not IEV.

7. Conclusion

The kinds of cases I've discussed involve a wide range of cognitive effects on perception—including feature modification, feature priming, and selective attention. On the traditional way of framing the debate, we have to ask which of these types of influences *really counts* as penetration. EEV, however, offers a unified perspective on all of them: cognition influences perception by potentiating or biasing sets of perceptual processes. The different effects play out due to the specifics of the case.

Similarly, many discussions in the debate suggest that there is a particular *level* at which perception is encapsulated ("early vision"), versus where it is penetrated ("late vision"). The kinds of effects I have talked about, in contrast, range across distinct levels of perception, from basic feature percepts to objects in contexts. Theorists also differ on whether penetrating influence must be *direct* or whether it can be indirect, where "directness" is described differently by different theorists (see, e.g., Raftopoulos, 2009). EEV, as near as I can tell, is entirely agnostic about how directly the interaction comes about.

EEV, then, doesn't play by the same rules as much of the philosophical discussion of cognitive penetration, and I take this to be a good thing.   If I am right, then there is no deeper question to be asked about the relationship between cognition and perception other than determining the ways in which EEV-style relationships are brought about in particular contexts. This increase in clarity is a pragmatic gain, although I fully admit that EEV itself would require considerable development and refinement before it could genuinely be said to be a mechanistic account of the interaction.   How this particular development will go, and what the ramifications of the EEV approach are for the broader epistemological issues surrounding cognitive penetration, I leave up for future discussion.

References

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583-609.

Barsalou, L. W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences, 22*(4), 577-609.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245-283). Oxford: Oxford University Press.

Brunel, L., Goldstone, R. L., Vallet, G., Riou, B., & Versace, R. (2013). When seeing a dog activates the bark: multisensory generalization and distinctiveness effects. *Experimental psychology, 60*(2), 100.
Delk, J. L., & Fillenbaum, S. (1965). Difference in perceived colour as a function of characteristic colour. *The American Journal of Psychology, 52*(2), 255-265.
Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
Firestone, C., & Scholl, B. (Forthcoming). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*.
Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.
Fodor, J. A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT press.
Fodor, J. (2007). The revenge of the given. In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 105-116). Malden, MA: Blackwell.
Fodor, J. A., & Pylyshyn, Z. W. (2015). *Minds without meanings: An essay on the content of concepts*: MIT Press.
Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.
Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69-78. doi: 10.1002/wcs.26
Heck, R. (2007). Are there different kinds of content? In J. Cohen & B. P. McLaughlin (Eds.), *Contemporary debates in philosophy of mind* (pp. 117-138). Malden, MA: Blackwell.
Herz, R. S., & von Clef, J. (2001). The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions? *Perception, 30*(3), 381-392.
Heurley, L. P., Milhau, A., Chesnoy-Servanin, G., Ferrier, L. P., Brouillet, T., & Brouillet, D. (2012). Influence of Language on Colour Perception: A Simulationist Explanation. *Biolinguistics, 6*(3-4), 354-382.
Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*(8), 985-1018.
Huettig, F., & Altmann, G. T. (2011). Looking at anything that is green when hearing "frog": how object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology, 64*(1), 122-145.
Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*(12), 91-97.
Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology, 22*(3), 333-347. doi: 10.1080/02643290442000130
Kulvicki, J. (2005). Perceptual content, information, and the primary/secondary quality distinction. *Philosophical Studies, 122*(2), 103-131.
Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: the role of race categories. *Journal of Experimental Psychology: General, 135*(4), 501.
Lupyan, G. (2014). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*.

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.

Millikan, R. G. (2000). *On clear and confused ideas: An essay about substance concepts*: Cambridge University Press.

Mishra, R. K., & Marmolejo-Ramos, F. (2010). On the mental representations originating during the interaction between language and vision. *Cognitive Processing, 11*(4), 295-305.

Mole, C. (2015). Attention and Cognitive Penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 218-238). Oxford: Oxford University Press.

Nanay, B. (2014). Cognitive penetration and the gallery of indiscernibles. *Frontiers in psychology, 5*.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *Journal of Vision, 8*(5), 11-16. doi: 10.1167/8.5.13

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature reviews. Neuroscience, 5*(4), 291-303. doi: 10.1038/nrn1364

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*(3), 793-797.

Purves, D., Shimpi, A., & Lotto, R. B. (1999). An empirical explanation of the Cornsweet effect. *The Journal of Neuroscience, 19*(19), 8542-8551.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *The Behavioral and Brain Sciences, 22*(3), 341-365; discussion 366-423.

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*: MIT press.

Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neural science inform philosophy?* : Mit Press.

Rock, I. (1983). *The logic of perception*.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Clarendon Press.

Siegel, S. (2012). Cognitive Penetrability and Perceptual Justification. *Noûs, 46*(2), 201-222.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153-156.

Stokes, D. (2012). Perceiving and desiring: A new look at the cognitive penetrability of experience. *Philosophical Studies*, 1-16.

Stokes, D. (2014). Cognitive penetration and the perception of art. *Dialectica, 68*(1), 1-34.

Toribio, J. (2015). Visual experience: rich but impenetrable. *Synthese*. doi: 10.1007/s11229-015-0889-8

Wu, W. (2013). Visual spatial constancy and modularity: Does intention penetrate vision? *Philosophical Studies, 165*(2), 647-669.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224-245).

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive science, 28*(4), 611-619.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171.