This is a penultimate draft. Please cite the version published in the *Consciousness and Cognition*, https://doi.org/10.1016/j.concog.2017.12.003.

When is cognitive penetration a plausible explanation?

Valtteri Arstila

valtteri.arstila@utu.fi

Department of Philosophy, University of Turku, 20014 Turku, Finland

Turku Institute for Advanced Studies, University of Turku, 20014 Turku, Finland

Abstract: Albert Newen and Petra Vetter argue that neurophysiological considerations and psychophysical studies provide striking evidence for cognitive penetration. This commentary focuses mainly on the neurophysiological considerations, which have thus far remained largely absent in the philosophical debate concerning cognitive penetration, and on the cognitive penetration of perceptual experiences, which is the form of cognitive penetration philosophers have debated about the most. It is argued that Newen and Vetter's evidence for cognitive penetration is unpersuasive because they do not sufficiently scrutinize the details of the empirical studies they make use of—the details of the empirical studies are crucial also when the studies are used in philosophical debates. The previous does not mean that cognitive penetration could not occur. Quite the contrary, details of the feedback connections to the visual perceptual module and one of the candidates presented by Newen and Vetter suggest that cognitive penetration can occur in rare cases.

Keywords: cognitive penetration, visual perception, cognition, reentrant processing

1 Introduction

Albert Newen and Petra Vetter argue for the cognitive penetration of perceptual experience, a phenomenon in which cognitive states or processes "directly influence our perceptual experience" when certain conditions (e.g., attention, the object or scenario causing the visual input, and perceptual conditions) are kept constant (2017, p. 26). Moreover, they argue that the burden of proof is on those who reject that cognitive penetration occurs, and they consider the relationship between perception and cognition in the light of cognitive penetration.

Several points raised by Newen and Vetter invite extensive discussion, but I will limit my commentary in two ways. First, Newen and Vetter present two types of arguments for cognitive penetration: those based on neurophysiological considerations, and three specific candidates for cognitive penetration based on psychophysical experiments. I will mostly focus on the first group of arguments, although I comment on the latter group as well. This is because the neurophysiological arguments that Newen and Vetter put forward are new within the philosophical debate about cognitive penetration, whereas the latter group of arguments is more "business as usual" and some of the arguments they put forward have been presented (and opposed) before.

Second, Newen and Vetter discuss two different claims: the *strong impenetrability claim*—the claim that "all processes forming our visual experience are cognitively impenetrable" (Newen & Vetter, 2017, p. 27)—and the *weak impenetrability claim*—the claim that early visual processes are cognitively impenetrable. I will focus only on the former. On the one hand, this is because I agree with Newen and Vetter that the weak impenetrability claim is implausible for empirical reasons. On the other hand, almost all philosophers who write about cognitive penetration focus on perceptual experiences, as do Newen and Vetter in their definition of the phenomenon cited above. Yet, it is only the strong impenetrability claim that concerns perceptual *experiences*; the weak claim concerns *early visual processing,* and no empirically plausible theory of conscious experience maintains that a mere activation of early visual processing realizes our experiences. Accordingly, elaborating on the weak impenetrability claim misses the point that most philosophers address.

It should be noted, however, that the strong impenetrability claim is ambiguous in two ways.[1] First, the claim can be interpreted as inclusive of the weak impenetrability claim or independent of it. The first interpretation is based on the fact that when Newen and Vetter introduce it, they refer to Firestone and Scholl (2015) who, in turn, argue against the cognitive penetration of visual processing in all its forms, including unconscious visual processing. In this paper, however, the strong impenetrability claim is interpreted in the second way. One reason for this is that the claim concerns visual experiences and not all early visual processing play a role in forming visual experiences. This entails that the claim does not apply to all early visual processes. Moreover, this interpretation concurs with the fact that Newen and Vetter think that the strong impenetrability claim could (in principle) incorporate the evidence against the weak impenetrability claim. If the strong claim were to contain the weak claim, then this would not work, as the incorrectness of the weak claim would entail the incorrectness of the strong claim.

A more contentious issue as regards Newen and Vetter's description of the strong impenetrability claim—"all processes forming our visual experiences"—is that the description lends to two interpretations too. On the narrow interpretation, the claim comprises only the realizers of visual experiences. On the broad interpretation, the claim comprises also the processes that precede

[1] I thank an anonymous reviewer for pressing me on this point as well as for several other constructive suggestions.

and have causal influence on the realizers of visual experiences. In addition to cortical processes, these processes would include pre-cortical processes and perhaps even processes one might prefer to think as non-visual. In this commentary, which focuses mainly on Newen and Vetter's neurophysiological considerations, the description is interpreted in the narrow sense because they themselves focus on the direct top-down neural connections to the "visual perceptual module". Thus, the narrow interpretation allows us to assess the issue their neurophysiological considerations concerned, namely the nature of the described direct connections to visual perceptual module and their effects on perceptual experiences. The narrow interpretation also concurs with the fact that Newen and Vetter emphasize the directness of cognitive influence on perceptual experiences (i.e., the quoted definition at the beginning).[2]

The main objective of this commentary is not to argue that cognitive penetration occurs or does not occur. Instead, by focusing on the shortcomings of Newen and Vetter's argumentation, the objective is to illustrate how crucial it is (also for philosophers) to pay attention to the details of empirical studies if they wish to use empirical results and theories in the debate concerning cognitive penetration. Although cognitive penetration might appear plausible when general claims are considered, it can be implausible when one looks at the details or the context behind the generalized claims. Moreover, even if cognitive penetration is the best explanation for a given phenomenon, which in fact is possible for some of the cases discussed below, the details suggest to what extent the phenomenon can be generalized.

2 Neurophysiological considerations

Newen and Vetter's objection against the weak and strong impenetrability claims is based on three different kinds of claims related to neurophysiology:

(i) All brain areas are heavily interconnected with other areas of the brain and much of this is recurrent in kind.[3]
(ii) The processing in "higher-level" cortical areas can begin at least as early as the processing in the primary visual cortex, and the former can thus influence the latter much faster than previously thought.
(iii) Auditory stimuli influence processing in the primary visual cortex in a category-specific manner.

These three claims are well-supported by the empirical evidence, and my disagreement with the authors does not concern the veracity of these claims, but rather their relevance for the strong impenetrability claim. Newen and Vetter, of course, argue that these three claims make the strong

[2] This does not mean that Newen and Vetter must be committed to the narrow interpretation and in fact, at the end, I present reasons why they should not do so. Nonetheless, in keeping with the spirit of their argumentation, the neurophysiological considerations in the next section will concern only the narrow interpretation—the systematic investigation of indirect influences of cognitive states to perceptual experiences influences is beyond the scope of this paper and must wait until another time.

[3] In recurrent processing, the information is transmitted from "higher-level" (or later) areas back to "lower-level" (or earlier) areas such as the primary visual cortex.

impenetrability claim untenable.[4] This, they argue, is because the high interconnectedness between each brain area makes the perceptual module "smart" (illustrating category-specific processing) and wide (e.g., incorporating visual and auditory processing) when it should be an impenetrable and encapsulated perceptual module (2017, p. 29). According to Newen and Vetter (2017, p. 32), this "presupposes an implausible version of a module."

Focusing on the plausibility of the impenetrable and encapsulated perceptual module is a wrongheaded way to approach the topic at hand however. This is because the truth of the strong impenetrability claim does not depend on the possibility of there being a plausible notion of such a module. Instead, its truth depends on whether the source of the possible penetration can be considered cognitive and whether the penetration influences perceptual experience rather than mere parts of the module.

The third claim above, for example, demonstrates the influence of non-visual processes on the processing in the primary visual cortex. But even if such influence does exist, it does not follow that non-visual processes bring about changes in perceptual *experience*. In fact, I will ignore the third claim since it is mostly based on a study that "does not directly show influence on visual perception *per se*". (Newen & Vetter, 2017, p. 31, their emphasis)[5]

What I want to propose is that if one approaches the strong impenetrability claim in terms of neurophysiology, it is better to first focus on the neural correlates of visual experiences and then consider the reasons why those neural correlates are or could be influenced by other (sub)cortical areas. In practice, this means that three interrelated questions must be addressed: *First*, what are the neural correlates of the contents of our perceptual experiences? This question concerns the notion of the visual perceptual module that one should use regardless of whether the module is impenetrable and encapsulated. *Second*, is the processing within this module impenetrable? If it is, then the strong impenetrability claim holds. If it is not, then the *third* question arises: in this case, is there a reason to think that this penetration of the perceptual module is cognitive in nature?

The answers to the first two questions are straightforward. Beginning with the first question—the one concerning the notion of the visual perceptual module—many prominent vision scientists hold that the neural correlates of the contents of our visual perceptual experiences[6] are confined to the occipito-temporal cortex, although such experiences may also involve the occipito-parietal-temporal "hot-zone" (for recent reviews, see, for example, Koch, Massimini, Boly, & Tononi, 2016; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013). It is unquestionable that this module is a complex recurrent network. In other words, contrary to what one might think based on Newen and Vetter's argumentation, these scientists maintain that there is a sensible notion of the visual perceptual module. Moreover, this notion does not contradict the views of the scientists that Newen

[4] The weak impenetrability claim is directly contradicted by the third claim and it also untenable in light of the discussion related to the other two claims.

[5] The other two cases that Newen and Vetter very briefly mention here can be explained in two well-established ways. First, the *attention-shift interpretation* holds that attentional factors change perceptual experiences. However, changes caused by attention are not cases of cognitive penetration. Second, the *judgment interpretation* holds that our judgments about our experiences change, but the perceptual experience remains the same. In this case, perceptual experiences do not change and hence this is not a case of cognitive penetration. (For these interpretations, see Stokes, 2013)

[6] The neural correlates of the contents of consciousness (in this case experiences of, say, color, visual motion, etc.) must be separated from the neural correlates of consciousness. While finding the former has been considered straightforward, finding the neural correlates of the state(s) of consciousness is a matter of considerable debate.

and Vetter themselves refer to. Petro, Vizioli and Muckli (2014), for instance, also separate higher visual areas from other "higher" cortical areas. The authors also refer to Scholte, Jolij, Fahrenfort and Lamme (2008), scientists who argue that feedback connections from the higher-level *visual* areas are required for surface segregation. Scholte and his colleagues also explicitly state that their study did not concern recurrent processing from areas outside the visual cortex. This concurs with Lamme's (2006) distinction between *localized* and *widespread* recurrent processing: whereas the former is confined within the visual cortex and brings about visual perceptual experiences, the latter also includes higher areas (the frontal cortex and prefrontal cortex) and enables cognitive accessibility to the processing occurring within the visual cortex.

While there is thus a notion of the visual perceptual module which many find sensible enough to be used in theories, the processing within this module is not thought to be impenetrable or informationally encapsulated. On the contrary, as already illustrated by Lamme's notion of widespread recurrent processing, there are recurrent processes in the visual cortex that originate from outside the visual cortex. Others agree (e.g., Koch et al., 2016; Kravitz et al., 2013; Petro et al., 2014). This means that the answer to the second question—the one concerning whether the processing within the visual perceptual module is impenetrable—is negative.[7] Thus, the truth of the strong impenetrability claim depends on the answers given for the third question, which is whether the penetration of the perceptual module (and not merely the primary visual cortex) is cognitive in nature.

To address this question, we must determine what cortical and subcortical areas have feedback connections to the visual perceptual module, and whether the functions these areas serve are such that the feedback could be counted as cognitive penetration. Newen and Vetter's discussion provides little help on these details as most of their comments are very general in kind (e.g., "each brain area is connected to 66% of the rest of the brain"). They do provide two detailed examples of the feedback connections though, both of which concern the timing of neural processing. While the first example, the influence of V5 on V1, can be ignored because this influence occurs within the visual perceptual module, the second example (the influence of the frontal eye field (FEF) on many different areas of visual processing) is discussed shortly.

Newen and Vetter's comments are presumably general because they think that "the exact role of feedback connections is under-researched, e.g. with respect to the nature of the information they exactly convey and how exactly they are involved in visual processing" (Newen & Vetter, 2017, p. 29). Contrary to their claim, however, there appears to be enough information to assess the nature of this possible penetration of the visual perceptual module and thus address the third question properly instead of making the question a matter of the plausibility of some very general remarks. In short, in addition to connections to the visual dorsal pathway, which is not related to cognitive processes, the visual perceptual module (occipito-temporal network) receives feedback from four regions (Kravitz et al., 2013): The orbitofrontal cortex, the amygdala, the ventrolateral prefrontal cortex (including FEF) and the medial temporal area. This means that the visual perceptual module is heavily interrelated with only *some* cortical and subcortical areas. If cognitive penetration occurs, it happens via these feedback pathways and the penetration reflects the functions of these areas. Obviously, there can be pathways that are currently unknown and cognitive penetration can also occur indirectly (e.g., by influencing pre-cortical visual processing). Nonetheless, because these

[7] It is worth noting that philosophers too have argued that there is a sensible notion of perceptual module even though the module is not informationally encapsulated and impenetrable (e.g., Lyons, 2015).

four feedback pathways and their functions exhaust and extend the ones discussed in the papers cited by Newen and Vetter (i.e., Gilbert & Li, 2013; Petro et al., 2014), in what follows, I will briefly discuss each of these areas and the question of whether they constitute cognitive penetration (more details can be found in the cited papers).

The first two areas having bidirectional connections with the visual perceptual module are associated with the processing of emotional stimuli. The first, *the orbitofrontal cortex*, has been implicated to be involved with the processing of rewarding stimuli. *The amygdala*, the only subcortical area having bidirectional pathways with the visual perceptual module, has been associated with the processing of threatening stimuli as well as other emotional stimuli such as rewarding stimuli. The feedback projections from these two areas enable reward processing to influence the processing of the primary visual cortex (for several examples, see Petro et al., 2014). More importantly, however, the processing in the amygdala changes valence of various stimuli. That the amygdala has feedback connections to most parts of the visual perceptual module (and not, say, only the primary visual cortex) is also significant. Thus, there are good reasons to believe that the activity of the amygdala can influence visual perceptual experience. Consequently, it appears a likely possibility that the valence of stimuli can influence perceptual experience and that this constitutes a case of cognitive penetration when the influence is due to the activity of the amygdala. However, proving this to be the case might be difficult (if not impossible) in practice because the amygdala also regulates attention and, by definition, changes in perceptual experiences caused by spatial attention are not cases of cognitive penetration. Thus, it could be that the (possible) cognitive penetration effects of the amygdala on perceptual experiences can be separated from the attentional effects caused by the amygdala only in theory.

The areas of *the ventrolateral prefrontal cortex* that have feedback connections to the visual perceptual module contribute to the maintenance and manipulation of visual information. These functions refer to the guidance of attention, working memory and switching task sets—functions corresponding with those listed by Gilbert and Li (2013) to whom Newen and Vetter refer. It is not self-evident, however, that these functions would constitute cognitive penetration.

Simply put, the issue is this: although selective attention and working memory were once considered distinct, recent studies have shown that the two overlap extensively (for reviews, see Baddeley, 2012; D'Esposito & Postle, 2015; Fougnie, 2008). This is particularly the case concerning the manipulation and updating of information within working memory (including switching task sets) since these functions depend upon selective attention to the extent that most models of working memory regard the "central executive" required for these tasks as an attentional system. Given that the change in perceptual experience caused by spatial attention does not constitute cognitive penetration, the change in perceptual experience caused by the manipulation of working memory and task sets should not be regarded as cognitive penetration either. One might respond to this by arguing that the maintenance of information (and task sets) in working memory does not require attention. However, while this is true, such maintenance is also not a case of cognitive penetration because it does not change experiences—it merely maintains the experiences caused by stimuli, attentional shifts and task sets. Accordingly, in light of these considerations, the feedback from the ventrolateral prefrontal cortex to the visual perceptual module does not challenge the strong impenetrability claim.

It is worth noting that the previous conclusion also holds for FEF, the area of the ventrolateral prefrontal cortex which Newen and Vetter refer to in their argument for how higher areas can influence lower areas of visual processing. While it is true that FEF projects on many areas

of the visual cortex, Newen and Vetter ignore the fact that FEF serves the functions of controlling visual selective attention as well as planning and executing saccadic eye movements. Although the top-down influence of this higher area on the visual perceptual module is thus uncontroversial, it is almost equally uncontroversial that this influence does not constitute cognitive penetration. This is because FEF controls visual attention and different perceptual experiences caused by the different locus of spatial attention are not cases of cognitive penetration. Although shifting gaze is not explicitly excluded by Newen and Vetter (but it is, for example, by Stokes, 2012), they would presumably not claim that different perceptual experiences due to people looking at different parts of the stimuli or visual field in general constitutes cognitive penetration. After all, given the small receptive field of the fovea, looking at different parts of the large stimuli or the visual field essentially amounts to looking at different stimuli.

In the paragraphs above, cognitive penetration was excluded partly because the possible changes in perceptual experiences would be due to spatial attention, attention towards a particular location in a stimulus or visual field. This form of attention was highlighted because, like many others working on the topic of cognitive penetration, Newen and Vetter too write only about spatial attention. Accordingly, the previous argumentation does not appear to pertain to another major form of attention, namely to feature-based attention in which attention is directed towards a feature dimension (e.g., motion vs. color) or feature attribute (e.g., red vs. orange). Hence one response to the argumentation above is to maintain that cognitive penetration is compatible with feature-based attention, as suggested by Fiona Macpherson (2012) and Dustin Stokes (2017), and that the cortical areas in question control not only spatial attention but also feature-based attention.

This response is not successful in this case though. On the one hand, this is because the attentional factors related to the discussed areas of the ventrolateral prefrontal cortex are mostly spatial in nature. Task sets and spatial attention, for example, are closely coupled (Longman, Lavric, & Monsell, 2016) and the activity of FEF is mainly related to spatial attention (Ramkumar et al., 2016). On the other hand, the effects of feature-based attention are mostly the result of bottom-up processing originating from the visual cortex (Theeuwes, 2013). Indeed, the rare cases in which feature-based attention modulates processing in the visual perceptual module in a top-down manner, the modulation originates from the *ventral prearcuate* region of the prefrontal cortex and proceeds to the visual perceptual module by guiding eye movements and spatial attention processes in FEF (Bichot, Heard, DeGennaro, & Desimone, 2015; Zhou & Desimone, 2011). That is, the top-down influence of feature-based attention is intrinsically linked to spatial attention, which excludes cognitive penetration.[8]

It is also not evident that feature-based attention differs from spatial attention as regards the issue of cognitive penetration in the first place. On the one hand, it is possible that the reason why some (but not all) philosophers write about spatial attention is merely because only spatial attention relates to the well-known and much debated candidates for cases concerning cognitive penetration (e.g., those involving bistable figures). On the other hand, the reasons Stokes and Macpherson give for their view are not particularly persuasive from an empirical point of view. Stokes cites three empirical studies as evidence for the separation of feature-based attention and spatial attention. His argumentation is unpersuasive, however, because two of these studies concern saccades. As discussed

[8] See also (Leonard, Balestreri, & Luck, 2015) for the results showing that the evidence for the independence of feature-based attention from spatial attention is not as strong as it has been taken to be.

above, saccades are controlled by FEF, whose activation is mainly related to spatial attention. The third study involves neurophysiological evidence compatible with Theeuwes' claim that the effects of feature-based attention are mostly the result of bottom-up processing. Hence, Stokes does not provide any reason to think that feature-based attention has a top-down influence without spatial attention shifts.[9] One of Macpherson's reasons, in turn, relates to a specific case she considers, which is a study by Delk and Fillenbaum (1965).[10] In this study, subjects were shown an orange patch. Macpherson holds that orange is a combination of phenomenal yellow and red, and according to her, in situations involving cognitive penetration, the subjects attend to the redness component of the orange patch. As the component is a feature of non-spatial feature (i.e., color), this cannot be a form of spatial attention. This reason is questionable, however, because there is no evidence that—in addition to a feature dimension (e.g., motion vs. color) and a feature attribute (e.g., red vs. orange)—feature-based attention can be directed to the feature of a feature when colors are attended (e.g., redness of orange).[11] Without evidence that this can occur, there is no reason to accept Macpherson's claim that feature-based attention would be compatible with cognitive penetration in real life—a conclusion that concurs with the evidence suggesting that feature-based attention and especially its top-down effects are intrinsically related to spatial attention. Empirical evidence, therefore, does not support the view that feature-based attention would somehow differ from spatial attention with regard the issue of cognitive penetration.

Finally, the visual perceptual module also relays and receives feedback information to and from *the medial temporal lobe*, which is essential for long-term object and object-context memory. In more detail, the visual perceptual module projects mainly to the perirhinal cortex and, to a smaller extent, also to the parahippocampal cortex, the entorhinal cortex and the CA1 (cornu ammonis 1) of the hippocampus. As regards the feedback connections, the hippocampus's strongest output connections are to the entorhinal cortex, and there exist much less significant output connections to the prefrontal cortex and the parahippocampal cortex. The entorhinal cortex, in turn, projects to the perirhinal cortex and the parahippocampal cortex. That is, the hippocampus and the entorhinal cortex do not have (at least significant) direct feedback projections to the visual perceptual module. However, such connections exist for both the perirhinal cortex and the parahippocampal cortex.

The perirhinal cortex, which is important for the recognition of familiar objects, forms the major boundary between the visual perceptual module and the medial temporal lobe. This makes it the more interesting of the two areas with feedback connections to the module, especially as the traditional view, according to which the perirhinal cortex and the rest of the medial temporal lobe only support memory functions, has been challenged. The challenge comes from studies showing that a lesion on the perirhinal cortex impairs performance in tasks in which subjects must discriminate

[9] Stokes' argumentation is unpersuasive also because these studies did not demonstrate that feature-based attention influences perceptual experience, something which must happen in cognitive penetration, but only the recollection and identification of stimuli and the reaction times to stimuli.

[10] Macpherson argues also that feature-based attention is *conceptually* compatible with cognitive penetration. Given that my focus here is on empirical issues and real-life possibilities, I will not address this argument.

[11] It is questionable also because of problems related to the credibility of Delk and Fillebaum's results, which Macpherson aims to explain (i.e., there might not be anything to explain, see section 3), and the validity of the evidence for the color orange possessing a component of redness, which Macpherson's argument necessitates (for a recent critical review on the matter, see Arstila, 2017b).

between different stimuli. These findings have been taken to suggest that the traditional view is false and the perirhinal cortex is an extension of the visual perceptual module with the function of representing conjunctions of visual features (Graham, Barense, & Lee, 2010; Murray, Bussey, & Saksida, 2007). This would mean that the perirhinal cortex is also a perceptual area, and there would be no justified reason to separate the memory and perceptual functions it serves. If this is correct, then our memories would influence our perceptual experiences without being mediated by, say, attentional factors. As a result, the strong impenetrability claim would be false. More recently, however, this alternative interpretation of the function of the perirhinal cortex has lost much of its appeal. This is because of findings such as those showing that the performance of patients with perirhinal cortex lesions is not impaired if they are allowed to use a pencil rather than rely on memory alone (Knutson, Hopkins, & Squire, 2013), and thus lesions on the perirhinal cortex do not impair perception. Accordingly, the results previously interpreted as suggesting, contrary to the traditional view, that the perirhinal cortex plays a role in visual perception are now only thought to illustrate the memory demands of the tasks involved (e.g., Knutson, Hopkins, & Squire, 2012; Suzuki & Naya, 2014).

The parahippocampal cortex has only weak connections to the visual perceptual module, but these connections might enable cognitive penetration. In humans, the parahippocampal cortex processes spatial information related to the peripheral aspects of visual scenes—it responds very little to information presented in the fovea. Accordingly, if feedback from the parahippocampal cortex to the visual perceptual module causes cognitive penetration of perceptual experience, then this influence would be limited to the peripheral visual field.[12]

Hence, the influence of memory on visual perceptual experience due to direct links between the medial temporal area and the visual perceptual module appears to be very limited. This conclusion in fact concurs with the claims made by Petro et al. (2014), whom Newen and Vetter refer to, since most of the memory-related examples Petro et al. provide concern either working memory or retrieval of memories in studies concerning activation of the primary visual cortex (not perceptual experiences). Indeed, the only study that Petro et al. cite that might illustrate how long-term memory influences perceptual experience concerns scene extrapolation (i.e., Chadwick, Mullally, & Maguire, 2013), which concurs well with the functional properties of the parahippocampal cortex discussed above.

To summarize, Newen and Vetter (2017, p. 29) argue that "each brain area is connected to 66% of the rest of the brain". Because of such dense interconnectedness "the claim of a functionally encapsulated and impenetrable perceptual module is simply not plausible, neither for the weak nor for the strong impenetrability claim." (2017, p. 29) I have argued that such a sweeping general claim misses the point as regards the strong impenetrability claim. This is because the possibility that the module is not encapsulated does not mean that cognitive penetration occurs. On the one hand, the influence of other (sub)cortical areas on the module may not result in change in the perceptual experience. On the other hand, it could be that the influence is of the kind that does not amount to cognitive penetration. That is to say, even if part of the feedback processing of some part of the visual perceptual module originates from the area external to the module, and even if this did change

[12] This would also explain why this kind of cognitive penetration has largely gone unnoticed: most of studies present the stimuli on the fovea and the parahippocampal cortex does not influence the processing of this part of the visual field.

perceptual experience, it does not necessarily amount to the cognitive penetration of the perceptual experience.

Contrary to Newen and Vetter's claim, there also exists enough information to assess the nature of the feedback connections to the visual perceptual module. This information suggests that if cognitive penetration takes place, it takes place on a much smaller scale than Newen and Vetter's argumentation implies. To begin with, most of the feedback processing within the visual cortex is local recurrent processing. (This means that much of the predictive coding is local in this sense as well.) Accordingly, even if some part of the visual cortex has connections to 66% of the other cortical areas, almost all the feedback connections are confined within the visual perceptual module. The examples Newen and Vetter provide for the cognitive penetration of perceptual experience were either part of the local recurrent processing (the influence of V5 to V1) or did not amount to cognitive penetration (the influence of FEF to the module).

Some of the feedback connections do originate from outside the visual perceptual module though. Even so, most of these do not amount to cognitive penetration because they involve changes in either attention or the direction of gaze, or do not appear to change perceptual experience. In light of current knowledge about neurophysiology and the cortico-(sub)cortical connections, the most promising candidates for cases of cognitive penetration relate to peripheral scene perception and stimuli involving valence. Notably, our memories and knowledge do not appear to have direct means of influencing our perceptual experience.

What I have argued concurs with the view presented by Kravitz et al. They (2013, p. 43) argue that much of the visual perceptual module "does not receive direct feedback from the output targets [except from the amygdala and some also extend from FEF]" and this is "why learning and task effects within the occipitotemporal network [i.e., the visual perceptual module] are generally small". As a result, most parts of the module only receive indirect feedback from the external sources. The authors also speculate that this "lack of direct feedback might allow these areas to more faithfully reflect the statistics of visual experience, creating stimulus representations that are more general and capable of contributing to adaptive action in many different contexts." (Kravitz et al., 2013, p. 43)[13]

Three caveats need to be noted regarding the claims I presented here however. First, undiscovered connections could always still exist, so the situation may change in the future. Second, previous argumentation has focused on the bidirectional connections, thus ignoring the one-way connection from the auditory cortex to the primary visual cortex. Yet, as mentioned above, such a connection has not been shown to bring about changes in visual perceptual experiences. Third, and most importantly, the above consideration follows Newen and Vetter by emphasizing the direct feedback connections to the visual perceptual module, for the reasons explicated in the introduction. Accordingly, the possibility remains that perceptual experiences are changed through indirect connections. One such possibility, and one which has also been suggested to take place in multimodal perception, is that one modality influences the processing of sensory information of another modality before the primary sensory cortices become involved.

[13] Conversely, learning does slightly influence the processing of the visual perceptual module over *long periods* of time—a claim supported by well-controlled studies (Op de Beeck & Baker, 2010). This raises the question of how such influence should be understood. For example, if the changes brought about by learning and other cognitive states are due to focusing upon different things, should the effects be classified as effects of attention (i.e., not a case of cognitive penetration) or effects of cognitive states (i.e., a case of cognitive penetration)?

3 Psychophysical case studies

Newen and Vetter present three cases for which they think cognitive penetration provides the most plausible explanation: how color concepts influence color perception, how memorized visual templates influence scene perception, and how posthypnotic suggestion influences the perceived color of digits. Their argumentation for cognitive penetration can be objected to, however, because they explicitly exclude perceptual learning from their consideration. They do so because perceptual learning cannot explain their last example (Newen & Vetter, 2017, p. 32). This line of reasoning is highly unusual because candidates for cases of cognitive penetration differ considerably, and thus it is common to explain different candidates by different means.[14] To make things worse, it has even been argued that perceptual learning accounts for their first candidate (Arstila, 2016), and it could possibly be a factor in the explanation for the second candidate as well (perceptual learning can occur after presenting a stimulus only once). Yet, rather than taking this plausible alternative explanation into account, or even explaining why their three candidates could not be explained by different means, Newen and Vetter simply resort to this unusual line of reasoning.

At this point, however, there is no need to consider perceptual learning in relation to Newen and Vetter's first two examples. Instead, it is enough to simply note that these examples are much less convincing candidates for cognitive penetration than Newen and Vetter argue them to be—after all, cognitive penetration is the best explanation for some of them because the other (more) plausible explanation is excluded from the outset. (Similarly, Aristotle's view on gravity can be considered plausible if one ignores Newton's and Einstein's theories of gravity.) More pressingly, and in keeping with the main objective of the commentary, I want to highlight the fact that the details of two studies cited by Newen and Vetter make the cognitive penetration interpretation of the first candidate problematic (regardless of the perceptual learning interpretation).

The first study is that of Delk and Fillenbaum (1965). Ignoring problems in replicating its results, an issue which oddly enough has not made philosophers weary of using it in their arguments, the detail deserving of attention is that Delk and Fillenbaum report that typically red objects were (allegedly) perceived redder than objects not typically red, even if both had the same surface properties. The second study is that of Hansen et al. (2006), and its results have been better received than those of Delk and Fillenbaum's study. Hansen et al.'s results showed that the typical color of familiar objects influenced the color perception of grey images of those objects. However, while such an effect occurred for blue and yellow objects, the authors did not study red objects. That is, contrary to Newen and Vetter's (as well as Macpherson's) argumentation, this study lends only circumstantial support to Delk and Fillenbaum's results.

Notably, later studies by the same research group and the interpretation they provide for their results puts even this support in doubt (Olkkonen, Hansen, & Gegenfurtner, 2008; Witzel, Valkova, Hansen, & Gegenfurtner, 2011). To begin with, the earlier paper, which is partly based on the same data as the one cited by Newen and Vetter, reported that the memory color effect on the image of a strawberry, a red fruit, was small or the opposite of what Delk and Fillenbaum's results would suggest (depending on the nature of the stimuli). The later paper, which included a heart shape figure similar

[14] To give three examples, Churchland's candidate involving bistable figures is explained with the attentional-shifting interpretation, Siegel's candidate involving recognitional concepts is explained in terms of the judgement interpretation, and part of Stokes' candidate involving the subjective necessity of money can be accounted for by the magnitude perception interpretation.

to that in Delk and Fillenbaum's study, reported opposite effects for the heart figure and Coca-Cola logo, as well as no memory color effect for a fire extinguisher. Moreover, the authors explain their results in terms of color constancy processing and, corresponding with their results, explicitly suggest that the color perception of typically red objects should not be influenced by our past experiences with these objects. Thus, there appear to be very good reasons to doubt Delk and Fillenbaum's results and, subsequently, that their study exemplifies cognitive penetration.

It is also not clear how cognitive penetration could account for these more comprehensive results (especially Witzel et al., 2011) either. This is because the explanation would need to account for the fact that only some of the stimuli were influenced by past experiences. For example, why would our knowledge of the blueness of a smurf influence our color experiences of smurfs, but our knowledge of the redness of a Coca-Cola sign would not influence our color experiences of the sign? Without an explanation accounting for this discrepancy, especially when there appears to be no significant differences in our cognitive states as regards these two items, the cognitive penetration interpretation is question begging.

The lesson that we can draw from the consideration above is that a plausible candidate for the case of cognitive penetration should satisfy two criteria: First, it should be based on a study whose results are relatively uncontroversial and do not undermine the cognitive penetration interpretation of the results. Second, if the phenomenon in question has been investigated in other studies, the results of these studies should also provide support for the cognitive penetration interpretation when one looks at the details (and not merely the general claims) of the studies. At best, a study is plausible when considered in a broader context as well. Of course, satisfying these criteria does not prove that the candidate in question is a case of cognitive penetration, for the results can also have other equally plausible interpretations. But if the two criteria mentioned here are not satisfied, then one should question the original results rather than ignore the details and claim that the results are due to cognitive penetration.

The final candidate for a case of cognitive penetration discussed by Newen and Vetter—namely Cohen Kadosh and colleagues' (2009) study in which a post-hypnotic suggestion changes the perceived color of grapheme—satisfies the first criterion. Moreover, even though there are objections to the interpretation Cohn Kadosh and colleagues provide for their results (as regards the functional similarity between hypnotic and developmental synesthesia), objections have not been presented against the claim that posthypnotic suggestions can induce color experience (Anderson, Seth, Dienes, & Ward, 2014). This candidate is also plausible in the broader context of hypnotic suggestions, as these have been shown to influence the perceptual experiences of both hypnotized and awake subjects (McGeown et al., 2012) and these results have been interpreted to imply cognitive penetration (Arstila, 2017a). Accordingly, there are good reasons to think that the candidate provided by Cohen Kadosh and colleagues satisfies the second criterion as well. Thus, in my view, this is the most convincing and only plausible candidate for a case of cognitive penetration presented by Newen and Vetter.

Two issues need to be mentioned as regards this candidate though. First, the candidate (and other candidates involving hypnotic suggestions) does not in fact establish that the strong impenetrability claim is false. To remind, the narrow interpretation of the claim we have employed, and the neurophysiological considerations pertaining to it, concerned only the direct top-down neurophysiological connections to the visual perceptual module. Moreover, the examination of such connections in section two suggested that the existing connections cannot bring about the kind of

effects these candidates describe (e.g., to make white paper to appear red). Consequently, assuming that the candidates and the neurophysiological considerations are sound, the only way to resolve the tension between the two appears to be that Newen and Vetter's candidate involves indirect causal connections (e.g., top-down connections to pre-cortical processes) through which cognitive states influence perceptual experiences.[15] In this case the influence of cognitive states on perceptual experiences would not happen through the kind of connections the narrow interpretation concerns and thus the candidate would not have a bearing on the issue of the strong impenetrability claim (as it was understood here). However, as I agree with Newen and Vetter in that the candidate describes a phenomenon we would like to call cognitive penetration—after all, even if indirectly, cognitive states influence our perceptual experiences in this candidate—the candidate and the neurophysiological considerations together suggest that the narrow interpretation of the strong impenetrability claim is too narrow.

Second, these candidates involve hypnotic suggestions and the studies used highly suggestible participants (such people make up only part of the population[16]). Accordingly, the kind of cognitive penetration implied by these studies is likely to take place only in exceptional situations for a small percentage of people. Thus, although the fact that it can occur is significant in itself, these studies do not provide reasons why (for most intents and purposes) we could not operate based on the view that cognition and perception are separate.

4 Final thoughts

To summarize, Newen and Vetter aim to "present the most striking evidence" for cognitive penetration. The evidence they present against the weak impenetrability claim is sound and much of it has been known for some time. However, their evidence against the strong impenetrability claim—against cognitive impenetrability in the way philosophers usually understand the phenomenon—turned out to be largely unpersuasive and unsuccessful in weakening the strong impenetrability claim.

This commentary focused mainly on the evidence related to neurophysiological connections, as such an approach is new in the discussion concerning cognitive penetration. It was argued that the evidence Newen and Vetter considered here missed the issue since they focused on the impenetrability of the perceptual module rather than the nature of the penetration of the module and the effects thereof. Accordingly, their claims about the interconnectedness of cortical areas were unpersuasive and unsuccessful in placing the plausibility of the strong impenetrability claim in doubt because, as general claims, they did not differentiate local and widespread recurrent processing. Then again, none of their detailed examples of feedback connections suggested that cognitive penetration occurs.

Newen and Vetter fared a little better with their arguments related to psychophysical studies: Two out of the three candidates for a case of cognitive penetration were implausible because

[15] This explanation is possible also because the candidate itself does not necessitate any direct link between cognitive states and penetrated perceptual experiences—by definition, the reported effects of post-hypnotic suggestions to experiences take place after the subjects are wakened from hypnosis, not at the time when the suggestions are given. Indirect causal connections remain a possible mechanism in the cases involving hypnotic suggestions too, as the possible effects of cognitive penetration took 30-120 seconds to occur (Arstila, 2017a).

[16] The prevalence rates of highly hypnotic suggestible people is around 10–18% of the general population, depending on the criteria (Barnier & McConkey, 2004).

they reject a plausible alternative explanation from the outset and ignore the details of the studies they make use of. The remaining candidate, which concerned posthypnotic suggestions, turned out to be a plausible candidate, so it is unfortunate that Newen and Vetter dedicated only one paragraph to it.

It is worth emphasizing that the previous criticism concerns Newen and Vetter's argumentation for cognitive penetration, and highlights the fact that their argumentation does not sufficiently scrutinize the details of the empirical studies they make use of. What I have tried to underscore is that the cognitive penetration interpretation is plausible only when the results it aims to explain are relatively uncontroversial, when it can account for the details of the results, and when the results concur with other possible studies on that matter. Hence, even if my criticism is successful, it does not follow that the cognitive penetration of perceptual experiences could not occur. In fact, I suggested that (i) the influence of valence on perceptual experience due to the activity of the amygdala and (ii) influences to the experiences of the peripheral visual field due to the parahippocampal cortex constitute cases of cognitive penetration. That is, the strong cognitive impenetrability claim, as it was interpreted here, is incorrect. Although it was not argued for, the weak impenetrability claim is likely to be incorrect too given the extensive connections from the amygdala to the visual cortex. Finally, I also suggested that the candidates that involve (post-)hypnotic suggestions constitute cases of cognitive penetration. However, these candidates differ from the cases of cognitive penetration above in that the candidates are likely to involve indirect neurophysiological connections to the visual perceptual module. Thus they refute the strong impenetrability claim only if the broader interpretation is adopted.

Let me end by expressing my disagreement with Newen and Vetter's claim that the burden of proof is on those who reject cognitive penetration. This appears objectionable for the simple reason that it is often easier to prove that something occurs than prove a general negative claim that something does not occur. For example, while one plausible case of cognitive penetration (or, say, the existence of ghosts) is enough to prove that cognitive penetration exists (or that ghosts exist), the non-existence of cognitive penetration cannot be proven in the same way. Instead, proving the claim that cognitive penetration does not happen appears to require the provision of plausible explanations for each candidate case of cognitive penetration. Hence, in my view, the burden of proof should be on those claiming that cognitive penetration does occur, because proving that cognitive penetration takes place requires only one successful candidate, whereas proving cognitive impenetrability requires the successful rebuttal of all candidates for cognitive penetration. Despite our disagreement on this issue, it should be noted that Newen and Vetter's argumentation illustrates their readiness to carry the burden of proving that cognitive penetration happens—and their example involving posthypnotic suggestion is a candidate that should be taken seriously by those rejecting the possibility of the cognitive penetration of perceptual experiences.

References

Anderson, H. P., Seth, A. K., Dienes, Z., & Ward, J. (2014). Can grapheme-color synesthesia be induced by hypnosis? *Frontiers in Human Neuroscience*, *8*(April), 220. http://doi.org/10.3389/fnhum.2014.00220

Arstila, V. (2016). Perceptual Learning Explains Two Candidates for Cognitive Penetration. *Erkenntnis*, *81*(6), 1151–1172. http://doi.org/10.1007/s10670-015-9785-3

Arstila, V. (2017a). Cognitive penetration, hypnosis and imagination. *Analysis*, *0*(0), 1–9. http://doi.org/10.1093/analys/anx048

Arstila, V. (2017b). What makes unique hues unique? *Synthese*. http://doi.org/10.1007/s11229-017-1313-3

Baddeley, A. D. (2012). Working Memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. http://doi.org/10.1146/annurev-psych-120710-100422

Barnier, A. J., & McConkey, K. M. (2004). Defining and identifying the highly hypnotizable person. In M. Heap, R. J. Brown, & D. A. Oakley (Eds.), *The highly hypnotizable person: Theoretical, experimental and clinical issues* (pp. 30–60). New York: Routledge.

Bichot, N. P., Heard, M. T., DeGennaro, E. M., & Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron*, *88*(4), 832–844. http://doi.org/10.1016/j.neuron.2015.10.001

Chadwick, M. J., Mullally, S. L., & Maguire, E. A. (2013). The hippocampus extrapolates beyond the view in scenes: An fMRI study of boundary extension. *Cortex*, *49*(8), 2067–2079. http://doi.org/10.1016/j.cortex.2012.11.010

Cohen Kadosh, R., Henik, A., Catena, A., Walsh, V., & Fuentes, L. J. (2009). Induced cross-modal synaesthetic experience without abnormal neuronal connections. *Psychological Science*, *20*(2), 258–265. http://doi.org/10.1111/j.1467-9280.2009.02286.x

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–42. http://doi.org/10.1146/annurev-psych-010814-015031

Delk, J. L., & Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *The American Journal of Psychology*, *78*(2), 290–293.

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *4629*(2016), 1–77. http://doi.org/10.1017/S0140525X15000965

Fougnie, D. (2008). The relationship between attention and working memory. In N. B. Johansen (Ed.), *New Research on Short-Term Memory* (pp. 1–45). New York: Nova Science Publishers, Inc.

Gilbert, C., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*. Retrieved from http://www.nature.com/nrn/journal/v14/n5/abs/nrn3476.html

Graham, K. S., Barense, M. D., & Lee, A. C. H. (2010). Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia*, *48*(4), 831–853. http://doi.org/10.1016/j.neuropsychologia.2010.01.001

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–8.

Knutson, A. R., Hopkins, R. O., & Squire, L. R. (2012). Visual discrimination performance, memory, and medial temporal lobe function. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(32), 13106–13111. http://doi.org/10.1073/pnas.1208876109

Knutson, A. R., Hopkins, R. O., & Squire, L. R. (2013). A pencil rescues impaired performance on a visual discrimination task in patients with medial temporal lobe lesions. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *20*(11), 607–10. http://doi.org/10.1101/lm.032490.113

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, *17*(5), 307–321. http://doi.org/10.1038/nrn.2016.22

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. http://doi.org/10.1016/j.tics.2012.10.011

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501.

Leonard, C. J., Balestreri, A., & Luck, S. J. (2015). Interactions between space-based and feature-based attention. *Journal of Experimental Psychology: Human Perception and Performance*,

41(1), 11–16. http://doi.org/10.1037/xhp0000011

Longman, C. S., Lavric, A., & Monsell, S. (2016). The coupling between spatial attention and other components of task-set: A task-switching investigation. *The Quarterly Journal of Experimental Psychology*, *218*(April), 1–28. http://doi.org/10.1080/17470218.2015.1115112

Lyons, J. (2015). Unencapsulated modules and perceptual judgment. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception. New philosophical perspectives* (pp. 103–122). New York: Oxford University Press.

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.

McGeown, W. J., Venneri, A., Kirsch, I., Nocetti, L., Roberts, K., Foan, L., & Mazzoni, G. (2012). Suggested visual hallucination without hypnosis enhances activity in visual areas of the brain. *Consciousness and Cognition*, *21*(1), 100–116. http://doi.org/10.1016/j.concog.2011.10.015

Murray, E. A., Bussey, T. J., & Saksida, L. M. (2007). Visual Perception and Memory: A New View of Medial Temporal Lobe Function in Primates and Rodents. *Annual Review of Neuroscience*, *30*(1), 99–122. http://doi.org/10.1146/annurev.neuro.29.051605.113046

Newen, A., & Vetter, P. (2017). Why cognitive penetration of our perceptual experience is still the most plausible account. *Consciousness and Cognition*, *47*(1), 26–37. http://doi.org/10.1016/j.concog.2016.09.005

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), 1–16.

Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in Cognitive Sciences*, *14*(1), 22–30. http://doi.org/10.1016/j.tics.2009.11.002

Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in Psychology*, *5*(OCT), 1–8. http://doi.org/10.3389/fpsyg.2014.01223

Ramkumar, P., Lawlor, P. N., Glaser, J. I., Wood, D. K., Phillips, A. N., Segraves, M. A., & Kording, K. P. (2016). Feature-based attention and spatial selection in frontal eye fields during natural scene search. *Journal of Neurophysiology*, *116*(3), 1328–1343. http://doi.org/10.1152/jn.01044.2015

Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. F. (2008). Feedforward and Recurrent Processing in Scene Segmentation: Electroencephalography and Functional Magnetic Resonance Imaging. *Journal of Cognitive Neuroscience*, *20*(11), 2097–2109.

Stokes, D. (2012). Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philosophical Studies*, *158*(3), 477–492.

Stokes, D. (2013). Cognitive Penetrability of Perception. *Philosophy Compass*, *8*(7), 646–663.

Stokes, D. (2017). Attention and the cognitive penetrability of perception. *Australasian Journal of Philosophy*. http://doi.org/10.1080/00048402.2017.1332080

Suzuki, W. A., & Naya, Y. (2014). The Perirhinal Cortex. *Annual Review of Neuroscience*, *37*(1), 39–53. http://doi.org/10.1146/annurev-neuro-071013-014207

Theeuwes, J. (2013). Feature-based attention: it is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130055–20130055. http://doi.org/10.1098/rstb.2013.0055

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *I-Perception*, *2*(1), 13–49. http://doi.org/10.1068/i0396

Zhou, H., & Desimone, R. (2011). Feature-Based Attention in the Frontal Eye Field and Area V4 during Visual Search. *Neuron*, *70*(6), 1205–1217. http://doi.org/10.1016/j.neuron.2011.04.032