UNIVERSITY OF TARTU

DEPARTMENT OF PHILOSOPHY AND SEMIOTICS

Sander Vaus

Cognitive Penetrability of Perception in Predictive Brains

Master's thesis in philosophy

Supervisor: Professor Dr. Daniel Cohnitz

Tartu 2016

Contents

1 Introduction

In this paper I will provide a defense for the hypothesis that the perceptual states that we humans enjoy can in certain cases be penetrated by our cognitive states. I will also give an account of the mechanism by which I propose this phenomenon might occur based on a computational framework of brain functioning commonly called ‘predictive coding’. A similar attempt was made by Gary Lupyan in his 2015 article entitled “Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems”. However, I contend in agreement with Fiona Macpherson (2015) that Lupyan misses the mark. Despite that, I believe that Lupyan was on the right track with connecting the debate to the predictive coding model of perception, and will provide a case for cognitive penetrability that avoids the pitfalls of his.

To get there, I will first, in section **2** turn my attention to the various ways in which the cognitive penetrability debate has taken place and to the different specifications of the cognitive penetrability hypothesis that have emerged from it. This requires an explication of both the mental states that we call ‘perceptual’ and the ones that we call ‘cognitive’, as well as the outlining of what kind of a relation between them would constitute ‘penetration’ in a philosophically interesting sense. This paves the way for an approach wherein the philosophically motivated concept of cognitive penetrability of perception can be tested on the ground of cognitive science.

In section **3** I will delineate the two general ways in which the cognitive penetrability hypothesis could be argued for – one turning on the actual brain processes underlying the mental states in question, and the other turning on the phenomenal content of the perceptual experience. I will show how the original formulation of the hypothesis as presented by Zenon Pylyshyn (1999), while useful, is conceptually problematic and how Lupyan, in following the path carved by Pylyshin, faces the same problem as he. I will conclude that the more reasonable way to approach the debate is to look at it from the angle of perceptual experiences, rather than brain functioning. I will further draw a distinction between high-level and low-level phenomena that constitute perceptual experiences, and will argue that engaging with the low-level phenomena makes more sense insofar as we are to retain the philosophically interesting aspects of the entire debate. I will then outline more specifically the required relation between perception and cognition for penetration to obtain and will provide analyses of two purported cases of cognitive penetrability, so constrained. In assessing the experiments and the attempted recreations of them, I will draw out the specific ways in which it would be permissible to deem them as presenting genuine cases of cognitive penetrability.

In section **5**, I lay out an account of perception that encapsulates the main mechanism by which I believe cognition can in fact penetrate perception. This is the “predictive coding” view of brain processing, according to which the brain engages in some form of Bayesian inference wherein top-down predictive models of the environment are tested against bottom-up error signals of phenomena that don’t match these predictions. I will show how this process unfolds in the case of perception, and then make a case, using an experimental find, for how this process can be informed by one’s cognitive states. This will constitute a case of one’s cognition directly changing one’s perceptual experience. With this done, I will turn back to the experiments on color perception discussed beforehand, and will show by the example of these cases how the predictive coding framework can be utilized to explain the phenomena found in these experiments. Finally, I will anticipate and respond to challenges that could be raised against my take on this debate.

2 Making Sense of the Debate

In order to assess the validity of the cognitive penetrability hypothesis, great care must be taken to lay out exactly the framework within which this assessment should take place. This is especially the case as there is considerable disagreement among both philosophers and psychologists over what would constitute cognitive penetrability of perception as well as over what is at stake if such an effect is in fact real.

2.1 What is at stake?

It is rarely assumed that cognitive and perceptual states are the products of wholly distinct mental processes that cannot be affected, or “contaminated”, by each other. The challenge is to propose an account of such affectedness that does not turn out to be entirely trivial. In the most general terms, it is a question of whether the non-perceptual content of our mental states could directly bear on what we perceive in a way that, even in the most minimal sense, would render us prone to seeing, hearing or otherwise perceiving that which we want, believe or intend. There are various reasons for why philosophers and psychologists would be inclined to pursue this challenge. For one, it raises epistemological questions pertaining to the reliability of sense-data and potential ways to mitigate the situation if the reliability of perception could be thus problematized. Furthermore, it is relevant insofar as we are interested in attaining a more veridical picture of the human nervous system and its workings.

Something to keep in mind, however, is that even if it turns out that perceptual systems are in some instances cognitively penetrable, this would not necessarily establish their unreliability. Depending on the details of how such penetrability might occur, its effect on perception could very well be the opposite. Given that we live in unpredictable environments wherein our sensory organs are constantly getting bombarded by an incessant amount of external stimuli, mechanisms by which to filter it all and to do it in an ecologically conservative way are naturally required so that all this information could be used to guide our behavior in ways beneficial to us. This could well mean omitting and transforming some of the incoming information, as well as filling in the bits that seem to be missing – processes that could be cognitively mediated. While all this might decrease the reliability of the system in terms of it giving us the most veridical and detailed picture of the world possible, the flipside is that without such mechanisms, the physical limitations of the brain would render the whole process of perception to be much slower, less adaptable and ultimately less reliable in terms of it giving us a passable, yet efficiently obtained, picture of the world.

2.2 A preliminary distinction between perception and cognition

What is meant by cognitive penetrability of perception? First we must assume the existence of mutually exclusive states of cognition and of perception. Perceptual states area easily the mental states caused by and about the changes in our various sensory organs. Vision allows us to have visual percepts, audition gives rise to auditory percepts, and so on, where 'percept' denotes a consciously felt state that appears to carry information about the external environment and one's physical body.[1] Things are, of course, rarely so simple. In more cases than not, any conscious percept is the product of various different sensory modalities working together, and oftentimes, interfering with each other (see for example James *et al*. 2002, Kayser and Logothetis 2008, Stein and Stanford 2008). A famous example of that is the McGurk effect, a paradigmatic case of visual sensory information interfering with auditory information, wherein seeing certain mouth movements that don't match those normally accompanying the sounds being heard cause people to mishear the auditory information (McGurk and MacDonald 1976). In other words, different perceptual modalities do not operate with just the kind of information most pertinent to any given modality, but may also take in account information from other sources in order to encode the perceptual scene available to a perceiving agent. The question is then whether a source of such information could also be non-perceptual; namely, one's cognition. Even though cross-modal perceptual penetration is well documented, establishing the penetration of perception by cognition has been decidedly less successful.

It is not immediately apparent and agreed upon what would constitute 'cognition' in this debate. For one, in various psychological literature, it is often assumed from the outset that perception *is* a kind of cognitive state. So to make sense of the entire issue at all, we need a much more specific conception of 'cognition', one that denotes a class of mental processes as *opposed* to perception. Having once established what it is exactly that is supposedly penetrating

---

[1] I am not at this point taking a stance on the representationalist/anti-representationalist debate over the contents of perceptual states. However, I *am* assuming that our perceptual apparatus has, in part, evolved in response to the informational demands of the environment – whether this is best unpacked in terms of it representing the actual environment, or something else, I leave open.

perception, the next step would be to determine also what might exactly constitute this kind of penetration.

There is quite a variety of mental phenomena that could reasonably be called a ‘cognitive state’. Some paradigmatic examples include states with propositional contents like beliefs and desires. These would also be the prime targets for the cognitive penetrability hypothesis, since it is exactly the proposed phenomena of one’s perception being susceptible to the effects of one’s beliefs, goals, intentions and the like that those invested in the debate are concerned with. Insofar as we assume a perceptual state to be something that is directly caused by the changes in one’s sensory organs that are sensitive to signals from the apparent outward environment, we can conceptualize cognitive states as something that do not necessitate this kind of a connection. So, while we can speak of perceptual states as being either appropriate or hallucinatory (a false perception) depending on whether it is elicited by the kind of sensory activity and its relation to the environment that is deemed appropriate for the given percept, no such requirements are extended to cognitive states. A belief state can be entertained in a variety of sensory situations wherein the veracity or the appropriateness of the belief state does not turn on the immediate activity of any sensory neurons at the time of entertaining that belief state. This would amount to a functional distinction between perception and cognition.

A further distinction could be drawn along the lines of the phenomenal content of perceptual experiences. This would be the first-person subjective experience of being in a perceptual state felt by the perceiving agent – the distinctive subjective quality or what-it’s-likeness of it. Cognitive states like intentions and beliefs are often taken to lack this kind of a phenomenal character.[2] Whether this is actually the case or not is of no consequence for the argument that I am going to make for cognitive penetrability. However, this aspect of perception *is* of crucial importance for a large part of it, and it seems wholly unproblematic to concede that cognitive states, at the very least, lack the *kind* of phenomenal content shared by perceptual states. I hold this to be due to the functional difference between cognition and perception outlined above – insofar as perceptual states are, at least in part, caused by the immediate apparent environment and seem to carry information about it, the experienced phenomenal

---

[2] There is no unanimous agreement on this matter, and the debate over the question of whether cognitive states are also accompanied by some special kind of cognitive phenomenology (the what-its-likeness of being in a cognitive state) is still going strong. *Cognitive Phenomenology*, edited by Bayne and Montague (2011) provides an excellent overview of this debate.

content of it is easily an effective way to present that information to the perceiver to make use of in guiding their actions within that environment. Cognitive states, not requiring such an immediate connection to the environment, would simply not benefit from having phenomenal content of this kind. Rather, the kind of content most pertinent for a cognitive state such as a belief would be that of a proposition, the truth value of which should remain constant regardless of how it feels to the cognizer.

Thus far, the distinction drawn between cognition and perception turns on the differences of their purported function and content. Another way to conceptualize this matter would be to look at the way in which these states relate to each other. To recapitulate, the distinctive content of a perceptual state is the phenomenal quality it possesses in virtue of it encoding the apparent causes for the changes in one's sensory organs. For a cognitive state, on the other hand, the content in question is propositional and, while it can be about the world, it need not have such an immediate connection with it. A way in which to make sense of the relation between cognition and perception as a functional whole could then be that perception is something that provides an agent with pertinent information about the immediate environment, whereas cognition is a means by which the agent can further make use of that information (by forming beliefs about the world, memorizing aspects of it, making decisions pertaining to it, and so on). Cognition is thus informed by perception, with information encoded by the latter constituting a source of input for the functioning of the former. If perception is cognitively *impenetrable*, then this is an asymmetric relation – the products of cognition not factoring into the workings of perception as a source of input. I will argue that this is not the case.

To reiterate, in what is to follow, we can tentatively take cognitive states to be mental processes with propositional content. Furthermore, to clearly set them apart from perceptual states, we can also assume them to be lacking in phenomenal quality (of the perceptual sort). This easily sets them apart from perceptual states, which *do* possess phenomenal character and can be argued to *lack* propositional content.

2.3 Two ways of hypothesizing the cognitive penetrability of perception: brain functioning vs perceptual experience

Working with this preliminary definition, we can now tackle the question of in what sense we could speak of perception as being penetrable by propositional attitudes. The hypothesis of cognitive penetrability of perception was first formulated in Pylyshyn (1999) as a claim that a perceptual system is cognitively penetrable if "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: 343). In the same publication, Pylyshyn makes the claim that a large proportion of the visual system, which he calls early vision, is *impenetrable* by knowledge, beliefs or expectations to the point of it being susceptible to making errors even if one is fully aware that their perceptual experience is in fact erroneous (1999: 414). A lucid example of that, according to Pylyshyn, is the great amount of perceptual illusions readily experienced by most humans wherein, despite knowing that one is experiencing an illusion, the illusion nonetheless persists.

A problem with Pylyshyn's account is that the way he sets up the cognitive penetrability hypothesis and the way he argues against it for early vision are at odds with each other. He bolsters his claim with experimental findings on visual perception, yet, the function computed by the early areas of the visual system isn't to produce a perceptual experience – this is attained only once the neuronal signals from the early visual areas are processed by higher visual areas and, ultimately, non-visual areas (namely the prefrontal cortex). It could be argued that the visual experiencing of a low level visual phenomenon is nonetheless indicative of low level processing insofar as the functions computed by higher levels of the visual stream bear no significance on the low level features of the visual scene. I will briefly consider this possibility in section **3.1.1**, but for now, the apparent mismatch between the impenetrability of early vision claim made by Pylyshin and the way in which he defends it, highlights the two significantly different ways in which the cognitive penetrability of perception could be argued for.

On the one hand, we can look at the actual functioning of the brain in the process of perception. Assuming the modularity of the brain in some minimal sense and the empirical brain-imaging data in support of it, it is possible to distinguish between the areas of the brain that encode perceptual data, and areas that are responsible for mental states with propositional content. We can then look at whether and how the activity in the latter modulates the activity in the former. While the cognitive penetrability hypothesis does not necessarily rest upon the assumption that it is the lowest parts of any perceptual stream (e.g. early vision) that have to be penetrated by the activity in cognitive brain areas for the hypothesis to hold, showing that this *does* indeed happen would clearly seem quite an attractive way to establish cognitive penetrability. If, after all, the earliest parts of the perceptual process can be shown to be cognitively penetrable, and higher levels of the process are informed by the activity in the lower levels, then the entire process is so penetrable. I will argue that this approached is ridden with conceptual difficulties, engaging with which already necessitates talk of perceptual experiences, and thus it would make more sense to conceptualize the entire debate in these latter terms.

On the other hand, then, we can look at the perceptual experience itself, and by conducting psychological experiments or just lucid self-reflection establish whether such an experience can be shaped in some interesting sense by our cognitive states. This is the way in which the cognitive penetrability of perception debate has largely unfolded. The experience in question could be that of simple, low level perceptual phenomena, or of more complex, high level phenomena. What is important here is to motivate an account of cognitive penetrability that makes use of a relation between propositional attitudes and perceptual experience that is philosophically interesting, non-trivial and semantically coherent. Cases wherein some cognitive process inadvertently causes perception to malfunction, resulting in an altered perceptual experience that, nonetheless, bears no semantic relevance to the contents of either the cognitive or the perceptual state, are not good enough. I will further elucidate this point in section **3.2.1**.

2.4 The state of the debate and a way forward

Despite the different ways in which the cognitive penetrability debate can be conceptualized, the way the hypothesis is either defended or argued against follows roughly the same pattern. On the one hand, we have the proponents of the hypothesis who present either intuitively accessible or experimentally elicited cases wherein perception appears to be cognitively penetrable. In response, the opponents either press on or further constrain the working conceptualization of the hypothesis so as to discount such cases on the grounds that they show something altogether distinct from cognitive penetrability. Which kind of cases would count as possible candidates for cognitive penetrability naturally depends on how the hypothesis is conceptualized. However, stricter conceptualizations also call for more rigorous cases, which is why the attempts to establish the hypothesis in the strongest possible sense and the arguments surrounding this task tend to converge on findings from cognitive sciences.

Cognitive science in itself, however, is not enough to solve the matter, as the hypothesis of cognitive penetrability of perception is, properly understood, not a scientific hypothesis. As ‘cognition’, ‘perception’ and the prospect of one penetrating the other are all high-level concepts that allow for considerable variance in how they are used within this debate, simply presenting an experimental finding without proper motivation for how these concepts are utilized will not be very convincing. This is evidenced by Firestone and Scholl (2015) who, in their article “Cognition does not affect perception: Evaluating the evidence for ‘top-down’ effects”, encapsulate the main concerns to which purported cases of cognitive penetrability are susceptible to. Most telling about this, however, is the fact the very title of the article rings quite strange considering that the article is devoted to analyses of cases wherein cognition *does* in fact affect perception. The issues is, rather, that the way in which this takes place, according to them, is not in the philosophically interesting sense of cognitive *penetrability*. For example, whether an instance of one’s perception being affected by one’s cognition is mediated by a shift in one’s perceptual attention or not does not in itself discredit the claim that one kind of a mental state was affected by another kind. Regardless, insofar as the hypothesis is constrained in a way that discounts intuitively accessible cases from being relevant to the claim, engaging with research from cognitive sciences is a good way forward for testing more specific claims with appropriate constrictions (e.g., whether cognition can affect perception without also causing a change in the perceiver’s perceptual attention).

In addition to testing more specific cognitive penetrability claims, the constrictions of which are hashed out on philosophical grounds, there is another crucial project for which engaging with cognitive science is required – outlining the mechanism by which cognitive penetrability, if feasible, could actually take place. For something to be considered as feasible, it should also be shown to be actually *possible*, and this is where the studying of the brain will do much more than philosophical musings over high-level conceptual claims. Such research can roughly go in two ways: (1) a bottom-up project of mapping the nervous system and the flow of information therein, and (2) a top-down project of developing computational models for specific brain functions. The latter approach has met considerable success in the field of computational neuroscience wherein a successful model of a given brain function is considered to be one that gives results that are similar to and produce phenomena that also arise in normal brain functioning. In the next section I will argue that it is approach (2) that makes the most sense for providing a feasible account for the mechanism by which perception is cognitively penetrable. I will follow Lupyan (2015) in that it is the predictive coding framework of brain functioning that accommodates the most feasible mechanism for it, but I will argue that Lupyan does not make use of it in the right way.

3 Is Perception Cognitively Penetrable?

I will start off this section by further motivating the distinction between establishing the cognitive penetrability hypothesis along the lines of brain processes vs perceptual experiences. I will conclude that while the proper understanding of perceptual experience necessitates engaging with what is known of brain functioning, establishing the feasibility of the cognitive penetrability hypothesis must still take place in terms of the perceptual experience. Thus, the most sensible way of going forward in this debate is to look at the perceptual experience itself and to conceptualize the features therein as instantiations of what goes on within the brain on a neuronal level.

It useful to distinguish between perceptual phenomena and perceptual processing. The former of the two is best studied in an agent-dependent experimental setup, whereas investigating the latter, insofar as we assume low-level brain processing to not be accompanied by conscious states, would be more privy to empirical research of brain anatomy and function. We can then speak of two ways to either establish or reject the cognitive penetrability hypothesis – one that turns on the consciously experienced perceptual states that any successful perceptual apparatus should give rise to and a functional account that looks at the flow of information within any stage of perceptual processing as the function of physicochemical processes within the nervous system. Consciously felt perceptual experiences on the one hand, the mechanism for attaining said experiences on the other. This distinction is also made by Machery (2015).

3.1 Cognitive penetrability of the perceptual process

Looking at perception as a process that unfolds within the brain and conceptualizing the cognitive penetrability of perception hypothesis as a task of discovering neuronal activity and pathways that would indicate the right kind of informational flow that could establish the hypothesis faces many difficulties. The brain is an intricately connected system, where signals from different areas of it are routinely sent to and picked up by other areas. It is true that outlining areas of it that are mostly correlated with the kind of phenomena attributable to the various perceptual modalities as well as to different cognitive processes is possible, implying the existence of segregated and localized hubs for different kinds of informational processing (see e.g., Achard *et al*. 2006). However, these areas are not informationally encapsulated from one another. Activity in perceptual areas naturally activates brain areas correlated with beliefs, whereas mental activity such as imagining can utilize areas correlated with belief in order to spike activity within the perceptual system (Agnati *et al.* 2013). In other words, the lines are heavily blurred. However, even if this was not the case, establishing cognitive penetrability by these means would still remain a difficult task. It is well known, for example, that neuronal signaling flows from the visual cortex to higher brain areas correlated with conscious thought, and it is also known these higher areas also send signals to the visual cortex (DiCarlo *et al.* 2012). The fact of information being exchanged, however, does not in itself establish penetrability in any interesting sense. Signals from the prefrontal cortex to the optic nerve are needed for the conscious control of one's eye movements, yet simply looking at different things, which might be a cognitively mediated activity, does not count as cognitive penetration.

In the most general terms, the problem here is that we simply do not know enough about the human neuroanatomy and its workings so as to either establish or discount cognitive penetrability on these grounds. In other words, while it is known that non-perceptual and presumably cognitive areas of the brain routinely affect the neuronal activity of the perceptual areas, we do not know whether this constitutes penetrability when certain conceptual constraints are put in place. Perhaps the most notable of these constraints would be the necessity to discount cases wherein changes in a person's perceptual experience can be simply attributed to shifts in their attention. I will motivate the need for constraints such as this in section **3.2.1**, but before that I will turn again to the challenge put forth by Pylyshin's original conceptualization of the cognitive penetrability claim, and discern whether it could be met by means other than the unfeasible project of mapping human brain neuroanatomy just discussed.

3.1.1 Cognitive penetrability of early vision

Recall that Pylyshyn (1999) claimed early vision to be cognitively impenetrable. Lupyan (2015) attempts to argue for the cognitive penetrability thesis from just this angle. By reference to a multitude of empirical findings, he seeks to show that early vision is in fact cognitively penetrable, as well as providing an account of the mechanism by which this takes place, which turns on the predictive coding model of perception. However, in her review of Lupyan's article, Macpherson (2015) rightly points out that Lupyan doesn't really meet this challenge. The empirical results used by Lupyan are of the variety of psychological experiments on perception which test the conscious perceptual awareness of test subjects in various borderline cases. An example would be masking experiments, wherein the test subjects' awareness of some near-threshold visual stimuli is either subdued or enforced by the experimental setup. In other words, by engaging with Pylyshyn, Lupyan fails to appreciate the very problem that also hounds Pylyshyn's article.

Why engaging with experiments on perceptual experiences is a conceptually challenging way to establish cognitive penetrability of early vision is simply for the fact that early vision is paradigmatically understood as something of which a perceiving agent cannot be aware of. The experiments that Lupyan references do deal with low-level visual phenomena like color or even the mere fact of seeing anything at all, but this does not reflect that which is considered to be the early levels of the visual perceptual system. While it is true that it is early visual areas that encode low-level visual stimuli, the very awareness of anything visual at all already presumes that the system has run its course from early vision all the way to forming a conscious visual percept. Since the latter of those is normally understood to involve higher cortical areas, specifically the prefrontal cortex, which's role in perception is not considered a part of *early vision*, the move from such experimental findings to establishing cognitive penetrability of early vision is not so obvious. This is not to say that the approach is inherently wrong – it would just require further work to establish a link between experimental findings pertaining to low-level visual percepts and early vision itself. In other words, the phenomena studied in the experiments cited by Lupyan may very well not even be phenomena tied to early vision.

As previously stated, Pylyshyn's conceptualization of the cognitive penetrability hypothesis turns on the purported function that a given perceptual system computes. For early vision, this would be to encode the light stimulus hitting the retina into simple visual features for the higher visual areas to further operate with. Producing consciously experienced percepts is not a part of that. We can certainly expect perceptual experiences of low-level visual phenomena to provide us with clues as to what is going on within the early visual system, yet the only direct way of assessing the penetrability of early vision in itself would presumably still have to take place in terms of empirically mapping the neuroanatomical processes underlying early vision. Again, we know that top-down neuronal signaling from higher brain areas to early perceptual areas is something that take place, but the concern that we lack the understanding of what this signaling exactly constitutes, still stands.

Despite these problems, I hold that Lupyan had the right idea – in order to avoid a general conceptual confusion over what is going on within the cognitive penetrability debate, providing an account of an actual mechanism by which cognitive penetrability could obtain, would be immensely useful. Such an account should be informed by what we know about the neuroanatomy of the human brain. There is good reason to believe that the predictive coding model of perception, that he proposes would best account for purported cognitive penetrability phenomena, is generally correct. Furthermore, some features of such a model (heavy reliance on top-down modulation of perceptual processing) make it an especially attractive way to frame such phenomena. I contend that this is the right way to go, and the final section of this paper is devoted to making a case for just that in a way that would avoid some of Lupyan's pitfalls. Before turning to this task however, I will clarify the general way in which the cognitive effects on perception would count as cases of penetration and review some of the most discussed evidence in support of that.

3.2 Cognitive penetrability of the perceptual experience

Framing the hypothesis of cognitive penetrability of perception in terms of the perceptual experiences acquired by the workings of our perceptual apparatus necessitates talk of the content of perceptual experience. Something that sets this content apart from the content of propositional attitudes is the first-person experience, or, what-it's-likeness of being in a perceptual state. We can call it the phenomenal content of perceptual experience, which is the special kind of content that is taken to underlie the so called "explanatory gap" of consciousness conceptualized by Levine (1983). The idea is roughly that, in physicalist terms, the experience of seeing something e.g. red can be explained as light at the wavelength of about 650 nm stimulating the retina, causing certain patterns of neuronal activation in the nervous system. This, however, says nothing as to how it feels for the first-person observer to see the color red. The question of cognitive penetrability can then be posed as the question of whether one's cognitive states can alter the quality of one's perceptual experience all the while nothing else in a given act of perception changes.

While the content of a perceptual experience can be said to be constituted or at least accompanied by its phenomenal quality to the perceiver, it is not entirely clear what should be considered the reach of this phenomenal content. We can distinguish between high-level and low-level perceptual phenomena. Low-level phenomena would be simple features of the perceptual scene such as color, visual edges and simple sounds. High-level phenomena, on the other hand, are complex combinations of low-level phenomena that allow for object recognition. In terms of the cognitive penetrability debate, cognition-effects on low-level perceptual phenomena are privy to experimental research wherein a positive result would be something akin to certain propositional states eliciting a change in, for example, the color of a perceived object. High-level phenomena, on the other hand, function rather in intuitively accessible proposed cases for cognitive penetrability. A simple example would be considering whether there is a distinct phenomenal difference in looking upon the face of one's friend, and looking upon the very same face if it belonged instead to a stranger. Intuitively, this seems to be the case, and insofar as the change is mediated by knowing that the face belongs to a friend or not, it seems to be a case of cognitive penetrability of perception. However, there is reason to discount such high-level phenomena from even being a part of the phenomenal content of one's experience. It can be argued that in such cases one's perceptual experience itself is constituted purely by lower-level phenomena and the conceptual or categorical higher-level recognition is added by non-perceptual mental processing (e.g., increased salience) (see Briscoe 2015).

Discerning whether this is the case or not cuts across an entirely different debate in the philosophy of mind, one that I would presently rather not engage with. Furthermore, for the purposes of the cognitive penetrability debate, the low-level approach also seems to make much more sense. It is exactly the possible susceptibility of low-level perceptual phenomena to cognitive modulation that seems to be the philosophically interesting question here. While high-level object recognition is certainly considered to be a part of the perceptual process in neuroscientific literature, pressing for cognitive penetrability from that angle would be to fail to engage with any interesting sense of cognitive penetrability. It is not controversial that seeing the face of a friend feels distinct from seeing the face of a stranger, and to defend the cognitive penetrability hypothesis from this angle wouldn't amount to much more than defending the hypothesis just for the sake of defending it. This would not be unlike problematizing the entire distinction between perception and cognition so as to trivially establish penetrability, and while there might be empirical reasons to do so (considering again the intricate connectivity of the brain and its functioning), this would simply ignore the interesting questions raised by the hypothesis, even if it were to turn out that the distinction between the states underlying it is, at best, analytically useful.

This also brings us back to why the constraints imposed upon either kinds of mental states and the required relation between them for penetrability to obtain are best hashed out on a conceptual level. This is a task to which I will presently turn to.

3.2.1 The relation between the cognitive and the perceptual

Let us first turn to formulating the required relation between propositional attitudes and perceptual systems so that the penetrability of the latter by the former could obtain, recalling this to be a crucial part of the original cognitive penetrability hypothesis put forth by Pylyshyn. Embedded in the task of underlining this relation is also the matter of delineating the correct kind of causal mechanism between one class of states and the other.

First, this relation must exclude cases wherein the propositional content of a cognitive state has no bearing on how it affects a perceptual state, even if it does in fact do that. An example used by Macpherson (2012) is that of having a belief that one has an important examination coming up, which causes stress and which in turn brings on migraine and the sensation of flashing lights in one’s visual periphery that accompany severe cases of the condition. While it is true in this scenario that a belief-state was responsible for causing a shift in one’s perceptual experience, this is not so in the interesting sense of cognitive penetration. That no interesting semantic coherence is obtained in this case is clear from the fact that the propositional content of the relevant belief-state bears no relevance to the content of the acquired experience. There seems to be no reason for why beliefs about important exams should be accompanied by experiences of flashing lights. Furthermore, the stress reaction and the consequent migraine could have just as well been induced by beliefs with entirely different propositional contents. For some beliefs, the accompanying experience of flashing lights might even seem pertinent (e.g., a highly stressful belief of being susceptible to being kidnapped by aliens who fly airships adorned in flashing lights), but insofar as the causal mechanism for why an experience of a certain kind is obtained does not turn on the actual content of the cognitive state (e.g., it is known instead to be caused by a migraine), cognitive penetrability of perception is not obtained.

To reframe the problem in terms of the previously stated function of perception as something that provides agents with information about the apparent environment, the worry is that the experience of flashing lights caused by a migraine is the outcome of what we would rather call the malfunctioning of the perceptual system. The visual system is under heavy duress during a migraine and the visual phenomena caused by it is the result of neurophysiological activity very much unlike during normal perception (see Kowacs *et al*. 2015). Similarly, if stressful beliefs were to cause a person to experience a stroke that leaves half their body paralyzed, we wouldn't call the consequent lack of feeling in the person's appendages an instance of one's tactile perceptual system being penetrated by a cognitive state.

The desired relation between cognition and perception should also resist trivial cases of indirect causality. For example, one's desire to eat chocolate can bring about one's experience of tasting chocolate in case the desire actually causes the person to go and eat chocolate. I call cases like this trivial, because nobody denies that cognitive states do in this way shape our perceptual world. That cognitive states can contribute causally to bringing about different perceptual experiences is certainly not to say that the latter is *penetrated* by the former. This distinction is sometimes more difficult to see. Take for example the picture of the duck-rabbit made famous by Wittgenstein, or the Necker cube. In both cases, it is possible to attain distinctly different visual percepts of one single picture. It seems like employing the concept of a duck allows one to see the picture as a duck, and the other way around with the concept of a rabbit. In the case of the Necker cube, it seems that one can almost exert one's will to see the cube as either facing up and right or down and left. However, a careful look at the matter reveals that the shift in the visual experience can be attributed to a shift in one's visual attention. Seeing either picture in one way or another is attained by essentially looking at different parts of the image. This means that the shift in the perceptual experience is due to a shift in the actual visual sense-data. This is, of course, completely normal and doesn't constitute a case of cognitive penetrability of perception.[3]

A prevailing feature among the aforementioned cases of cognition affecting perception is that something about the very situation of perceiving changes. In the case of migraine-induced flashing lights it's the atypical functioning of the retinae and the associated neuronal wiring. In

---

[3] Insofar as the stated goals for investigating the relation between perception and cognition in this debate stand, this still seems by all rights an epistemologically interesting phenomenon to study. However, I will not address this.

the case of attentional shifts, it's the scene of perception that changes. The real force of the cognitive penetrability hypothesis, however, lies in the proposal that a shift in one's perceptual experience could happen purely by virtue of some cognitive activity, all else left unchanged. I will use a formulation of this idea as put forward by Macpherson (2012):

> Perceptual experience is cognitively impenetrable if it is not possible for two subjects (or one subject at different times) to have two different experiences on account of a difference in their cognitive systems which makes this difference intelligible when certain facts about the case are held fixed, namely, the nature of the proximal stimulus on the sensory organ, the state of the sensory organ, and the location of attentional focus of the subject. (Macpherson 2012: 29)

In the next subsection, I will analyze two of the most prominently discussed experiments that seem to support cognitive penetrability in the sense conceptualized by Macpherson.

3.2.2 Empirical evidence for cognitive penetrability of the perceptual experience

Some of the strongest evidence in support of cognitive penetrability of the perceptual experience has emerged from research on color vision. In a classic study, Delk and Fillenbaum (1965) found that test subjects tend to match reddish-orange cut-outs of normally red objects (e.g. heart shapes, lips) to a redder backgrounds than they do for normally non-red objects (e.g. mushrooms, bells) and basic geometric shapes. Olkkonen, Hansen and Gegenfurtner (2008) found in a series of experiments that subjects, when asked to make achromatic color changes to photographs of fruits, are more likely to systematically shift the spectrum to the opposite direction of the fruits' normal color, away from the grey point. Both results suggest that there is a biasing effect, wherein the subjects' perceptual judgments are partly modified by either their knowledge or memory of the normal colors of the observed objects. It is, however, not certain that the test subjects' judgments reflect a genuine shift in the quality of their perceptual experience. This worry is accounted for by concerns pertaining to the experimental setup of each study as well as by the difficulties faced in attempts to recreate the results.

There could be reason to doubt that the results of the aforementioned experiments reflect genuine features of the test subjects' perceptual experiences. It may very well be that it is merely the perceptual *judgments* of participants that are sensitive to certain extra-perceptual factors, whereas the underlying perceptual experience itself could easily remain static across the tested conditions. The tasks presented to the test subjects in both experiments are by all rights difficult tasks and it is possible that the demands of the experimental setup forced the participants to fail at the tasks in an effort to actually do well in them. Compliance with the presumed demands of the task is known to be a contributing factor in such psychological experiments (see Orne 1962). In the Delk and Fillenbaum study the participants had to look at the figures through a wax paper in a dimly lit room. Given the complexity of matching the ambiguously colored figures to backgrounds in such a setting, it stands to reason that the participants could have been inclined to employ their knowledge of the typical color of observed figures as a cue by which to call for a more red or a more orange background (the test subjects didn't adjust the backgrounds by themselves – they had to verbally instruct the experimenters to make necessary alterations). None of this necessarily implies an actual shift in the participants' color experience between different cases. Insofar as we are concerned with the epistemological question of the reliability of sense-data, this is still an interesting phenomenon, but not cognitive penetrability of perception in the sense of Macpherson's conceptualization of it. Is it fair to assume that the participants were systematically and grossly mistaken in their reports about their perceptual experience? It is difficult to say, but a look at some of the attempts to recreate the study will shed some light on the matter.

Interestingly, an attempt to recreate the Dilk and Fillenbaum study by Gross *et al.* (2014) was not met with success. In addition to an experiment conceptually in line with the original study, they also presented the participants with a forced choice setup of choosing which of the two figures shown, one of which was a conventionally red object and one of which wasn't, appeared more red. In one variation of this setup, the presented images were actually identical in their chromaticity. They also tested for an El Greco effect[4]. In all the experiments, no correlation between the conventional redness of an object and test performance was found. This is possibly due to significant differences in the experimental setup, a look at which is valuable for discerning what exactly the Dilk and Fillenbaum study showed and what it did not.

---

[4] The El Greco effect refers here to a phenomenon and a strategy conceptualized by Firestone and Scholl (2015) of testing various proposed top down effects on perception. It is named in reference to a now discredited hypothesis over why the figures depicted in El Greco's paintings appear strangely elongated – proposing that it was due to his distorted vision due to astigmatism. If that were true, however, the distortion effect would have presumably also applied to the way that a canvas would appear to him, essentially nullifying the effect. Firestone and Scholl proposed that the same principle should hold for cases of apparent top down effects of perception, wherein the distortion effect on one's vision would apply not only to the target stimuli tested by an experiment, but also to the means of measuring that very perceptual distortion.

Consider first the forced choice experiment where participants were presented with two figures, one conventionally red and the other not, and were asked to indicate which one of them appears redder. Gross *et al*. take the variation of this setup, where the figures were actually of identical color, to be an especially damning case against cognitive penetrability – the test subjects were, after all, given all the incentive to exhibit their bias towards deeming one of the images as redder, the conventionally red figure being presumably the prime target. This assessment does not seem to make much sense. After all, getting the subjects' perceptual experiences and perceptual judgments mixed up is something that we were trying to avoid. This kind of a setup, however, clearly favors producing a biasing effect in the participants' perceptual judgments. Barring any differences in their actual perceptual experience, they were nonetheless incited to make an error in their judgment of it. Be that as it may, by extension, this also proved to be a case where cognitive penetrability did not seem to occur, unless we assume the participants' judgments to actually override their experience in this case. Given this possibility, this does not seem to be a good experiment for testing the cognitive penetrability hypothesis.

Next, consider the 'El Greco' experiment. An attempt was made to elicit a measurable difference between cases where participants had to match a figure to a uniform background and cases where the figure had to be matched to another, bigger figure of the same sort. If there was a genuine top down effect at play, one would assume that in the latter case, it would also apply to the background (which would be e.g. a larger heart-shape, presumably susceptible to the same kind of bias that the to-be-matched smaller heart shape would be susceptible to, if a top down effect is in fact present). Since the original study turned on the perceived contrast between the color of a shape and its background, it is difficult to ascertain what the 'El Greco' experiment was even supposed to illustrate. If the top down effect found by Dilk and Fillenbaum was very robust and straight forward, then perhaps it would make sense to assume that e.g. a reddish-orange heart shape, despite appearing redder than it is, would be still ill-fitted to the larger, redder heart shape which, due to its shape, would appear redder still. There are many things to be concerned with pertaining to this setup, like it discounting the size of perceived objects, discounting the role of perceptual attention and so on. However, no El Greco result was found for the much simpler reason that the original top down effect was not reproduced at all. Still, the point stands that it would be far too hasty to expect the result of the original study to be quite so simple. At best, it showed a possible top down effect from shape to color perception *in the* *context* of it being contrasted with a uniform background of a similar chromaticity. Why this is an important distinction becomes apparent from my analysis of the final experiment run by Gross and colleagues that most closely replicates the setup of the original study.

Last, let's turn to the 'replication' experiment most in line with the original Dilk and Fillenbaum study. Instead of having the participants look at the colored shapes and backgrounds through a wax paper, in a dimly lit room, they presented the task on a computer screen. One of the reasons for the original setup was to blur the edges between the shapes and the background. Macpherson (2012) considered this a feature of the experiment not pertinent to its results, proposing that the same could be achieved by replicating the test on modern computer screens. Taking this cue, Gross and colleagues did just that, yet the top down effect reported in the original study was not found. In the 'replication' experiment, they initially did not in fact introduce any artificial edges to the shapes, which reportedly made the task of disambiguating the shape and the background entirely trivial. In response to this, a dark edge was drawn around each shape to further incite the possibility of color perception bias. I would argue that both Macpherson and Gross *et al*. have failed to appreciate the importance of edge blurring present in the original study and that the original setup was in fact much more pertinent for investigating color perception.

Within the human visual system, different processes pertaining to different dimensions of the visual scene are known to operate side by side, often with the effect of accommodating for, or interfering, with each other. For example, within the Kanizsa's triangle illusion (Appendix 1), the presence of shapes with appropriately lined up cutouts produces the perception of illusory contours to accommodate for an apparent bright triangle that seems to occlude the other shapes. Since no bright white triangle is actually present, we can assume this to be the case of the visual system completing one of its jobs in a limited information setting, i.e. producing the percept of a surface that would readily explain the seemingly cut out parts of the present darker shapes. Importantly, edge detection and color perception are two separate processes that are nonetheless informed by each other. I propose that the reason why the 'replication' experiment failed to produce the results of Dilk and Fillenbaum could be for the fact that in deferring to blur the edges between the figures and the backgrounds, the 'replication' experiment, unlike the original, didn't test just color perception, but also edge detection, while it is not entirely clear how the latter might interfere with the former. Given that the presence of edges can in fact produce an illusory perception of color difference in actually identically colored areas of the visual scene (Cornsweet illusion), this seems a rather alarming oversight. In short, none of the experiments conducted by Gross *et al*. are satisfactory in their aim to recreate the study by Dilk and Fillenbaum, both conceptually and otherwise. Regardless, they highlight the ways in which the findings of the original study, if correct, should *not* be interpreted.

A further concern that could be raised against the Dilk and Fillenbaum study is that even if such a setup produces a genuine shift in a person's perceptual experience, this could also be explained in terms of a shift in the perceiver's attention. While not spatial attention, it could presumably be the case that a person, upon looking at a swathe of orange color, could in different circumstances attend either the red or the yellow content of the orange color (being that it is a mixture of the two). Macpherson (2012: 45) disregards this idea on the grounds that for this to be a viable argument against cognitive penetrability in this case, more work ought to be done to establish why this wouldn't nonetheless count as a case of cognitive penetration. This is for the reason that the concern raised by those who would consider attention-mediated cases of apparent penetrability to not count as such, is because allocation of attention is something that occurs before perceptual processing. In other words, one's cognitively mediated attentional focus determines that which is to be perceived, all the while the actual process of perceiving remains itself unaffected. According to Macpherson, this would not be the case, as attention to the redness within a swathe of orange could take place only once perceptual processing has already gone under way. This response isn't entirely convincing, considering that a parallel argument could also be made for shifts in spatial attention not mediated by one's eye movements. One can, for example, shift the focal point of one's visual attention to objects in one's periphery, which brings about a change in the visual experience that could also only take place once the perceptual process is already well under way. Yet, choosing to attend to an object in one's visual periphery doesn't seem to be a case of cognitive penetrability of perception in any satisfying sense. Regardless, the second study I will discuss here seems immune to such a challenge.

Another study that found a link between color perception and the conventional color of the perceived object is due to Olkkonen *et al*. (2008). In this study, a significant trend of judging images of fruit slightly in the opposite spectrum of their normal color as being fully achromatic (greyscale) was found. The thinking goes, that the participants' memory of normal fruit colors affects their perceptual experience of fruit depictions in a way that e.g. a perfectly greyscale banana still appears somewhat yellow, and it is only when the color of it is shifted slightly to the opposite spectrum that this memory effect is overcome.[5] The effect was diminished with less realistic paintings of fruit and eliminated entirely with presentations in the form of simple outlines of the fruit. Now, it seems reasonable that the more realistic the presentation of the fruit, the more difficult the task would be. A more realistic presentation implies that the visual system has more sensory data to work with, which puts a higher demand on the system as a whole. This also makes the system more susceptible to making "errors", as in order to make sense of a complex visual scene, even if depicted in a photograph, various higher-level features like depth, illumination and color constancy are attributed to the scene. In the context of normal viewing conditions within an actual three-dimensional environment, these are all incredibly useful features of our perception. Since our perceptual apparatus doesn't discriminate between actual environments and realistic depictions of them, however, various interesting visual illusions can be elicited in carefully manipulated images. Illusory or not, the question is whether a setup like the one created by Olkkonen *et al*. could actually bring about this kind of a phenomenal experience and, importantly, whether this could be facilitated by the perceivers' cognitive faculties.

The Olkkonen *et al*. study is a recent one, and rather straightforward, giving little reason to doubt its results. Especially so considering that the authors themselves didn't frame the results in terms of the controversial cognitive penetrability hypothesis. Given the scarcity of experimental results that unproblematically support the hypothesis, however, it is not surprising that those in support of it have tried to make sense of the study in just those terms. Despite the allure, here, too there is a reason to doubt that the results reflect genuine phenomenal differences in the participants' perceptual experiences, and could instead be the result of an error in judgment. However, as before, it makes sense to resist such a conclusion and determine instead the extent to which an actual shift in the participants' phenomenal experience could have been obtained. If the reader were now to take a look at a grayscale photograph of a banana, it is

---

[5] Interestingly, the effect was the most prominent in regards to the images of yellow bananas (on the yellow-blue spectrum), while being much less pronounced in the case of red strawberries (on the red-green spectrum). This could be for the reason that the human visual system is simply less sensitive to blue and violet light, accommodating for the fact that the perception of slight blue tint on the images of bananas shifted slightly to the opposite end of their normal color spectrum seemed more likely to go unnoticed.

unlikely that they would report the presence of a subtle yellow tint therein. This, however, does not discredit the veracity of the perceptual reports by the participants in this experiment, as this was not the task that they were presented with. Instead, they had to manipulate a colored picture of the fruit so that it would appear achromatic to them. The fact that the task started off with a colored image is, I hold, a crucial aspect of it. I will provide my reasoning for this contention in the last subsection of the paper. As far as the challenge based on attentional shifts is concerned, however, this does not seem to be available here. Given that there is no color-content present within a greyscale image and no yellow-content within a swathe of blue color, the participants couldn't have been shifting their attention towards that which is not there.

A lesson to be learned from both of the experiments and the recreation attempt of the first one is that the contextual factors surrounding the studied target phenomena are not trivial. Why the reported effects were obtained in the original experiments and are not easily reproduced in experimental setups that don't rightly regard this context can be shown to be accountable by the mechanism that I will propose underlies the cognitive penetrability of perception.

4 Perception, Cognition and Predictive Coding

In this section of the paper I will motivate an account of perception that I hold to be the key in making sense of phenomena that could reasonably called cases of cognitive states penetrating perception. This is the “predictive coding” view of perception, which posits that the process of perception is largely made up of a mechanism of error correction, wherein predictions about the perceptual scene (i.e. the most probable candidate models for what is causing the occurring sensory stimulation) are tested against the incoming sensory activity. When the mismatch between the two exceeds a certain threshold, an error signal is produced and propagated up to higher perceptual areas where a new model is created and then sent down the perceptual stream to repeat the process. While the idea of perception as a process that produces the best guess about the state of the world given the available data is not a new one, having been alluded to already by Von Helmholtz in 1860, it is only recently that we has seen considerable empirical verification of the claim as well as neuroscientific data on the possible neuronal mechanisms underlying this process of predictive error correction within the brain (see den Ouden *et al*. 2012 and Clark 2013 for excellent surveys of that).

4.1 Predictive coding of perception

The predictive coding view of perception frames it in terms of a multilevel hierarchical computational process between higher level top-down predictions and bottom-up lower level error signals. By such an account, the perceptual system does not encode the whole perceptual scene in its entire capacity for the higher brain areas to make sense of, but rather it propagates up the particular perceptual stream only the information that is surprising to the system as a whole. This means that the higher levels of any perceptual system propagate a predictive model of that which the agent might be perceiving down the perceptual stream and in so far as the model is accurate, no new sensory information needs to be propagated up the perceptual stream, it is only the error signals or a mismatch between a prediction and the actual stimulus that is sent further into the brain.

Given the unpredictable, stochastic nature of the outward environment and the computational limitations of our brains, certain heuristics that help in reducing the strain put on our nervous system all the while retaining the benefits of getting useful information pertaining to the environment are immensely useful. On the lowest level, it generally makes sense to assume that objects persist in time and extend in space. Thus, to take the example of vision, when a retinal cell reacts to light of a certain wavelength, it would be useful to assume that whatever is responsible for producing or reflecting that light falls into the receptive field of not just that retinal cell, but also of those nearby. Furthermore, it is also useful to assume that the very thing responsible for causing this retinal activation will remain there in the nearby future. Thus, a model of such a thing can be built, and it is only when something unexpected in the activity pattern of retinal cells occurs, e.g. when some cells sensitive to the same wavelength are not activated or when the very cells that were activated stop being so, that any new information needs to be propagated up the perceptual stream, so that a new model could be built. This gives us a very economically efficient way of signaling the presence and the spatial dimensions of perceived objects. (Clark 2013: 4)

On a higher level, it also makes sense to assume that perceived patterns keep repeating themselves. When hearing a repeated, rhythmic string of sounds, it is useful to assume that this very same repetition will keep playing itself out. It is only when something unexpected occurs, e.g. the pattern breaks or the sounds stop altogether that new information becomes necessary. Error signaling of such omissions have in fact been found in the brain areas relevant for a given kind of omission (see den Ouden *et al.* 2012), illustrating how the perceptual system is fundamentally sensitive to surprising change, and not just sensory stimulation. On a higher level still, it is also useful to assume that learned associations between certain stimuli will also hold in the future. A sight of lightning is likely to be followed by the sound of thunder. Strawberries are likely to give off a sweet scent. The sound of somebody yelling ‘donkey’ is likely to be followed by the sight of a four-legged animal with large ears. A yellow traffic light is nearly always followed by a red light (Panichello *et al*. 2013). A way in which to employ such predictions to one’s advantage in the process of perception is for the brain to perform a kind of Bayesian inference, where top-down predictive models of the world function as prior hypotheses and bottom-up error signals function as new evidence by which the statistical likelihood of the veracity of the prior hypotheses are re-evaluated. Alternative hypotheses, if more likely, will then take precedence over the original ones, and it is the most likely hypothesis given one’s priors and one’s evidence that will come to be one’s view of the perceptual world.

The hierarchical build-up of the perceptual system is of crucial importance here. In the case of vision, the complexity of the visual experience is created through the intricate interplay of neuronal coalitions specialized to encode particular aspects of the visual scene. Light-sensitive retinal cells signal, via their pattern of activation signal the raw physical light data hitting the eye to the lateral geniculate nucleus (LGN) where the color and luminosity content of the visual scene is encoded. From there, the primary visual cortex (V1) gets its input and encodes the presence of oriented edges and other very low level visual phenomena. This is picked up by higher areas of the visual system, wherein more complex phenomena like movement and combinations of simple features get encoded. A very high level feature of the perceptual process – object recognition – takes place largely in the inferior temporal lobe (IT) that houses neurons responsible for recognizing complex objects within the visual field.

Top-down feedback signaling happens between each of the visual areas. The top-down effects of predictive coding are thus sensitive to the function that each of the levels of the visual stream compute, eliciting error signals pertinent to the kind of phenomena that each layer encodes. The inverse of this is that *expected* stimuli, i.e. that coinciding with the predictive model, elicit less activity (compared to unexpected stimuli) in the regions of the brain that would normally encode them – a simple, expected shape in an expected area of the visual field is accompanied by less activity in the parts of V1 under which's receptive field the area falls to, an expected face elicits less activity in the higher, face sensitive regions of the visual cortex, and so on. (See den Ouden *et al*. 2012 and Panichello *et al*. 2013)

The way by which top-down predictive models of the environment could modulate the activity in lower parts of the perceptual stream is by the joint processes of excitation and inhibition, wherein a coherent model of the perceptual world, insofar as it is not met with considerable error signals, reinforces the activity of the neuronal coalitions that encode this type of a perceptual scene, all the while inhibiting the activity of those that do not (den Ouden *et al*. 2012). A good example of such a process would be the case of binocular rivalry. In a binocular rivalry setup, a test subject is shown a different image to each of their eyes. The resulting perceptual experience, however, turns out to not be a merged image of the two, but rather one or the other, switching at a semi-regular interval. In terms of predictive coding, this is the result of one or the other model of the world being overwhelmed by the error signals from the eye that is shown an image not corresponding to the current model (Clark 2013: 4-5). Furthermore, which of the images is more likely to dominate the subject's perception can be enforced by introducing images with vastly different content, wherein the visually more stimulating image will occupy the majority of the subject's experience. This effectively drowns out the error signals produced by the retinal neurons that are picking up on the less stimulating image.

However, it is not just parts of the visual cortex that signal feedback to lower areas – feedback that factors into the workings of the visual cortex originates also from other areas of the brain. In fact, insofar as we take perception to be a process by which a consciously felt percept is created, activity in the pre-frontal cortex, informed by that in the visual cortex, is also required. And just as with the layers within the visual system, top-down feedback is present also here, with activity in the pre-frontal cortex directly modulating the workings of the visual parts of the brain. In addition to that, signals from other perceptual modalities are also known to affect the activity patterns of visual neurons (den Ouden *et al*. 2012). Next, I will discuss some ways in which bottom-up error signals can be ignored by the perceptual system in case the predictive model of sensory stimuli is deemed statistically more likely by the system as a whole.

4.2 Cross-modal predictive signaling

Predictive error signaling hasn't been found in just the perceptual system, but also in the areas of the brain responsible for action, language, memory, cognitive control, and motivation (den Ouden *et al*. 2012). In terms of the cognitive penetrability of perception debate, the question to ask then, is whether this kind of predictive inference occurs also between areas of the brain that presumably encode states with propositional contents that would constitute cognition and areas that encode perceptual information.

Cross-modal effects between different perceptual systems are quite easy to produce. The McGurk effect discovered by McGurk and MacDonald (1976) is a clear case of one's auditory system being overridden by the workings of the visual system. They found that identical sounds of verbalized words can be perceived in distinctly different ways depending on the seen mouth movements of a person, wherein the perceived sound correlated with what is commonly associated with the given movement of the mouth. In the predictive coding paradigm, this is easily explained by the prior hypothesis that links certain sounds with certain movements of the mouth (through learned association) suppressing the bottom-up auditory signals in order to produce a percept of a more likely external stimulus. This is also not a matter of the visual system dominating the auditory system simply due to it being more complex. The inverse may hold as well, as in cases where a single flash of light accompanied by multiple auditory beeps seems to flicker (Shams *et al*. 2000) or where two disks moving towards each other and overlapping are perceived as bouncing off each other when a beep is heard at the time of the overlap, while seeming to pass through each other when the beep is absent (Sekuler *et al*. 1997). In both cases, the expected association between the sound and the dynamics of visually seen objects takes primacy over the raw visual data in order to produce a more likely percept. This goes to show that perceptual systems are not informationally closed off to sources outside of its particular perceptual modality, yet this does not establish that a source of such information could also be cognitive. In order to get closer to this mark, a look at how semantic categorization can affect the workings of perception is pertinent.

Consider, again, the binocular rivalry setup described in the last subsection. Interestingly, there is a way to lessen the dominance of the more stimulating image by prior semantic priming. Panichello *et al*. (2013) present a case wherein, if the subject's right eye is presented with highly stimulating dynamic noise and the left eye is presented with a simple low-contrast written word (e.g. 'salt'), it is the image shown to the right eye that dominates one's perceptual experience. By increasing the contrast of the written word shown to the left eye, this suppression effect can eventually be broken. However, if prior to the experiment, both of the subject's eyes are shown a word that is semantically related to the one subsequently presented to the left eye (e.g. 'pepper'), the dominance of the dynamic noise percept is broken much more quickly, requiring a lesser increase in the visual contrast of the written word. In a somewhat similar experiment, Lupyan and Ward (2013) showed that a masked stimulus that is otherwise invisible to test subjects can be made visible by simply cuing a semantically relevant word. So an otherwise suppressed image of a zebra could be perceived by participants once the word 'zebra' was heard. Importantly, the verbal cue 'zebra' did not cause participants to hallucinate zebras in test trials where no zebra image was actually present, and it caused a further suppression for images that were semantically unrelated, like those of pumpkins.

It seems then, that the predictive coding done by the perceptual system can be informed by semantic categories. By employing semantic categorization, statistically more likely predictive models of the outward environment can be created, which can cause the relevant kind of error signals to gain more stock in reshaping the perceptual experience, while suppressing those that have less traction. This would amount to top-down activation of the neurons that encode semantically associated visual properties. Lupyan (2015) takes cases like this to be clear examples of cognition penetrating the early workings of the perceptual system, namely, early vision. However, matters are not quite so clear. It is quite possible that the apparently semantically mediated perceptual activity could instead be triggered by the contents of a purely lexical module of the brain, without beliefs, knowledge, concepts or other cognitive faculties being involved at all (Macpherson 2015: 578). In other words, it could be that simply hearing the relevant word triggered the appropriate activity in the visual cortex via lexical priming, which doesn't necessarily mean that the word had to be even understood by the agent. This would easily rule out cognitive penetrability, as no mental faculty that could presumably qualify as cognitive would even be involved in the process of attaining the target perceptual experience. Since the cited experiments didn't test for this possibility (e.g. by testing whether just thinking of something semantically related to the target or perhaps seeing an image of it would produce a similar effect), there is good reason to be vary. In light of this, I will now turn to delineating exactly what would constitute cognitive penetrability of perception within the predictive coding framework and present a case wherein this is obtained. Embedded in this task is also the specification of the mechanism by which I propose this to take place. Having done so, I will briefly show how the same approach can be applied to the experimental findings of cognitive penetrability discussed in the previous section.

4.3 Cognitive penetrability of perception in predictive brains

What would then count as cognitive penetrability of perception within brains that partake in predictive coding and error signaling? There is a tension here between two kinds of ways in which our cognitive apparatus could affect the workings of our perception. On the one hand there is a sense in which we might consider actual cognitive states entertained at the moment of perception factoring into that which is to be perceived. This would be the most alarming case of cognitive penetrability, depending on the details of which, a portion of the work done by perception could be shown to be susceptible to direct manipulation of the perceiver on a cognitive level (for better or worse). On the other hand, there is the possibility that one's cognitive mechanisms could reshape the very way in which one's perception operates in a way that doesn't necessitate active cognitive work of the relevant sort at the time of perception, which could actually work *against* direct cognitive control.

4.3.1 Direct cognitive penetrability of perception

There are many cases wherein the first possibility for cognitive penetrability is known not to obtain. These would be the kinds of perceptual illusions that persist even if the perceiver knows them to be illusions. However, the presence of negative cases does not disprove the possibility of positives. In fact, Vishton *et al.* (2007) seem to have found just such a positive case in a variation of an experiment run by Aglioti and colleagues. When Aglioti *et al.* (1995) reproduced the Ebbinghaus illusion (Appendix 2) in a three-dimensional setting with physical disks, they found that there is a mismatch between the perceptual estimations of disk sizes (that were changed between trials) and the maximum grip aperture (as measured by the distance between thumb and forefinger) employed by the participants when asked to grasp one of the disks. Even though, on average, the disk surrounded by smaller disks appears around 10% larger than an identically sized disk surrounded by larger disks (measured by the size-difference of the target disks in a variation of the experimental setup wherein the targets of actually different size were made to appear to be of the same size), the increase in maximum grip aperture exhibited a magnitude of only around 6%. This is accounted for by the separate visual pathways at work across different tasks, the ventral pathway for visual awareness and the dorsal pathway for visually guided grasping, wherein the effect of the illusion varies from one to the other.

The dual pathway model of visual processing posits the existence of two separate pathways by which visual tasks are completed within the brain. The dorsal, vision-for-action pathway is suited for visuo-motor tasks that involve orienting one's body within the visually represented outward environment, whereas the ventral, vision-for-perception pathway is more suited for identifying visual objects. This is a smart computational strategy, as the neuronal processes necessary for e.g., reaching out and grabbing a tool work within a frame of reference distinct from and susceptible to different invariances from processes that are required to identify the object as e.g., a hammer, and in fact, it has been found that the different pathways can actually interfere with each other's operations (Koch 2004: 212). The ventral pathway employs mechanisms that render object recognition invariant to distance, lightning conditions and so on, all in the service of perceiving given objects in similar ways across different viewing conditions. This also renders the workings of the vision-for-perception pathway to be slower than that of the dorsal, vision-for-action pathway utilized to get an accurate estimation of the proximity of external objects to one's physical body. This is partly also accounted for by the fact that the visual processing in the dorsal pathway operates at a non-conscious level and with a larger receptive field. Hence, the otherwise highly accurate task of estimating the distance of an ambiguous object in one's visual periphery can be undermined once the object is brought into one's focal point, upon which the activation of the ventral pathway that seeks to identify the object and the consciously mediated phenomenal experience accompanying it can cause the distance-estimation of the object to become much less accurate. The Aglioti *et al*. experiment with the three dimensional setup of the Ebbinghaus illusion presents a clear case where the effect is inverse – with operations of the dorsal pathway informing those of the ventral pathway in a way that causes the vision-for-perception pathway to overcome some features of its normal operation that are accountable for the presence of this illusion.

What Vishton *et al*. (2007) found, in a variation of the three dimensional reproduction of the Ebbinghaus illusion, that the very act of reaching for a disk significantly lessened the effect of the illusion on subsequent perceptual estimation of disk sizes for several minutes after the reaching trials, diminishing the illusion of one of the disks appearing larger upon perceptual estimation to a mere 5.74%. More importantly though, the effect of the illusion in perceptual estimation trials was diminished significantly also when the participants were merely informed that they would be subsequently required to grasp the disk that appeared larger. This makes sense, given that the activity patterns of the viso-motor neurons in the dorsal visual pathway are highly similar regardless of whether a given viso-motor activity is actually performed or merely imagined. Briscoe (2014) proposes that the way in which the workings of the dorsal pathway could inform the workings of the ventral pathway in this case could be by the evoking of a high level intention to act, by evoking beliefs concerning the action that the participant is required to perform, or by evoking relevant motor imagery. All those could exert influence on the perceptual estimation by changing the activity of the visual system that account for depth information such as disparity, vergence, accommodation, and relative size. Regardless of which of these possibilities is more likely, cognitive penetrability of perception would be obtained by all of them. I will propose a more general account of how this process could unfold within the framework of predictive coding, which, not being mutually exclusive from any of the aforementioned accounts, will, I believe, present a satisfying reasoning for this effect and establish it as a case of cognition penetrating perception.

The Ebbinghaus illusion, like many of such persistent perceptual illusions, makes a lot of sense within the framework of predictive coding. Predictive models, being informed by either hardcoded features of perception or by learned associations, encode on a neuronal level the most likely candidates for what might be causing a given sensory stimulation, as determined by some form of Bayesian inference. Why such illusions persist despite conflicting knowledge of their illusory nature, is simply that the illusionary percept is more likely given what is present in the perceptual scene. The Müller-Lyer illusion wherein identical lines are seen to have a different length depending on whether they are adorned by inward or outward pointing arrows at their ends, is probably due to the fact that in normal situations (in the external world, outside of the laboratory), lines adorned with arrow-like figures pointed outwards present convex surfaces that are closer to the observer, and the other way around for lines adorned with inward pointing arrows (Lupyan 2015: 558). When two objects cover the same visual angle, with one being nearer and the other being farther, it is the one that is nearer to the observer that must be objectively smaller. The visual system makes use of this inference and produces a model where one of the lines does indeed appear smaller. Given that this holds true in most viewing conditions, such a percept is a statistically highly likely model of the environment and any low-level error signals accounting for how much of the visual angle each of the drawn lines subtends are suppressed by this likely predictive model. In the case of the Ebbinghaus illusion, this effect could be explained by the fact that larger objects surrounding the target object create a scene wherein the target will appear to be closer to the observer and will, by the same logic, make it appear smaller than the identical object that nonetheless appears to be further away.

The effects of these features of the ventral visual pathway, however, are not shared to the same extent by the dorsal visual pathway. Given that the latter is much more suitable for estimating the actual size of seen objects, the illusion manifests at a lesser degree. I propose that the signals from the dorsal pathway to the ventral one that conflict with the perceptual estimation of disk sizes done by the ventral system reinforce the bottom-up error signals that go against the previously most statistically likely predictive model of the visual stimuli. This necessitates the production of a new, more accurate model, given all relevant evidence, which will become the new visual percept experienced by the perceiver, which is an experience distinct from the one that was had before. Given that the activity patterns of the neurons in the dorsal pathway are largely the same regardless of whether the motor task was carried out or not, this effect can be produced by merely intending, imagining, or believing that this kind of a motor task is to be carried out. All of these are cognitive processes that, by evoking activity in the dorsal visual pathway that modulates the activity of the ventral visual pathway and the perceptual experience attained therein, effectively penetrate perception. What is more, this is obtained without changes in the stimuli, the viewing conditions, or changes in visual attention and it is also attained in a way that does not cause the perceptual system to malfunction.

4.3.2 Indirect cognitive penetrability of perception

In this last subsection, I will turn back to the experiments discussed in section 3, and show them to be similarly accountable by the predictive coding framework of perception. Both of these cases, if proper cases of cognitive penetrability, obtain their effects in an indirect way. That is, it is not a cognitive state entertained at the moment of perception that is accountable for the change in the perceptual experience, but rather cognitively mediated bias that is learned over an extended period of time.

In the Dilk and Fillenbaum study with cardboard cutouts, it was found that participants are prone to matching the ambiguously orange-red cutouts of conventionally red objects to redder backgrounds than not conventionally red objects. A replication attempt of the experiment by Gross *et al*. highlighted the importance of blurred edges between the target and the background and showed the effect to be highly dependent on the presented task, showing that people do *not* tend to simply see conventionally red objects as more red than they are. Turning to the predictive coding framework, we can say that even if there is a high level prediction about conventionally red shapes appearing more red than conventionally non-red objects, this is easily overcome by bottom-up error signals from low-level visual neurons sensitive to color. However, under difficult viewing conditions and given the complicated task of matching an ambiguously colored object to a background that doesn’t actually match, this high-level prediction could take more precedence. Since it is statistically more likely that some conventionally red object is closer in its chromaticity to the slightly-redder background than to the slightly-more-orange one, the error signals about a mismatch between the slightly-redder background and the object could simply have less traction within the perceptual system when compared to the error signals elicited by a mismatch with the slightly-more-orange background.

In the Olkkonen *et al*. study, it was discovered that if people were asked to manipulate photos of fruits so that they would appear perfectly greyscale, they tended to overshoot to the opposite end of the spectrum; the implication being that the perfectly greyscale image of the fruit still appeared to have a subtle tint of its former color. This, again, is not reproducible by simply looking at greyscale photographs of fruit. Again, a possible high-level prediction of a banana being yellow, if present, seems to be overruled by low-level bottom-up error signals to the contrary. However, the same might not hold for visual tasks wherein a yellow banana is actually already perceived and upon careful manipulation loses its normal color. A possible high-level predictive model about bananas being yellow would then initially be reinforced by the lack of error signaling to the contrary. Subtle changes in the actual chromaticity and the error signals pertaining to it could then be drowned out by the reinforced model. It could well be that this kind of subtle manipulation, in unison with the low-level predictions about objects retaining their appearance throughout time, could account for the fact that a perfectly greyscale image of a banana attained by this kind of a process could not produce the error signals necessary to reach the threshold wherein the working model would be discarded.

In both of these cases, categorical knowledge of objects and the features pertaining to them seem to have some stock on the perceptual experience of perceiving agents in specific circumstances. This would amount to cognitive penetrability of perception, albeit in a manner not quite so direct as with the case of the three dimensional Ebbinghaus illusion. Yet, if predictive error signaling is one of the main mechanisms employed in brain functioning, as has been argued, cases like this highlight the general means by which cognitively mediated high-level predictions about the outward environment could, in favorable circumstances, penetrate the phenomenal experience of perceptual states.

4.4 Replies to possible objections

Finally, I will consider a few objections that could be raised against my treatment of the cognitive penetrability debate.

First, I should consider the possibility of problematic attentional shifts that would trivialize the shift in perceptual experience pertaining to the modification of the Ebbinghaus illusion. There is no reason to think that the change in the reported perceptual experience is attributable to changing patterns of eye-movements, as the task of assessing which of the target circles is bigger remains the same, regardless of whether one is asked to subsequently grab the larger one or not. A more feasible challenge would then be to propose some sort of non-spatial attentional shift. Since the change in perceptual experience turns on the perception of size, there should be some feature of the image that allows for participants to focus on something that would allow for them to assess the targets as being of identical size, something that we could perhaps call the sameness-content of the image. However, nothing of the sort seems to be present. As before, the means by which the participants are to assess whether one of the circles is bigger don't change between the cases wherein they are asked to subsequently grab the larger of them and the ones where no such instruction is given. Furthermore, it is not apparent how any sort of attentional shift could account for the finding that after performing the visuo-motor task, the participants' visual estimation persisted for only several minutes. Surely, if they found a way by which to focus their attention for better results, they wouldn't forget it in such a short time.

A second concern could be that the Ebbinghaus case presents a very specific kind of a setup wherein the modulation of one's perception by cognition is mediated by the dorsal visual pathway, rendering it a special case that doesn't establish cognitive penetrability in different cases. While this is essentially true and insofar as I am interested in establishing cognitive penetrability as such and not just presenting one strange case where this can happen, it is a serious concern. However, I believe there are good reasons to resist this. For one, recent studies have shown how the ventral and dorsal visual streams are, in their normal functioning, much more interconnected than such a concern would imply. For example, while it is true that the locations of objects are encoded largely by the dorsal pathway, both the ventral *and* the dorsal pathway have in fact been found to routinely contribute to shape perception (Zachariou *et al*. 2014). Despite that, I would resist the conclusion that it is only via the special kind of interplay between both pathways that cognitive penetrability of perception might occur. Given that there is ample evidence for the predictive coding model of the brain working in most of its mechanisms and the fact that it neatly explains a lot of perceptual phenomena, including purported cases of cognitive penetrability, I would cautiously propose that this kind of a computational mechanism is utilized in a more widespread manner. This is especially so in the plausible case that cognitive penetrability is a feature of perception that, instead of placing the perceiver in an epistemological abyss, provides them with an efficient way of gleaning pertinent information from the apparent environment. Something that would presumably contribute to the biological fitness of a perceiving organism.

Related to the last concern, one might also be unsatisfied with my treatment of the two experiments from color perception. The concern here is that my conceptualization of what is going on in these cases using the model that I presented could be simply question begging. However, I don't believe my analyses of the experiments were unfair to their opponents, and even then, my goal was mainly to outline *how* cognition could be penetrating perception in these cases, assuming that this is in fact a case. I then showed how the features of these experiments could be easily explained by my proposed mechanism for cognitive penetrability as, more than anything, an illustration. While this certainly does not constitute a knockdown argument in favor of framing these experiments as cases of cognitive penetrability of perception, no aspect of my thesis claim rests on this.

5 Conclusion

In this paper I provided a defense for the hypothesis that one’s perception can be cognitively penetrable. By attempting to bring more clarity into this argument, I argued that while the conceptual details pertaining to it are best lain out upon philosophical reflection, the testing of cognitive penetrability claims as well as outlining a possible mechanism for it is a project that has to engage with cognitive science.

I showed that the argument ought to be unpacked in terms of discerning whether cognitive states, understood as mental states with propositional content, can directly modulate the phenomenal content of one’s perceptual experience in a way that everything else pertaining to the perceptual situation remains fixed. Furthermore, perception, upon being penetrated by cognition, should retain its function so as not to simply malfunction, for it is likely that if perception is indeed cognitively penetrable, this could very well be a function of perception immensely useful to an organism that has to navigate their immediate environment.

I also argued that the phenomenal content of the perceptual experience in question should be that of low-level perceptual phenomena, as this would be the sure way to appreciate the philosophically interesting question that arise from this debate.

Upon outlining the proper relation between perception and cognition that should hold for the former to be penetrated by the latter, I turned to analyzing two purported cases of cognitive penetrability from the field of color perception. By engaging with the counterarguments mounted against these findings, I outlined the way in which perception could be said to be cognitively penetrated in these cases.

Finally, I presented and motivated the predictive coding framework of perception according to which perception is a largely top-down process of testing prior hypotheses or models of one’s perceptual world against bottom-up error signals of stimuli not accommodated by the given model. Via this probabilistic inference, certain stimuli can be inhibited and others enforced. I then considered, in light of this, more purported cases of cognitive penetrability. I showed how the approach taken by Lupyan of attempting to establish cognitive penetrability in terms of categorical knowledge mediated by predictive coding does not unproblematically establish the hypothesis. I presented instead another case that, utilizing the dual pathway model of vision, seems to establish the penetrability of one’s perceptual experience by propositional attitudes. I then showed how this is accounted for by the predictive coding model of perception, wherein a mental state with propositional content can be shown to directly affect the probabilistic distribution of one model of the world taking primacy over others. I then employed the same strategy for proposing the potential ways in which the aforementioned experiments on color perception could in fact constitute cognitive penetrability, albeit in a less direct way.

Abstract

The debate over whether perception is cognitively penetrable, in order to remain philosophically interesting and relevant, should be unpacked in terms of discerning whether propositional states can directly modulate the low-level phenomenal features of one's perceptual experience. For this, it should also engage with cognitive science, and it is by proposing a scientifically feasible mechanism of how cognitive penetrability could obtain that there is even a reason for considering it. The predictive coding model of perception provides a framework within which such a mechanism can be motivated. By framing perception as a process of probabilistic inference, wherein top-down models of the world are tested against bottom-up error signals, we can discern whether propositional attitudes could affect the inhibitory and excitatory signaling of sensory neurons in a way that either promotes or inhibits the top-down predictive model that constitutes one's perceptual experience, all the while preserving the necessary kind of relation between perception and cognition. In a three dimensional variation of the Ebbinghaus illusion, this can in fact be shown to happen, establishing cognitive penetrability of perception and a working mechanism by which this could take place.

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs, *The Journal of Neuroscience*, **26**(1): 63-72.

Aglioti, S., DeSouza, J. F., and Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand, *Current Biology*, **5**(6): 679–685.

Agnati, L. F., Guidolin, D., Battistin, L., Pagnoni, G., and Fuxe, K. (2013). The Neurobiology of Imagination: Possible Role of Interaction-Dominant Dynamics and Default Mode Network, *Frontiers in Psychology*, **4**: 296. DOI 10.3389/fpsyg.2013.00296

Bayne, T., and Montague, M. (2011). Cognitive Phenomenology: An Introduction, *in* T. Bayne and M. Motangue (eds.), *Cognitive Phenomenology*. New York and Oxford, Oxford University Press, pp. 1-34.

Briscoe, R. E. (2014). Do intentions for action penetrate visual experience, *Frontiers in Psychology*, **5**:1265. DOI 10.3389/fpsyg.2014.01265

Briscoe, R. E. (2015). Cognitive Penetration and the Reach of Phenomenal Content, *in* J. Zeimbekis and A. Raftopoulos (eds.) *The Cognitive Penetrability of Perception: New Philosophical Perspectives*. Oxford, Oxford University Press, pp. 174-199.

Carruthers, P. and Veillet, B. (2011). The Case Against Cognitive Phenomenology, *in* T. Bayne and M. Motangue (eds.) *Cognitive Phenomenology*. New York and Oxford. Oxford University Press, pp. 35-55.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Sciences*, **36**(3): 181-204.

Delk, J. L., and Fillenbaum, S. (1965). Differences in Perceived Color as a Function of Characteristic Color, *The American Journal of Psychology*, **78**(2): 290-293.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, **73**(3): 415-434.

Firestone, C., and Scholl, B. J. (2014). Can you *experience* ‘top-down’ effects on perception?: The case of race categories and perceived lightness, *Psychonomic Bulletin and Review*, **22**: 694-700.

Firestone, C., and Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for ‘top-down’ effects, *Behavioral and Brain Sciences*. DOI 10.1017/S0140525X15000965

Gross, S., Chaisilprungraung, T., Kaplan, E., Memendez, J. A., and Flombaum, J. (2014). Problems for the Purported Cognitive Penetration of Perceptual Color Experience and Macpherson’s Proposed Mechanism, *The Baltic International Yearbook of Cognition, Logic and Communication*, **9**:1-30. DOI 10.4148/biyclc.v9i0.1085

Helmholtz, H. von. (1860/1962) *Handbuch der physiologischen optik, vol. 3*, J. P. C. Southall (ed. & trans.), Dover. (Original work published in 1860; Dover English edition in 1962).

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., and Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas, *Neuropsychologia*, **40**(10): 1706–1714.

Kayser, C. and Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure and Function*, **212**(2): 121–132.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*, Denver, Roberts & Company Publishers.

Kowacs, P. A., Utiumi, M. A., and Piovesan, E. J. (2015). The Visual System in Migraine: From the Bench Side to the Office, *Headache: The Journal of Head and Face Pain*, **55**: 84-98.

Levine, J. (1983). Materialism and qualia: the explanatory gap, *Pacific Philosophy Quarterly*, **64**: 354-361.

Lupyan, G., and Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness, *Proceedings of the National Academy of Sciences*, **110**(35): 14196–14201.

Lupyan, G. (2015). Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems, *Review of Philosophy and Psychology*, **6**(4): 547-569.

Machery, E. (2015). “Cognitive Penetrability: A No-Progress Report” in Zeimbekis and Raftopoulos (eds.) *The Cognitive Penetrability of Perception: New Philosophical Perspectives*. Oxford: Oxford University Press, pp. 57-74.

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.

Macpherson, F. (2015). Cognitive Penetration and Predictive Coding: A Commentary on Lupyan, *Review of Philosophy and Psychology*, **6**(4): 571-584.

Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications, *American Psychologist*, **17**: 776-783.

Olkkonen, M., Hansen, T., and Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes, *Journal of Vision*, **8**(5): 1-16.

Vishton, P. M., Stephens, N. J., Nelson, L. A., Morra, S. E., Brunick, K. L., and Stevens, J. A. (2007). Planning to reach for an object changes how the reacher perceives it, *Psychological Science*, **18**: 713–719.

Den Ouden, H. E. M., Kok, P., and de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation, *Frontiers in Psychology*, **3**:548. DOI 10.3389/fpsyg.2012.00548

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

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

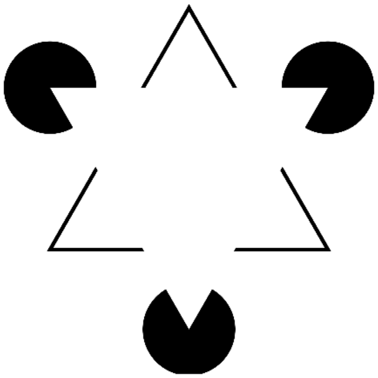
Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, **9**: 255–266.

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.

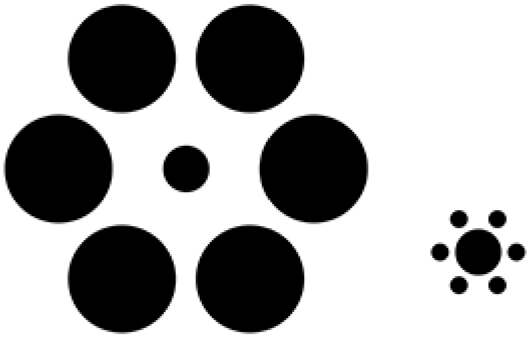
Zachariou, V., Klatzky, R., and Behrmann, M. (2014). Ventral and dorsal visual stream contributions to the perception of object shape and object location, *Journal of Cognitive Neuroscience*, **26**(1): 189-209.

Appendices



Appendix 1. Kanizsa's triangle

The lined up cutouts produce an illusory percept of a bright triangle.



Appendix 2. Ebbinghaus illusion

The central circle on the left appears smaller than the one on the right whereas they are actually of the same size.

Non-exclusive licence to reproduce thesis and make thesis public

I, Sander Vaus,

(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

Cognitive Penetrability of Perception in Predictive Brains,

(title of thesis)

supervised by Professor Dr. Daniel Cohnitz,

(supervisor's name)

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 02.05.2016