Empirical Perspectives on the Cognitive Penetrability of Perception

Piotr Litwin

University of Warsaw,
Faculty of Psychology
piotr.litwin@psych.uw.edu.pl

Abstract

The problem of the cognitive penetrability of perception pertains to whether perceptual processing may be impacted by higher-order cognitive processes. It may be understood in a twofold sense: 1) whether what a perceptual system computes may be altered in a way that is semantically coherent to one's cognitive states; 2) whether perceptual experience may be influenced by cognitive processes. It has been argued that the cognitive penetrability problem is not scientifically tractable since we have no direct access to other persons' subjective experiences and, therefore, we have to rely on their reports, which are mediated by higher-order processes (e.g., judgments). In this paper, I analyze the scope of methods harnessed in contemporary work on the penetrability of perception, focusing particularly on methods from experimental psychology and neuroscience. Among them, I indicate the most promising techniques and paradigms, as well as those that are inadequate to successfully tackling the problem. I also discuss experimental results which unequivocally suggest direct influences on perception which have not been addressed to date by supporters of the view that perception is impenetrable. I also describe the predictive processing theory of cognition, focusing on how it contributes to our understanding of cognitive penetrability, and discuss scientific results validating the theory. In conclusion, converging empirical evidence seems to suggest that perception is cognitively penetrable and we seem to be at the dawn of the ultimate solution of the problem.

Keywords: cognitive penetrability of perception; early vision; visual matching; performance-based measures; physiological reflexes; predictive processing.

1. Introduction: What Is the Cognitive Penetrability of Perception?

Cognitive penetrability is a term that covers two related but distinct problems in the philosophy of perception (Macpherson, 2016). The classic, "cognitive" interpretation pertains to the problem of "early vision" (Pylyshyn, 1999)—a hypothetical encapsulated part of the visual system that produces basic visual representations. According to the orthodox standpoint, the output of early vision is believed to be determined solely by the physical properties of an attentionally gated sensory signal and rigid principles governing the processing of information. Higher-order processes, like memory, judgment or knowledge take these basic visual representations as an input but cannot influence processing itself; this means that early vision is cognitively impenetrable. If cognitive penetrability occurs, then the effect of computations in the early visual part should be "sensitive, in a semantically coherent way, to the organism's goals and beliefs, . . . altered in a way that bears some logical relation to what the person knows" (Pylyshyn, 1999, p. 343). The cognitive penetration problem in this sense has been narrowed to visual perception (possibly due to the fact that a substantial part of the psychological work on perception involves vision), but there are good reasons to extend it to other sensory modalities (Lupyan, 2015), e.g., audition (Brogaard & Gatzia, 2015) or even simple multisensory experiences since "highly circumscribed modes of interaction" (Pylyshyn, 1999, p. 364) may be wired into the perceptual module that governs the well-examined crossmodal effects (Firestone & Scholl, 2016).

The "phenomenological" interpretation of the cognitive penetrability thesis relates to the problem of higher-order influences on the content of the conscious percept—whether what we perceive can be altered in a semantically coherent way to one's cognitive or affective states (Macpherson, 2016). Siegel (2012) provides a precise definition that indicates that perception is cognitively penetrable if two subjects (or one subject in two different moments) can have different visual (perceptual) experiences while attending to the same stimuli in the same external conditions (and having equally efficient sensory organs; Macpherson, 2012) as a result of differences in cognitive or affective states. Macpherson (2016) also stresses the importance of the presence of causal semantic links between each of the steps between a belief (or another higher-order state) and a corresponding perceptual experience. For example, a migraine causing flashing lights in the visual experience of the sky as a consequence of the belief that aliens are attacking Earth does not count as a case of cognitive penetration, since only the causal link persists between belief and migraine, and the semantic link is lacking.

The problem is crucial from the situated cognition perspective. It is believed that the penetrability of perception ("theory-laden observation"; Fodor, 1983) should threaten the independence of observation and make us feel epistemologically insecure, since no correspondence relation between perception and the world remains. Modularity of perception is believed to be a necessary instrument to retain this link. Others suggest both that it would be non-adaptive for organisms to waste their mental resources representing meaningless objects, and that cognitive sculpting of perception is ecologically as well as epistemologically desirable (Goldstone et al., 2015). The incoming sensory signal is inherently

ambiguous and uncertain, so if higher-order processes—not only expectations or knowledge, but also learned emotional associations, affective states, interoceptive information or motivations—could enhance perceptual processing, then they should influence it (Lupyan, 2015). Empirically speaking though, it is not clear whether the problem still persists. Some claim that the problem is not yet solved due to the strict methodological requirements it demands, but it could be in the future (Firestone & Scholl, 2016a); others seem to even doubt that it is solvable and suggest that contemporary research encounters pretty much the same obstacles as the New Look movement (Machery, 2015). On the other hand, some researchers claim that the convergence of evidence from behavioral science, neuroscience and social cognition allows us to take cognitive penetration as a fact (Vetter & Newen, 2014; O'Callaghan et al., 2017). The problem may be additionally obscured by the constantly accelerating mass of sometimes average quality scientific reports—direct influences on perception are frequently implausible or even incapable of accounting for the reported effects (Firestone & Scholl, 2014, 2016a).

In this article, I aim to address the issue of whether the cognitive penetrability of perception is a scientifically tractable problem. Firstly, I try to clarify the problem by addressing the subtle interplay between attention and perception, arguing that some attentionally mediated modulations of early vision activity may reflect genuine cognitive influences on perception. Therefore, even though one should definitely control the orientational subsystem of attention (the locus of attentional focus) while running experiments on the penetrability of perception, the "no attentional effects" condition should perhaps be revised. Subsequently, I summarize the empirical state of the art, focusing on evidence from experimental psychology and neuroscience. I critically evaluate the methods and paradigms used in contemporary psychological science; essentially, I am interested in which methods provide data that allow credible statements about genuine effects on perception. As a result, I indicate the most promising experimental paradigms as well as those inadequate to tackling the problem. I conclude by briefly outlining the predictive processing theory of cognition, focusing on the specific sense of cognitive penetrability it provides and describing experimental paradigms to test precise PP predictions.

2. Fuzzy Borders between Attention and Perception

Before we proceed to the experimental methods, there is a foggy territory that should perhaps be cleared: the delicate relation linking perception and attention, which is of the greatest importance in the context of the problem. According to the traditional view, perception and attention should be sharply distinguished. Pylyshyn (1999) underlines that "allocation of attention to certain locations or certain properties [is executed] prior to the operation of early vision." Attention is directed to a given part of a perceptual scene by selecting the input for processing but does not influence processing itself. Although it is certainly cognitively guided (and what we attend to certainly determines what we perceive), counting such attentional modulations as cases of cognitive penetrability trivializes

the problem; this would lead to the absurd conclusion that—as cognitively guided perceptual selection processes—intentional closure of one's eyes, or intentional eye movements should also be counted as cases of penetration (Lupyan, 2015). Accordingly, the "no differences in attention" condition (empirically speaking, no differences in endogenous attention between participants, which means that they should focus on the same spot in a visual field) is frequently indicated as a precondition for cognitive penetrability (e.g., Siegel, 2012; Toribio, 2015).

Defenders of the orthodox view of perception claim that it is the information selection process ("gating") that is often biased by cognition. For example, attentional shifts between vertexes of the Necker cube lead to shifts between percepts (Toppino, 2003); attention may also play an analogous role in other ambiguous figures (Long & Toppino, 2004). Therefore, studies claiming that perception is penetrable on the basis of the preference for one interpretation of an ambiguous figure are susceptible (e.g., Balcetis & Dunning, 2006) as it is the locus of attention that is biased, rather than perception. As a result, distinct sensory material is provided which ends up being processed in a particular way. It has also been suggested (Firestone & Scholl, 2016) that post-attentional effects of increased perceptual sensitivity (when objects are perceived "as if" they are brighter or of higher contrast after the sensory space in which they occur has been focused on; e.g., Carrasco et al., 2004; Ling & Carrasco, 2006) are caused by the larger amounts of perceptual information processed. In this case, attention may be thought of as a miner who dug out more perceptual resources in a given part of a perceptual field, or as a "spotlight" that may be "directed" to a given part of the perceptual landscape, providing more or less information for perceptual processing (Lupyan, 2015).

Nevertheless, attention seems to be a sophisticated process and some of its manifestations may escape common-sense "selection" or "spotlight" theories. Firestone & Scholl (2016a, p. 36) admit that "attention may interact in rich and nuanced ways with unconscious visual representations to effectively mold and choose a "winning" percept—changing the content of perception rather than merely influencing what we focus on." We can point to the biased-competition theory (Desimone & Duncan, 1995) as a broader alternative theory that may explain a wide variety of such attentional influences (endogenous vs exogenous, covert vs overt, conscious vs unconscious etc.; Marchi, 2017). The theory starts with the familiar notion that cognitive systems struggle to cope with an overabundance of information and must allocate resources to the processing of the most salient objects. However, information is not selected with the use of an internal filter (operating in a specified locus) or a spotlight-like gating mechanism. Particular objects and features compete for neural representations in perceptual subsystems and attention is understood as a *result* of these conflict resolution/selection processes (Mole, 2015)—the object which is selected for processing is the one to which attention is "deployed." The outcome of the selection processes in a particular subsystem is then spread throughout the sensorimotor network and may be used as an input (be it top-down, bottom-up or lateral), thus biasing competition processes in other subsystems. It is important to note that, in this view, attention is de-reified and is not to be taken as an "internal eye" which is governed by a homunculus-like entity so as to pick up the

proper material for processing (see Anderson, 2011 for a powerful critique of the reification of attention in contemporary cognitive science); here, attention is simply a result of the selection processes. For example, our perceptual systems evolved in such a way that some stimuli are more perceptually salient, which means that they have properties that tend to skew perceptual processing in their favor (e.g., red color, high luminance). As a result, they "grasp" (exogenous) attention (winning competition processes in lower-level perceptual subsystems) and influence competition processes at higher levels of the cognitive hierarchy, clearing the way to consciousness for a given object.

However, competition is not constrained to the perceptual system. Higher-order processes (e.g., knowledge about visual search task demands, expectations or motivation) may influence perceptual processing indirectly via resolution of conflicts between stored representations of objects held in working memory. The competition is biased towards objects important in the context of the task and this outcome is projected to lower-level perceptual subsystems to enhance processing of particular features that these objects tend to have. For example, attention is well known to bias perception in the rivalry condition in favor of the attended stimuli, prolonging their dominance (van Ee et al., 2006). Attending to a specific feature (different orientation of overlapping gratings) biases neural activity in the V1 cortex towards this feature (Kamitani & Tong, 2005; for a review of feature-based attention effects on perceptual processing in early visual areas see Carrasco, 2011) and reverses the neural response attenuation associated with repeated presentation of a stimulus as early as in V1 (Kok et al., 2012b). Attending to a given category of an object may enhance processing of corresponding perceptual features, e.g., vertical lines among horizontal ones when attending to people (Oliva & Torralba, 2006) and may also warp global neural representations of other objects belonging to other irrelevant categories (Çukur et al., 2013). The studies performed by Ling and Carrasco (2006) and Carrasco et al. (2004) may also be reinterpreted in terms of biased competition theory. In endogenous covert attention (intentional attention towards a specific feature or point in a sensory space with gaze fixed on a different point) there is simply a top-down bias towards objects in a given part of a visual field. Therefore, less perceptual evidence is required to form a conscious percept, since weaker activations may suffice to win the competition as long as particular properties are promoted in selection processes (e.g., activations in neurons with particular receptive fields, receiving input from a given part of a sensory space).

Distinguishing "clearly attentional" effects from "clearly perceptual" ones may be challenging. The problem of attention has to be considered cautiously as attentionally-mediated top-down influences may possibly modulate processing of particular visual representations in a perceptual subsystem as early as in V1. It may be unwise to dismiss such effects as "merely attentional." It seems that if we accept biased competition theory, we should acknowledge such effects as genuine cases of cognitive penetration since the content of perception and neural operations within the bounds of early vision may be influenced in a semantically-coherent way with cognition (Mole, 2015). Therefore, during experimental examinations of cognitive influences on perception, we should instead maintain equal locus between subjects, e.g., by direct instructions pertaining to what to focus on, or by imposing

additional loads to prevent attention from being deployed away from the distracting task (Firestone & Scholl, 2016a). In biased-competition theory terms, this would mean that we should maximally constrain the external sources of bias. Note that this means that one should equalize between-subject differences in the orientational subsystem of attention (to keep sensory material provided for perceptual processing still), but the remaining subsystems—alerting (achieving and maintaining a state of high sensitivity to upcoming stimuli) and control (responsible for processing incongruent stimuli or conflict and maintaining behavioral control; Posner, 2008; Yin et al., 2012)—are left unaddressed. Moreover, higher-order processes (e.g., motivation) are certainly involved in how these subsystems work. It seems though that the former may be intricately interconnected with perception (how "maintaining high sensitivity to upcoming stimuli" differs from preferential processing is unclear) and the latter pertains more to the execution of response.

3. Empirical Ways to Deal with "Early Vision" on the Neuronal Level

Since some light has been shed on the confusing perception–attention distinction, we may proceed to the penetrability of perception problem. Neither of the two interpretations of the problem ("cognitive" or "phenomenological") is easy to deal with. From a neuroscientific perspective, early vision was originally defined functionally (Cecchi, 2014) as a part of the visual system that computes basic visual representations that could be described using geometry terms and that does not require access to memory to perform its tasks (Pylyshyn, 2003). These representations of basic visual properties include surface, color, texture, orientation or motion representations (Raftopoulos, 2014). The original neuroanatomical definition provided by Pylyshyn (1999) seems though to be too sparse, including parietal areas or the inferior temporal cortex (Cecchi, 2014), and areas engaged in multisensory integration (Pasalar et al., 2010) and in comparing incoming visual input with representations held in working and short-term memory (Miller et al., 1991). Therefore, neuroanatomically, the demarcation line between "early" and "late" vision is not yet clearly established. Most researchers seem to endorse the temporal conception, regarding visual processing that encompasses the first 100ms after stimulus presentation as an impenetrable early stage (e.g., O'Callaghan et al., 2017), although the arbitrariness of this measure raises serious doubts about its accuracy. The main merit of this measure lies in its applicability: any evidence of changes in processing in the first 100ms after stimulus presentation as an effect of varying cognitive or affective states counts as a case of cognitive penetration. Speaking neuroanatomically, any changes in feedforward processing up to the extrastriate cortex caused by cognitive or affective biases should be taken as such cases. Claiming effects on perception on the basis of observations of increased activity in the visual cortex seems to be an established practice; for example, Gao and Scholl (2013) suggest that increased activity in the MT+ brain region is an argument supporting direct perception of animacy. However, one may be misguided by the simplicity of this approach. A substantial amount of evidence suggests that feedforward activations are not sufficient to evoke conscious awareness that may require local recurrent interactions that

bind and organize perceptual features (Lamme, 2004). Since these interactions are responsible for generating representations of basic visual properties (Raftopoulos, 2014), early vision possibly includes these interactions, therefore the temporal criterion should be extended up to 120–150ms. It is restricted to the first 100ms for the sake of certainty, but this creates the risk of overlooking some early vision modulations.

On the basis of the temporal criterion, and due to the fact that it is the lowest-level visual cortical part that receives input directly from thalamic afferents, the striate cortex seems to be the best bet in the context of the penetrability of perception dispute. Therefore, most arguments (both for and against) focus on changes in V1, but there are some reasons to consider as highly problematic such interest in V1: 1) only 20% of the variance of V1 activations may be explained by perceptual input (Carandini et al., 2005); 2) a lot of sensory stimulation on the retina is not processed at all by V1; 3) V1 receives much more top-down input (and from a broader part of a visual scene) from higher-level visual (V2, V3 or V5) or cognitive areas than bottom-up or lateral input (Muckli & Petro, 2013). Muckli and Petro (2013) suggest that the function of V1 may be to integrate feedback projections conveying cognitive or context information pertaining to the given visual scene with the incoming signal. Such an interpretation may suggest that early vision does not exist since no neuroanatomical structure fulfills the theoretical requirements pertaining to its architecture. This would mean that perception is either cognitively penetrable (since there is no encapsulated part dedicated to perceptual processing) or the problem is ill-posed (since early vision does not exist; Macpherson, 2016). Nevertheless, supporters of the orthodox standpoint argue that ubiquitous feedback and lateral projections may be wired into a broader "perceptual module" and taking context or crossmodal effects as examples of cognitive penetrability trivializes the problem (Firestone & Scholl, 2016).

Unanimous proof of the cognitively guided modulations in V1 would certainly settle the debate in favor of cognitive penetrability, since no plausible (at least for vision) neuronal locations of a perceptual module would remain. Nevertheless, it should be noted that since it is very challenging to accurately determine the exact projections—not to mention the accurate mapping from mental functions to neuronal structures and activations (Stokes, 2013)—the established criteria for early vision (100ms, modulations in the V1–MT pathway) are still an unsatisfactorily and uncertain neuroscientific praxis.[1] Therefore, most of the scientific reports on cognitive modulations of early visual cortex activity are discussed in chapter 4 in juxtaposition with experimental data pertaining to perceptual experience.

Cecchi (2014) analyses modulations of V1 activity in the context of "architectural cognitive penetration," which he defines as a process in which the structure or behavior of the perceptual system is affected by the cognitive system and, as a result, indirectly influences perceptual experience. He distinguishes synchronic cognitive penetration—cognitively

---

[1] This will hopefully change in the near future with the further development of methods of analysis that allow one to infer about causality (e.g., analysis of effective connectivity or dynamical causal modelling; Stephan & Friston, 2010; Friston, 2011; Friston et al., 2003; Friston et al., 2017)

influenced (e.g., via attentional guidance) modulation of the very beginning of visual processing—and diachronic cognitive penetration. The latter is defined as a long-term adaptation of neural organization that speeds up processing of visual properties to which cognitively guided attentional resources are continually deployed. This neural plasticity is more than just a mere adaptation of the visuomotor system to a changing perceptual environment since it is a result of a cascade of gradual synchronic cognitive penetrations which ultimately consolidate. Counting diachronic influences as cases of cognitive penetration seems particularly controversial and relies on whether we accept attentionally mediated modulations as a genuine cognitive impact (Macpherson, 2016; for considerations concerning fuzzy borders between attention and perception, see chapter 2).

4. Empirical Ways to Deal with the "Early Vision" on the Phenomenological Level

Pursuing a solution to the cognitive penetration problem on a neuroanatomical basis seems to be a tough nut to crack, but untying the knot of phenomenology may be even be more challenging. Since we do not have direct access to other persons' subjective experiences, we have to rely on verbal responses, psychophysical measures, performance in tasks related to perception of stimuli, and so on (Seth, 2008). Due to this indirectness, it may be very difficult to distinguish genuine effects on perception (on the directly captured experiential qualities or "how the objects are perceived") from effects on higher-order processes such as judgment (on which properties are ascribed to perceived objects on the basis of the act of perception; Firestone & Scholl, 2016b). Such perceptual judgments may not be convergent with what is actually perceived: for example, a banana in a dark room may be judged to be yellow in the absence of any experience of yellowness or an attacking tiger may be judged to be closer than it really is instead of being perceived closer. Perceptual judgments do not have to be an effect of conscious reasoning and may be predicated automatically (Stefanucci & Storbeck, 2014).

According to the definition, arguing for penetrability requires data that unequivocally implies effects on perception. Nevertheless, Firestone and Scholl (2016a, 2016b) point out that confusing perception and judgment is a common problem in contemporary research on higher-order influences on perception which are claimed to be found on the basis of overgeneralized declarative measures. Employing Likert scales (Banerjee et al., 2012), distance estimations (Cole et al., 2013) or adequateness assessments (Caruso et al., 2009) dramatically enhances the risk of judgment mediation; indeed, the obtained results are frequently proved to reflect effects on judgments. This was elegantly shown by Firestone and Scholl (2014), who replaced numerical scales with items prone to putative perceptual effects. Participants were asked to recall situations in which they performed moral and immoral deeds and, subsequently, to judge the lightness of the room. Interestingly, Firestone and Scholl applied gray patches instead of the numerical values used in the original study (Banerjee et al., 2012). Therefore, the putative perceptual darkening should affect the object of study (lightness perception) as well as the measuring device (gray patches); if the effects of thinking about non-ethical actions on perception were real, then the veil

of darkness would have overlaid the whole room participants were sitting in, including the patches. As a result, the patches would have also seemed darker and the effects would have canceled each other out since the same patches would have matched the lightness of the room in both groups. The averages for both groups should have been equal. That was not the case though: Firestone and Scholl replicated the original results (the "immoral" group chose darker patches as matching the perceived lightness), which means that the results arose from the experimenter effect or response bias. On the basis of the results discussed, one must not claim that they reflect direct effects on perception (that would be an example of the reasoning error Firestone and Scholl call the "el Greco fallacy"[2]).

4.1. Visual Matching

Balcetis (2015) proposes two alternative methods as remedies for these problems: visual matching and performance-based measures. In a visual matching procedure, participants adjust their position to the target (or the experimenter's position in relation to themselves) to match the distance between markers (or the distance from themselves to the target). If the distance between the participant and the target is shorter than between the targets, we may infer that the perceived distance is lengthened (and the other way around: if the real distance is larger, we infer that the perceived distance is shortened). However, the visual matching procedure seems problematic since it assumes that the perceived distance shortens or lengthens selectively to the object of interest and the rest of a perceptual scene is undistorted, including similar objects. For example, it has been shown that apertures are judged to be shorter when participants hold a horizontally oriented stick (Stefanucci & Geuss, 2009). Participants watched a tape being continuously elongated by the experimenter and were asked to say "stop" when they thought that the length of the tape matched the width of the aperture. Those who held a stick said "stop" significantly sooner. Nevertheless, claiming that such results reflect direct effects on perception relies on the assumption that perception of the aperture is penetrated but perception of the tape is adequate. However, it is unclear whether visual processing of basic properties (such as perception of width or horizontal lines), even if penetrated, could be biased selectively to the object. It seems possible since cognitive states may differ between people looking at the aperture (participants were told that they would walk through it) and the tape (walking through it is impossible). One may argue that judgments could have been biased by a series of selective "micro-penetrations" of the aperture during multiple attentional switches between the tape and the aperture. Nevertheless, the results obtained by Firestone and Scholl (2014), who replicated the experiment using a second aperture instead of a tape, support the simpler explanation that this procedure does not protect against response bias.

---

[2] The name of the reasoning error comes from the (popular at the beginning of twentieth century) theory that the elongated silhouettes on el Greco paintings were caused by his astigmatism. In this case though, el Greco would also have perceived the canvasses he used as longer, so elongated people painted on elongated canvasses would have seemed perfectly normal to an outside observer. Therefore, the theory is obviously wrong.

It is worth mentioning that Balcetis (2015) discusses visual matching procedure solely in the context of distance perception. However, visual matching may also be applied in psychophysical studies on color or lightness perception (Hansen et al., 2006; Witzel et al., 2011). Levin and Banaji (2006) asked participants (exp. 1) to adjust the luminance of the discolored faces of black and white people. Precisely, the participants' task was to increase or decrease the level of luminance of the adjustable face (e.g., white) so it would match the luminance of the target face (e.g., black). Participants tended to "overadjust" faces, picking objectively darker white faces as matching the lightness of the black faces (and vice versa, the effects were pronounced in both directions), as if their perception was cognitively penetrated by knowledge pertaining to racial differences. Moreover, similar results were obtained for merged faces (50% black and 50% white; exp. 2). Participants were asked to adjust the lightness of a gray rectangle to match the lightness of an ambiguous face that was previously presented as either black (in juxtaposition with a white face) or white (in juxtaposition with a black face). The obtained results were analogous: participants chose lighter samples to match "white" faces in comparison with "black" faces. Such procedures are particularly interesting since they are immune to the el Greco fallacy: the target and the adjustable probe differ in the direction of the expected top-down bias (exp. 1) or belong to distinct categories (alleged top-down influences affect the appearance of a face, but not of a gray patch; exp. 2).

Firestone and Scholl (2015, 2016a) criticized the study, pointing to the fact that the obtained results may still be caused by lower-level differences. For example, even though the mean luminance is equalized, the hue distribution differs in black faces (darker cheeks) and white faces (darker eyebrows and eye socket regions). Distinct distributions may bias the global impression of the lightness of the face. To examine this possibility, Firestone and Scholl (2015) blurred faces in such a way that they could not be recognized as belonging to different races (however, they did not apply a forced-choice measure—see below) and observed that, nevertheless, the blurred black face was significantly more often pointed to than darker ones. However, Baker and Levin (2016) responded with an experimental report suggesting that when the forced-choice measures are also applied to the race, the blurred faces are significantly more often correctly assigned, therefore it seems that some information concerning the race may be preserved despite the blurring. Moreover, Baker and Levin (2016) replicated the experiment with reversed colors of the blurred faces. They reasoned that the originally black faces should be described as lighter after the reversal if it was the hue distribution that influenced the choice. However, that was not the case: the reversed black faces were perceived as neither lighter nor darker (were selected as lighter or darker at random). Interestingly, reversed faces were correctly assigned to their race, which suggests that information pertaining to race is somehow read from the blurred faces.[3]

---

[3] It is also worth noting that the strict methodological requirements elaborated by Firestone and Scholl (2016a) may be non-reconcilable; for example, F&S demand that the studies should avoid presentation of the various objects (since they may differ in lower-level properties) and, at the same time, they point to the need for "amazing demonstrations" of the irresistibility of the phenomena of cognitive penetration that may be seen "in front of

Color matching procedures are particularly interesting since they go far beyond simple judgments and follow strict psychophysical rigor. Witzel et al. (2011) examined the influence on perception of the color diagnosticity of the object, which they defined as the strength of the association between the object and its usual color. In the first stage of the study, they isolated a set of highly color diagnostic objects. In the main part of the experiment, participants had to adjust the colors of these objects to objective gray. To measure the effect of the knowledge about the object on its color appearance, the memory color index developed by Hansen et al. (2006) was employed. According to the logic of this measure, in the case of penetration achromatic adjustments should be shifted towards the color that is opposite to the associated color (e.g., towards blue in the case of a banana) since it is biased adjustment that equalizes the effect of the typical color projected on the achromatic object. Therefore, the equally saturated opposite color (in reference to the subjective point of gray) should appear to participants as perfectly gray. Positive memory color indices (indicating opposite achromatic adjustments) were obtained for ten of the fourteen tested objects. "The MCIs [were] highest for objects with typical colors that are close to the yellow–blue daylight axis, such as yellow, blue and violet" (Witzel et al., 2011, p. 43). This may be explained by the fact that these colors are particularly perceptually variable during the day due to changing daylight illumination and, therefore, the color projection mechanism possibly secures the constancy of these colors. Interestingly, a reverse effect was obtained for red objects. Neither the complexity nor abstractness of the objects affected the adjustments.

This study is particularly valuable since it seems to avoid the pitfalls described by Firestone and Scholl (2016): it provides (unfortunately rather subtle) visual demonstrations of the effect (Witzel et al., 2016) and the effect is ecologically plausible (Witzel et al., 2011). However, it may be argued that it reflects the adaptation of the perceptual system rather than top-down cognitive modulations *per se* (Firestone & Scholl, 2016). In the case of genuine higher-order influences, one could expect that the color diagnosticity effect on perception should not be mediated by the incidence of exposure to the given object; for example, the strength of association between the color blue and a Nivea cream tin may be very strong, even if one does not use the product very often.

4.2. Performance-Based Measures

Performance-based measures have recently gained interest. However, it seems that not all these measures are properly tailored to application in studies on cognitive penetration. Most of all, behavioral responses should come after stimulus presentation under strict temporal

---

one's eyes." However, it remains unclear to me how both of these requirements may be achieved together in the case of Levin and Banaji's (2006) study 2: the same stimuli had to be presented either separately (in varying cognitive or affective states) or at the same time, but the latter situation would have required a strange double-think, e.g., that the ambiguous face on the left is white and the ambiguous face on the right is black.

pressure as unnecessary extension of the time window between object presentation and response increases the probability of cognitive intervention in behavior guidance. For example, Cole et al. (2013) employed a beanbag toss in their study, but their results cannot be generalized on perception since the object was present in the perceptual environment for quite a while. There is no reason to suppose that the differences in the force applied to the toss were a result of perceptual distortions ("I tossed closer since I saw the object was closer") rather than from judgments ("I judged the target to be close, so I adjusted the power of the toss"). Importantly, it is also possible to use "performance-based measures in which subjects' success is tied directly to how they perceive the stimuli" (Firestone & Scholl, 2016a, p. 25), e.g., a visual search task. In the study conducted by Gao et al. (2009), participants observed geometric shapes floating around a computer screen with some of them ("wolves") regularly approaching ("chasing") others ("sheep"). Subsequently, participants were asked to control the "sheep" in order to avoid being caught (touched) by the "wolf." The researchers observed that participants' performance was an interesting U-shaped function of maximal angular deviation of the wolf's heading towards the cursor, with performance rates dropping steeply for less constrained movement directions (maximal angular deviation 90°–120°), and then rising for completely unconstrained movements (randomly moving, "incompetent wolves"). Scholl and Gao (2013) claim that these results imply that animacy is directly perceived (not just cognitively ascribed) since a "wolf" heading directly towards the "sheep" popped out from the crowd as animate, while these approaching the sheep in a less constrained manner were perceptually melted with randomly moving identical distractors. Since the task was performed under constant time pressure, it is unreasonable to posit the involvement of higher-order processes. Performance relied strongly on the basic visual properties (motion direction, with other lower-level visual properties of the perceptual scene being controlled) of the objects in a visual scene—on "subtle display parameters, in the form of a psychophysical function (and in ways that do not seem readily explainable by appeal to higher-level judgment)" (Scholl & Gao, 2013, p. 19).[4]

Moreover, performance-based measures may not be directly tied to the manipulated stimuli. Gayet et al. (2016) point out that it is beneficial to employ a response task (e.g., reporting the orientation of lines filling figures) which is orthogonal to the experimental manipulation (e.g., fear conditioning to the color of figures). If the fear-conditioned stimuli were more visible in a brief presentation or a perceptual suppression task, this effect should be generalized to other basic properties of these stimuli. The latter should be reported in order to avoid response bias.

Usage of orthogonal measures relies heavily on the problematic assumption that one has to be conscious of given visual properties in order to successfully classify them. Such an assumption may be countered by pointing to the blindsight phenomenon (Weiskrantz, 1986),

---

[4] Note that this study pertains to the directedness of the perception of animacy, not to the cognitive penetrability of the perception problem. So-called "rich views" on perception (according to which we can directly experience higher-order properties of percepts) may be perfectly coherent with traditional views that perception is cognitively impenetrable (see Toribio, 2015).

which arises from widespread lesions in V1 due to which patients are able to properly classify stimuli with regards to a given property (e.g., up or down motion direction) while simultaneously claiming that they are unaware of them and even expressing doubts about the reasonableness of such testing. However, when thoroughly examined, these patients may report partial awareness in correct trials. In a study employing the Perceptual Awareness Scale (Overgaard et al., 2008), the relationship between awareness and performance was quite normal, with random performance for "no awareness at all" and rates rising steadily with increasing clarity of the conscious percept. Therefore, the authors of the PAS scale assume that the ability to report the properties of a given stimulus should be correlated with how clearly it was perceived, although this assumption has so far been taken for granted (Sandberg & Overgaard, 2015). If we accept that assumption, we can juxtapose indirect performance-based measures against subjective measures like PAS to let us adjudicate whether effects are on perception (when performance is higher in one experimental condition) or on judgment (similar performance in both conditions) in cases in which stimuli in the experimental group are consistently judged as "more visible." Moreover, the correlation coefficients between visibility and performance, even if imperfect, should not differ significantly between groups. It is also worth mentioning that such tasks should come under strict temporal pressure in order to minimize judgmental contamination.

Interestingly, performance-based measures may also be applied for participants that are unaware of an experimental manipulation at all. In a study performed by Seitz et al. (2009), participants deprived of food and water received occasional drops of water while passively fixating their gaze at the center of a screen which contained no task at all, but on which differently oriented gratings were subliminally presented. The small portions of water were paired in time with gratings of a particular orientation (the trained orientations differed among participants). After 9 days of the training phase, participants proceeded to the experimental part in which they were repeatedly presented with gratings in eight different signal-to-noise ratios (from 0.05 to 0.2) and had to indicate which one of the two gratings was presented. Their performance improved after the training phase for trained orientation, but not for control orientations, even though they were not aware of either of them during the training phase (70% correct for lowest signal-to-noise ratio compared to random performance for control orientations). Moreover, "these learning effects were specific to the eye to which the stimuli were presented, a hallmark of early visual processing" (Seitz et al., 2009, p. 701). According to Marx and Einhäuser (2015), these results show that "rewards may exert a direct influence on perception by modulating perceptual representations without or in addition to attentional mechanisms." Indeed, it may possibly be taken as an example of architectural diachronic cognitive penetration (Cecchi, 2014; see above) since the neural organization of the visual cortex selectively adapted to a given grating orientation and processed it much more efficiently as a result of a cascade of synchronic penetrations; concurrent reward processing possibly repeatedly exerted a direct influence on unconscious lower-level visual processing (e.g., reward information influenced processing of the faint yet regular activations in early visual areas to build proper associations).

4.3. Physiological Reflexes

Performance-based measures are not the only behavioral measures we have at our disposal—some of them are unequivocally tied to what is perceived on the basis of our knowledge about human physiology. Naber et al. (2011) thoroughly tested two such measurement methods, thereby circumventing the need for both verbal and non-verbal intentional responses. In the binocular rivalry paradigm (Blake & Logothetis, 2002), two dissimilar (albeit carefully standardized) stimuli are presented separately to the eyes, resulting in dynamic alternations between the percepts. These stimuli may differ in properties crucial in the context of human physiology; for example, in luminance, which affects pupil size. Naber et al. (2011) showed that pupil size was dependent on what was perceived: at the moment when participants were reporting that they perceived a brighter stimulus, the pupil was contracted and gradually adjusted its diameter shortly before and after perceptual transitions. The same patterns of dynamical adjustments have been observed when both stimuli have been presented to both eyes (monocular rivalry condition), thus suggesting that pupil size adapts to perceptual changes since physical stimulation is held constant.

The kin method is based on optokinetic nystagmus (OKN), an eye movement pattern present when a stationary subject follows an object in motion, or a subject in motion follows a stationary object. It comprises two phases: during the slow phase, the eye smoothly pursues the stimulus and tries to keep the retinal image stable (e.g., when observing a tree through the window of a moving train). During the fast phase, the eye abruptly resets its position to a subsequent coherently moving object (e.g., another tree). A similar situation may be enacted in a binocular rivalry condition, in which two standardized stimuli moving in opposite directions are presented to both eyes (Fox et al., 1975): the direction of the slow phase movement indicates the perceived stimulus. It is worth mentioning that OKN overlaps more with reports (~90%) than pupil dilations (~65%); some researchers have suggested that this means that the former is more reliable (Naber et al. 2011), although it may also be caused by sharper transitions in the OKN variant compared to frequently interfused percepts composed of stimuli of varying luminance. Such frequently or partially switching percepts may be much harder to report in a discrete manner. The data from "simulated rivalry," in which perceptual switches are simulated by alternate presentation of identical stimuli to both eyes, support the "harder to report" view. Correlation between pupil size adjustments and reports increased to 83% in simulated rivalry, suggesting that both methods are reliable (88% for OKN; Frässle et al., 2014).

Imperfect overlaps are probably caused by failures in report tasks rather than shortcomings of the method. Real-time coverage of one's subjective experience is a tremendously demanding task that recruits higher-order executive functions (metacognitive processes) to categorize frequently ambiguous percepts and constantly monitor the adequateness of reports. It has been shown that activity in the frontal regions of the brain is absent when participants are asked to simply observe the rivalry with no task at all (Frässle et al., 2014), and physiological reflexes explain much more variance of the neural activity that underlies binocular rivalry than subjective reports. These data suggest that even though the method

has been initially validated in reference to subjective reports, its greatest advantage is that the content of perception may be monitored solely on the basis of shifting physiological reactions (Block, 2014), and these reflexes are much more reliable indicators of the subjective experience!

Optokinetic nystagmus has been harnessed for experimental work on motivational influences on perception (Marx & Einhäuser, 2015). Depending on the condition, reward or punishment (either continuous aggregation or outflow of compensation money) was assigned to the perception of gratings drifting in one direction, whereas perceiving other gratings had no influence on participants' compensation. On the basis of monitored eye movements, participants were given feedback whether they aggregated or lost money; gratings were surrounded by annuli that accordingly varied in width (e.g., their width expanded when gratings coupled with reward were perceived). It was found that rewarded (or non-punished) percepts were seen much more often: the scale of the effect of reward was similar to the scale of effect of attention, which is well known to bias perception in the rivalry condition (Meng & Tong, 2004; van Ee et al., 2006). Nevertheless, these two effects were independent. In an attentional load condition in which endogenous attention was equal among the participants, reward (or lack of punishment) still increased the dominance of the rewarded (or unpunished) percept.

In my opinion, this study remains the most convincing argument for the cognitive penetrability of perception. Basic visual properties (color, luminance or movement direction) were processed preferentially when associated with reward: dominance times for rewarded or unpunished percepts were extended, whereas participants showed no preference towards either of them when watching the same gratings prior to the experimental manipulation. Therefore, it is reasonable to think that all of the stimuli used had properties that are equally efficiently processed by the human brain, and the bottom-up biases were equated. The response (automatism on the level of eyeball movements) does not entail mediation by higher-order processes of any kind. What is more, these eyeball movements were passive, i.e., guided by the perceptual experience.

However, one objection may be raised here. One could perhaps disagree that the basic visual properties were processed in a different manner since color or movement of the gratings possibly seemed the same to the participants whether associated with reward or not (this was not controlled) and claim that it is the preferential *selection* of the stimuli rather than *processing*. In my opinion, neuroscience may contribute to solving this problem. Lateral geniculate nuclei (LGN) receive modulatory top-down projections from V1 and feedforward signals strictly from the retina (Cudeiro & Sillito, 2006). Therefore, they represent the first stage in the visual pathway at which cortical top-down inhibitory signals could affect selection processes (Saalmann & Kastner, 2011), e.g., via selective turning down of the LGN layers carrying the signal from the eye to which the punished stimuli are presented. In this context, we could conceive of perception as a "self-fulfilling prophecy," with the perceptual system just picking what it wants to perceive in congruence with higher-order cognitive demands. However, V1 modulations from the pulvinar have been

observed to bias competitions in the early visual cortex (Purushothaman et al., 2012). Such modulations could influence the activity of direction-sensitive (reacting to the particular direction of movement) cells in V1. "When the spatiotemporal context of a visual stimulus autonomously enhances its salience in conflict with behavioral or top-down goals, the lateral pulvinar can suppress neural responses to this stimulus in the early visual cortex, thus biasing the competition in favor of behaviorally relevant stimuli" (Purushothaman et al., 2012, p. 7). Observation of such inhibitory modulations would support attentionally mediated cognitive influences on perception in accordance with biased competition theory. Thirdly, direct linear top-down projections from the prefrontal cortex via the visual pathway would support direct cognitive penetrations (O'Callaghan, 2016). Dynamic causal modelling could certainly help nuance these models and comprehensively reveal which one of them explains the neuronal signals underlying motivational influences on bistable perception described by Marx and Einhäuser (2015).

Surprisingly, physiological reflexes are omitted in the recent influential paper by Firestone and Scholl (2016), who claim that none of the foregoing studies on higher order influences on perception meet strict methodological requirements, therefore, cognitive penetrability has not yet been proven. They indicate six major pitfalls that may be responsible for the widespread tendency to confuse effects on perception with other effects, and they provide six corresponding recommendations. Most of them have been discussed above (apply unique disconfirmatory predictions, distinguish perception from judgment, avoid response bias, control attention, carefully match the features not relevant to the higher-level factors, evoke intergroup differences in cognitive states rather than the stimuli presented) or pertain to the particular scientific problem (distinguish effects on perception from well-known effects on memory in studies on recognition; Firestone & Scholl, 2016). According to Firestone and Scholl, these high standards must be met: only the convergence of precise indicators may allow us to claim (on the basis of inference to the best explanation) that the effects are on perception. Physiological reflexes seem to offer more, opening a new promising way of scientific simplicity that does not require a trade-off of certainty.

5. Cognitive Penetrability of Perception and Predictive Processing Theory of Cognition

In the last section of the article, I would like to briefly outline one of the currently dominant theories of cognition (predictive processing [PP]) and consider the additional explanatory potential it bears in the context of the penetrability of perception problem (O'Callaghan et al., 2017). According to PP, perception is a constant interplay between incoming sensory evidence and top-down predictions based on the internally generated model of the world which develops via continuous interaction with the environment (Friston, 2010; Clark, 2013). Based on prior experience, predictions originate from very general and abstract expectations (represented in the highest levels of the model) and constrain more and more detailed predictions on the lower levels, down to early perceptual areas where they guide perceptual processing (Seth et al., 2011). In the face of incongruent sensory evidence, the model verifies its assumptions and reorganizes itself in order to explain

away prediction errors. As a result, the model constantly evolves to efficiently resolve ambiguities inherent in sensory signals and represent biologically significant objects and events in a way that best serves the survival of an organism.

Cognitive penetrability of perception is a natural consequence of PP, since the plethora of higher-order information from expectations, knowledge, memory, affective states, motivations and other modalities continuously sculpt the most probable interpretations of the incoming sensory signals (Lupyan, 2015). The very idea of perception as inference makes the idea of an impenetrable early perceptual system sound strange: one would have to posit the encapsulated part that outputs the basic visual representations which the rest of the system constantly tries to predict. Therefore, it seems that since the whole system contributes to the production of basic perceptual representations, one should conclude that early vision either does not exist or is cognitively penetrable (Macpherson, 2016); either way, the idea of an encapsulated part of a perceptual system is false.[5]

PP may provide plausible explanations for the described cases of putative cognitive influences on perception. For example, binocular rivalry (harnessed in a study performed by Marx and Einhäuser) has recently been neatly elaborated in PP terms (Hohwy et al., 2008). The alterations are supposed to happen because of a general, "ontological" expectation that two solid objects cannot co-exist in the same place, which constrains lower-level perceptual content (Clark, 2013). Therefore, the mixed percept is very unlikely to reflect the causes of the sensory signals, and only one stimulus is picked up for conscious experience; this leaves a substantial number of the bottom-up signals (associated with a suppressed stimulus) unexplained. As the domination time of the attractor extends, prediction errors associated with the repeller propagate up the hierarchy, therefore another hypothesis must be chosen to explain them away. The percept switches, but the new hypothesis is incapable of explaining away prediction errors associated with the preceding percept, and so on. In this view, reward and punishment associations form another source of information that skews perceptual processing towards preferable interpretations; however, the prediction errors associated with punishment stimuli are too precise to be fully muted and so the size of the effect is constrained.

---

[5] Note that, in this view, cognitive penetration is simply equated with perception which is "omnipenetrable," as all cases of perceptual inference are cases of cognitive penetration. Some philosophers find this view trivial (Macpherson, 2016) and aim to nuance the notion of cognitive penetration within a predictive processing framework (Hohwy, 2017). Hohwy (2017) describes cognitive penetration as a computationally suboptimal (too slow) Bayesian learning rate which is caused by inefficient higher-level models encoding general expectations on how to minimize error/optimize the learning rate [e.g., in new environments]. A learning rate that is lower than optimal means that bottom-up prediction error signals are dampened even though sufficiently invariable sensory evidence is present. As a result, the model's optimization rate is too slow and provides percepts excessively relying on top-down driven inferences. Such percepts may be understood as "false" in Bayesian terms. However, the simplified view on penetration within PP that perception is omnipenetrable seems to be sufficient for the purposes of the present article, which is focused on empirical perspectives on unambiguous rejection of the traditional, modular view.

However, an important contribution of PP theory to experimental work on the penetrability of perception is that it makes specific predictions not only about the existence of penetrability, but also about its scope and character. Firstly, attenuated feedforward processing is associated with predictions matching actual sensory evidence, since feedback modulations silence feedforward projections carrying a predictable signal. Therefore, diminished activation in early stages of perceptual processing should be incidental to the expected stimuli. An experimental study employing a repetition suppression paradigm has shown that this is indeed the case (Grill-Spector et al., 2006) unless the repetition itself is unexpected (Summerfield et al., 2008). It has even been suggested that neurons respond more to feature expectations than stimulus features themselves because, under high expectation of a face, FFA neurons respond strongly to stimuli belonging to other categories (e.g., a house; Egner et al., 2010). Secondly, an experiment carried out by Kok et al. (2012a) shows that expectations do not simply reduce overall neural activity in V1, but they also sharpen visual representations in this area. In the experiment, participants were presented consecutively with two gratings differing slightly in orientation; they were then asked to perform either an orientation or contrast task. In the orientation task, they repeatedly decided whether latter stimuli were oriented clockwise or anticlockwise in reference to the former stimulus. In the contrast task, their judgments referred to the brightness of the latter stimulus (whether it was brighter or darker than the former). The gratings were preceded by an auditory cue indicating with high (75%) probability the orientation of the subsequently presented gratings.[6] Unsurprisingly, expectations strongly suppressed neural activity in V1. Much more importantly, analyses employing MVPA (multivariate pattern analysis) showed that orientation classification accuracy on the basis of activity in V1 was improved when gratings were expected compared to when they were unexpected. Neurons sensitive to non-presented orientations were prevented from firing as they were "silenced" via top-down feedback projections conveying expectations. This representational sharpening was also largely correlated with participants' improvement in an orientation discrimination task (lowered discrimination threshold), suggesting that perception was also sharpened. Taken together, these results account for penetrability of perception in a specific way that is coherent with PP.

It is worth mentioning that the concept of attention, which is closely related to the concept of precision (Hohwy, 2012), is also nuanced within PP theory. Precision (which is also constantly predicted by the model) indicates the reliability of the bottom-up signal. The influence of the prediction errors on the subsequent predictions is weighted proportionally to how invariable these errors are expected to be. For example, in a foggy gathering dusk, sensory signals may be expected to be highly variable, therefore their influence on subsequent predictions is dampened—the posteriors will rely on the priors to a greater extent.

---

[6] Importantly, the task was orthogonal to the orientation expectation in order to distinguish the expectation effects in the visual cortex from noise generated by action preparation. In a task directly pertaining to the expectation manipulation, participants could be convinced with 75% certainty what the orientation would be, and thus prepare the motor response beforehand.

In this view, sensory signals predicted to have a smaller dispersion of variance (high precision expectations) are the ones to which we attend. High precision expectations may be derived both from the properties of the signal (its constant invariance—bottom-up attention) or selective alterations of the gain of error units in the context of the particular task or problem (top-down attention; Clark, 2013). Note that, similarly to the biased-competition theory, attention is de-reified and is taken as *a result* of the processing rather than as a mysterious cause (see Mole, 2015). Accordingly, such attentional effects should be understood as genuine effects on perception in the context of the cognitive penetration debate. Covertly attended stimuli are perceived as if they have higher contrast rates (Ling & Carrasco, 2006; Carrasco et al., 2004) not because of the spotlight mechanism but because when they are endogenously attended higher precisions are ascribed to the weak activations that were previously balancing on the verge of consciousness. This precision weighting is a key factor that streams these subtle activations through the cognitive system, making them capable of influencing posteriors and clearing their way to consciousness. It is also worth noting that, in line with PP, attention dampens the sensory attenuation of the expected signals, since the residual bottom-up signals have higher precisions (Kok et al., 2012b).

6. Conclusions

The cognitive penetrability of perception problem is certainly barely approachable from empirical positions. Since accurate mappings from mental functions to neuronal activities are lacking at the current stage of the development of neuroscience, scientific progress in solving this problem on the neuronal level is largely impeded (Stokes, 2013). Lack of direct access to other persons' subjective experience seems to condemn us to scientific compromises (relying on first-person reports etc.; Seth, 2008). However, the state of the art does not look as grim as is suggested by some prominent researchers on the topic (Firestone & Scholl, 2016; Machery, 2015). A wide variety of ingenious methods from physiology, experimental psychology and neuroscience may be applied in combination in order to distinguish genuine effects on perception from merely attentional (in a "spotlight" sense) or judgmental ones. Some of the reports harnessing these methods strongly imply that perception is penetrable (Marx & Einhäuser, 2015; Kok et al., 2012a; Witzel et al., 2011) and they have not hitherto been critically targeted by the supporters of the orthodox standpoint. The constant development of new apparatus and analysis methods (e.g., effective connectivity for fMRI and EEG data; dynamic causal modelling, simultaneous recordings in frontal and posterior brain areas) coupled with paradigms allowing for increasingly certain inferences about perceptual experience will possibly soon allow an unambiguous answer to the question of whether cognitive penetration occurs. I predict this answer will be affirmative.

Bibliography

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology, 2*, 246. doi:10.3389/fpsyg.2011.00246

Baker, L. J., & Levin, D. T. (2016). The face-race lightness illusion is not driven by low-level stimulus properties: An empirical reply to Firestone and Scholl (2014). *Psychonomic Bulletin & Review, 23*(6), 1989–1995. doi:10.3758/s13423-016-1048-z

Balcetis, E. (2015). Approach and avoidance as organizing structures for motivated distance perception. *Emotion Review, 8*(2), 115–128. doi:10.1177/1754073915586225

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology, 91*(4), 612–625. doi:10.1037/0022-3514.91.4.612

Banerjee, P., Chatterjee, P., & Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science, 23*(4), 407–409. doi:10.1177/0956797611432497

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences, 18*(9), 445–447. doi:10.1016/j.tics.2014.05.007

Brogaard, B., & Gatzia, D. E. (2015). Is the auditory system cognitively penetrable? *Frontiers in Psychology, 6*(643), 1072. doi:10.3389/fpsyg.2015.01166

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., & Rust, N. C. (2005). Do we know what the early visual system does? *The Journal of Neuroscience, 25*(46), 10577–10597. doi:10.1523/JNEUROSCI.3726-05.2005

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. doi:10.1016/j.visres.2011.04.012

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*(3), 308–313. doi:10.1038/nn1194

Caruso, E. M., Mead, N. L., & Balcetis, E. (2009). Political partisanship influences perception of biracial candidates' skin tone. *Proceedings of the National Academy of Sciences of the United States of America, 106*(48), 20168–20173. doi:10.1073/pnas.0905362106

Cecchi, A. S. (2014). Cognitive penetration, perceptual learning and neural plasticity. *Dialectica, 68*(1), 63–95. doi:10.1111/1746-8361.12051

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204. doi:10.1017/s0140525x12000477

Cole, S., Balcetis, E., & Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychological Science, 24*(1), 34–40. doi:10.1177/0956797612446953

Cudeiro, J., & Sillito, A. M. (2006). Looking back: Corticothalamic feedback and early visual processing. *Trends in Neurosciences, 29*(6), 298–306. doi:10.1016/j.tins.2006.05.002

Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience, 16*(6), 763–770. doi:10.1038/nn.3381

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222. doi:10.1146/annurev.ne.18.030195.001205

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience, 30*(49), 16601–16608. doi:10.1523/JNEUROSCI.2770-10.2010

Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science, 25*(1), 38–46. doi:10.1177/0956797613485092

Firestone, C., & Scholl, B. J. (2016a). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences, 39*, 1–77. doi:10.1017/s0140525x15000965

Firestone, C., & Scholl, B. J. (2016b). When do ratings implicate perception versus judgment? The "overgeneralization test" for top-down effects. *Visual Cognition, 23*(9–10), 1217–1226. doi:10.1080/13506285.2016.1160171

Fodor, J. A. (1983). The modularity of mind: An essay on faculty psychology. Cambridge, MA: A Bradford Book / MIT Press.

Fox, R., Todd, S., & Bettinger, L. A. (1975). Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Research, 15*(7), 849–853. doi:10.1016/0042-6989(75)90265-5

Frassle, S., Sommer, J., Jansen, A., Naber, M., & Einhauser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience, 34*(5), 1738–1747. doi:10.1523/JNEUROSCI.4403-13.2014

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. doi:10.1038/nrn2787

Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain connectivity, 1*(1), 13–36. doi:10.1089/brain.2011.0008

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage, 19*(4), 1273–1302. doi:10.1016/S1053-8119(03)00202-7

Friston, K. J., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., & Zeidman, P. (2017). Dynamic causal modelling revisited. NeuroImage. Advance online publication. doi:10.1016/j.neuroimage.2017.02.045

Gayet, S., Paffen, C. L. E., Belopolsky, A. V., Theeuwes, J., & van der Stigchel, S. (2016). Visual input signaling threat gains preferential access to awareness in a breaking continuous flash suppression paradigm. *Cognition, 14*9, 77–83. doi:10.1016/j.cognition.2016.01.009

Goldstone, R. L., de Leeuw, J. R., & Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition, 135*, 24–29. doi:10.1016/j.cognition.2014.11.027

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14–23. doi:10.1016/j.tics.2005.11.006

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368. doi:10.1038/nn1794

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology, 3*, 96. doi:10.3389/fpsyg.2012.00096

Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition, 47*, 75–85. doi:10.1016/j.concog.2016.09.004

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108*(3), 687–701. doi:10.1016/j.cognition.2008.05.010

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*(5), 679–685. doi:10.1038/nn1444

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012a). Less is more: expectation sharpens representations in the primary visual cortex. Neuron, 75(2), 265–270. doi:10.1016/j.neuron.2012.04.034

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & d Lange, F. P. (2012b). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex, 22*(9), 2197–2206. doi:10.1093/cercor/bhr310

Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks, 17*(5–6), 861–872. doi:10.1016/j.neunet.2004.02.005

Lee, S.-H., Blake, R., & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature Neuroscience, 8*(1), 22–23. doi:10.1038/nn1365

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135*(4), 501–512. doi:10.1037/0096-3445.135.4.501

Ling, S., & Carrasco, M. (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research, 46*(8–9), 1210–1220. doi:10.1016/j.visres.2005.05.008

Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin, 130*(5), 748–768. doi:10.1037/0033-2909.130.5.748

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. doi:10.1007/s13164-015-0253-4

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. doi:10.1111/j.1933-1592.2010.00481.x

Macpherson, F. (2016). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition, 47*, 6–16. doi:10.1016/j.concog.2016.04.001

Marchi, F. (2017). Attention and cognitive penetrability: The epistemic consequences of attention as a form of metacognitive regulation. *Consciousness and Cognition, 47*, 48–62. doi:10.1016/j.concog.2016.06.014

Martinez-Conde, S. (Ed.). (2006). *Progress in Brain Research: Vol. 155, Visual perception* (1st ed.). Amsterdam, the Netherlands: Elsevier.

Marx, S., & Einhauser, W. (2015). Reward directly modulates perception in binocular rivalry. *Journal of Vision, 14*(10), 502. doi:10.1167/14.10.502

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision, 4*(7), 539–551. doi:10.1167/4.7.2

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action (Oxford science publications: No. 27)*. Oxford, UK: Oxford University Press.

Miller, E., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science, 254*(5036), 1377–1379. doi:10.1126/science.1962197

Muckli, L., & Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Current Opinion in Neurobiology, 23*(2), 195–201. doi:10.1016/j.conb.2013.01.020

Naber, M., Frassle, S., & Einhauser, W. (2011). Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness. *PLoS ONE, 6*(6), e20910. doi:10.1371/journal.pone.0020910

O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., & Bar, M. (2016). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition, 47*, 63–74. doi:10.1016/j.concog.2016.05.003

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In S. Martinez-Conde (Ed.), *Progress in Brain Research: Vol. 155, Visual perception* (pp. 23–36). Amsterdam, the Netherlands: Elsevier.

Overgaard, M. (Ed.). (2015). *Behavioral methods in consciousness research*. Oxford, UK: Oxford University Press.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience, 31*(10), 1783–1790. doi:10.1111/j.1460-9568.2010.07193.x

Posner, M. I. (2008). Measuring alertness. *Annals of the New York Academy of Sciences, 1129*, 193–199. doi:10.1196/annals.1417.011

Purushothaman, G., Marion, R., Li, K., & Casagrande, V. A. (2012). Gating and control of primary visual cortex by pulvinar. *Nature Neuroscience, 15*(6), 905–912. doi:10.1038/nn.3106

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(03). doi:10.1017/s0140525x99002022

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.

Raftopoulos, A. (2013). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philosophical Psychology, 27*(5), 601–620. doi:10.1080/09515089.2012.729486

Saalmann, Y. B., & Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron, 71*(2), 209–223. doi:10.1016/j.neuron.2011.06.027

Sandberg, K., & Overgaard, M. (2015). Using the perceptual awareness scale (PAS). In M. Overgaard (Ed.), *Behavioral methods in consciousness research* (pp. 181–196). Oxford, UK: Oxford University Press.

Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron, 61*(5), 700–707. doi:10.1016/j.neuron.2009.01.016

Seth, A. K. (2008). Theories and measures of consciousness develop together. *Consciousness and Cognition, 17*(3), 986–988. doi:10.1016/j.concog.2007.08.004

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00395

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs, 46*(2), 201–222. doi:10.1111/j.1468-0068.2010.00786.x

Stefanucci, J. K., & Geuss, M. N. (2009). Big people, little world: The body influences size perception. *Perception, 38*(12), 1782–1795. doi:10.1068/p6437

Stephan, K. E., & Friston, K. J. (2010). Analyzing effective connectivity with fMRI. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 446–459. doi:10.1002/wcs.58

Stokes, D. (2013). Cognitive penetrability of perception. *Philosophy Compass, 8*(7), 646–663. doi:10.1111/phc3.12043

Storbeck, J., & Stefanucci, J. K. (2014). Conditions under which arousal does and does not elevate height estimates. *PLoS ONE, 9*(4), e92024. doi:10.1371/journal.pone.0092024

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*(9), 1004–1006. doi:10.1038/nn.2163

Toppino, T. C. (2003). Reversible-figure perception: Mechanisms of intentional control. *Perception & Psychophysics, 65*(8), 1285–1295. doi:10.3758/BF03194852

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

van Ee, R., Noest, A. J., Brascamp, J. W., & van den Berg, A. V. (2006). Attentional control over either of the two competing percepts of ambiguous stimuli revealed by a two-parameter analysis: means do not make the difference. *Vision Research, 46*(19), 3129–3141. doi:10.1016/j.visres.2006.03.017

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27*, 62–75. doi:10.1016/j.concog.2014.04.007

Weiskrantz, L. (1986). *Blindsight: A case study and implications (Oxford psychology series: no. 12)*. Oxford, UK: Clarendon.

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

Yin, X., Zhao, L., Xu, J., Evans, A. C., Fan, L., Ge, H., … (2012). Anatomical substrates of the alerting, orienting and executive control components of attention: Focus on the posterior parietal lobe. *PLoS ONE, 7*(11), e50590. doi:10.1371/journal.pone.0050590

Zeimbekis, J., & Raftopoulos, A. (Eds.). (2015). *The cognitive penetrability of perception*. Oxford, UK: Oxford University Press