COGNITIVE IMPENETRABILITY OF LOW-LEVEL PERCEPTUAL FEATURES

By

David Bitter

Submitted to

Central European University

Department of Philosophy

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Supervisor: Professor Katalin Farkas

Budapest, Hungary

2017

Copyright notice

Abstract

Is our perception of the world influenced by what we believe or desire the world to be? The converse is surely true: we often come to believe and desire what we do on the basis of what we see, hear, touch, taste, and smell. Might perception be informed by cognition in a similarly specific and direct manner? In particular, might cognition influence perceptual experience in a way that sustains some semantic or logical or rational coherence between the source cognitive content and the resultant perceptual content, whereby this influence is not mediated by changes in the stimuli, the state of the sense organs, or the allocation of attention?

If the answer is yes, then perception is cognitively penetrable. Assumedly, this would have interesting theoretical implications regarding mental architecture (philosophy of mind and psychology), the justificatory role of perception (epistemology), and the theory-ladenness of observation (philosophy of science).

The central thesis of this dissertation is that, a trending view in philosophy and psychology to the contrary notwithstanding, there is in fact little reason to assume that perceptual experience of low-level features such as color or lightness is cognitively penetrable. It is not claimed outright that such an effect is empirically impossible. But it is argued that the prima facie best evidence and the arguments that build on them fall short of providing a convincing case for cognitive penetrability. So the burden is on defenders of the view to explain why anyone should still entertain that the posited kind of influence actually occurs.

Closely connected to the negative thesis regarding cognitive penetrability are two positive theses. One is that cognitively impenetrable perceptual processing is plausibly richer than is commonly assumed. It is argued in particular that intra-perceptual form–lightness

associations and lightness contrast may explain away an apparent lexical effect that has been most widely proposed as a prime candidate of cognitive penetration.

The second positive thesis is that some of the best candidates of cognitive penetration may be better explained by changes in affect than by changes in perception or perceptual belief. It is argued in particular that even if the phenomenon of (hypnotically) suggested hallucination is not a sham, it still doesn't follow that subjects experience genuine perceptual hallucination or cognitive delusion. More plausibly, just as suggested analgesia may involve a dissociation between the sensory and affective components of pain, it only feels to subjects as if the suggested perceptual state of affairs really pertained.

The dissertation thus defends cognitive impenetrability of perceptual experience while expanding on both the intra-perceptual and the extra-perceptual. The intra-perceptual account carries special weight insofar as it underscores the philosophical relevance of cognitively impenetrable perceptual processes. The importance of the extra-perceptual account is that it exposes as false a dichotomy according to which all candidate effects are either perceptually or cognitively mediated. Thus, in addition to the general conclusion that cognitive penetrability of perception is empirically implausible, a crucial implication of the dissertation is that the very dialectic of the cognitive-penetrability debate itself requires mending.

Acknowledgments

Katalin Farkas, you are the best. Your superb and gracious guidance and support will forever be appreciated. Hanoch Ben-Yami and Howard Robinson, thank you for so many challenging discussions and your incessant personal kindness. Gábor Betegh, I am indebted to your wisdom that we should allow ourselves to dream big dreams. Thomas Nagel and Tamar Gendler, my gratitude goes out to you for having enabled my visits to your fine institutions. Ned Block, I am obliged for your temporary supervision. Chaz Firestone, Steven Gross, Edouard Machery, Fiona Macpherson, Brian Scholl, and Dustin Stokes, your comments and encouragement have positively facilitated my thought and work. Jerry Fodor and Zenon Pylyshyn, this dissertation is dedicated to your intellectual inspiration. My loved ones, thank you for all the moral support.

Table of Contents

1. Introduction

Is our perception of the world influenced by what we believe or desire the world to be? The converse is surely true: we often come to believe and desire what we do on the basis of what we see, hear, touch, taste, and smell. Could perception itself be informed by cognition in a similarly specific and direct way? If the answer is yes, then perception is cognitively penetrable.

It would follow that perception is in an important sense continuous with cognition; that the structure of belief formation may in some cases be circular; and that scientific observation may itself be theory-laden. Such consequences would have important implications for theories of mental architecture and hence philosophy of mind and psychology; foundationalist theories of justification and hence epistemology; and debates regarding the role of observational evidence in theory testing and hence philosophy of science and scientific practice in general. It is thus natural that there is growing interest in whether perception is cognitively penetrable.

So is it? On the face of it, the answer would seem to be yes. For example, there is the story of the desert nomad whose desire for water leads him to hallucinate a source of water where there is none. Or what about the amputee whose refusal to accept the loss of a limb results in her experiencing pain in a body part that does not exist? The list of candidate examples is long. Accordingly, it might seem evident that our perceptual experience is cognitively penetrated through and through.

On closer scrutiny, though, the situation is not so simple. For example, mirages are real optical phenomena caused by atmospheric conditions. People tend to see water in a desert where there is none because light from the sky is refracted by hot air above the desert surface. From a certain distance and angle, the effect indeed visually resembles a sheet of water on the

ground. So the source of the respective percept is not cognitive, even if the resulting water-related belief undoubtedly is. Hence, mirages are not good examples of cognitive penetration.

Nor is the source of phantom limb pain cognitive. True, illusory visual feedback of a phantom hand caused by the mirror reflection of an amputee's normal hand can lead to an alleviation of phantom pain (Ramachandran, Rogers-Ramachandran, & Cobb, 1995). But the effect is presumably a function of somatosensory–motor coupling between the normal and the phantom hand, which in turn is plausibly explained by cognitively unmediated Hebbian (associative) learning. So phantom limb pain is not a convincing case of cognitive penetration, either.

Of course, there *are* perceptual effects the source of which is cognitive. But in many if not all cases, the relation between the source cognitive state and the resultant perceptual state does not seem to be of the right kind. For example, perceptual effects mediated by changes in the stimuli (e.g., turning off the lights), the state of the sense organs (e.g., closing one's eyes), or the allocation of (spatial or feature-based) attention are widely disregarded as genuine cases of cognitive penetration. In these cases, a cognitive source seems neither necessary, nor sufficient for the relevant effects to occur. On the other hand, the changes in the stimuli, the state of the sense organs, or the allocation of attention seem both necessary and sufficient. Accordingly, once such factors are fixed, many candidate phenomena (e.g., voluntary switches between alternate perspectives of ambiguous figures such as the Necker cube) turn out not to be cases of cognitive penetration (Pylyshyn, 1999a, 2003; Raftopoulos, 2001, 2009).

A further issue is that many perceptual effects aren't arbitrarily sensitive to the contents of cognition. For example, selective attention might alter the impedance of the ear or the aperture of the eye. Yet such influences may only affect the amplitude of a transducer's response, rather than the stimulus property to which the response is specific (Fodor &

Pylyshyn, 1981). It is thus traditionally considered a crucial condition of cognitive penetration that the effect should sustain some semantic coherence or logical–rational relation between the penetrating cognitive content and the penetrated perceptual content (Macpherson, 2012, 2015; Pylyshyn, 1981, 1984, 1999a, 2003; cf. also Stokes, 2013).

Contrary to its prima facie plausibility, then, it is in fact hard to find a convincing empirical example of cognitive penetration, if by the term we mean:

(CogPen) Phenomenally conscious perceptual experience is cognitively penetrated $=_{df}$

(i) the locus of the effect is genuinely *perceptual*;

(ii) the source of the effect is genuinely *cognitive*;

(iii) the effect sustains some *semantic coherence* or *logical* or *rational relation* between the penetrating cognitive content and the penetrated perceptual content;

(iv) the effect is *not mediated* by changes in the stimuli, the state of the sense organs, or the allocation of attention.

It is against this backdrop that a particular set of psychological experiments has been singled out not long ago as providing a possibly most convincing case yet for cognitive penetration. In these experiments, subjects consistently judged a grayscale image of a prototypical white (Caucasian) face as lighter than an objectively equiluminant grayscale image of a prototypical black (African American) face. Interestingly, subjects also judged the lightness of a grayscale image of a racially ambiguous face differently when it was labeled 'White' rather than 'Black' (Levin & Banaji, 2006). It is especially in consideration of the latter result that the respective findings have been widely interpreted as demonstrating that even

relatively abstract beliefs or concepts such as regarding race can influence perceptual experience of relatively low-level visual properties such as color and lightness (e.g., Collins & Olson, 2014; Hugenberg & Sacco, 2008; Levin & Banaji, 2006; Macpherson, 2012; Stokes, 2013; Vetter & Newen, 2014).

However, on the face of it, there is even more convincing evidence for cognitive penetration: that related to the phenomenon of (hypnotically) suggested hallucination (Arstila, 2017; Bitter, 2015a, 2015b, 2017; Newen & Vetter, 2017). In hypnosis, typically, one person (the subject) is guided by another (the hypnotist) to respond to suggestions for changes in subjective experience, possibly involving alterations in perception, thought, emotion, or behavior (Green, Barabasz, Barrett, & Montgomery, 2005; Kihlstrom, 2008). For example, in a classic study, it was suggested to subjects under hypnosis that they would be presented colored pieces of paper. Although all pieces of paper were actually white, four out of five subjects claimed to see the suggested colors. When these subjects were asked about the color of the next piece of paper, they consistently claimed to see the opposite color of the earlier suggested color (e.g., blue after yellow). Based on an apparent ignorance of negative afterimages and complementary colors, it was argued that the latter responses reflected spontaneous negative color afterimages of genuine perceptual hallucinations (Erickson & Erickson, 1938; cf. also Arstila, 2017; Barber, 1959a, 1964; Rosenthal & Mele, 1952).

If the above claims are true, and color and lightness experience are indeed relevantly influenced by corresponding beliefs or concepts, then this has enormous consequences for the debate on cognitive penetrability. For if such low-level experiential features are cognitively penetrable, then quite possibly any or at least most of perceptual experience is also cognitively penetrable. As already noted, this has potentially important implications for issues such as concerning the justificatory role of perception, the theory dependence of observation, and of

course the continuity of perception with cognition and hence mental architecture. So it bears great relevance whether or to what extent the above claims are warranted (cf. Fodor, 1983, 1985; Lyons, 2005, 2015; Raftopoulos & Zeimbekis, 2015; Siegel, 2012, 2015; Stokes, 2015).

A central thesis of this dissertation is that, a trending view in philosophy and psychology to the contrary notwithstanding, there is in fact little reason to assume that perceptual experience of low-level features such as color and lightness is cognitively penetrable. It is not claimed outright that such effects are empirically impossible. But it is argued that the prima facie best evidence and the arguments that reference them fall short of providing a convincing case for cognitive penetrability. It bears emphasis that although the dissertation focuses on a restricted set of phenomena, this is the set based on recent proposals that apparently constitutes the strongest evidence for relevant infiltration of perception by cognition. Hence, if the pending argumentation is correct, the burden is on defenders of cognitive penetrability to explain why anyone should still entertain that the posited kind of penetration really occurs.

Closely connected to the negative thesis regarding the insufficient abductive plausibility of cognitive penetrability are two positive theses. One is that cognitively impenetrable perceptual processing is plausibly richer (even if not more sophisticated) than is commonly assumed. It is argued in particular that partly associative intra-perceptual processes more plausibly account for the alleged face–race lightness illusion mentioned above than top-down cognitive influences do.

Of course, the general idea that relatively simple modular functions may explain (away) prima facie sophisticated effects is not new (cf., e.g., Deroy, 2013; Fodor, 1983; Pylyshyn, 1999a, 1999b). However, no one before has shown how the idea might apply to the specific case of interest (Bitter, 2013a, 2013b, 2013c, 2014). Thus, in addition to a detailed analysis of

why a prima facie most powerful argument to the contrary is unwarranted, an important contribution in this regard is the development of a specific empirical hypothesis concerning how intra-perceptual form–lightness or shape–color associations may mediate an apparent lexical effect that has been most widely proposed as a prime example of cognitive penetration.

The second positive thesis of the dissertation is that various candidates of cognitive penetration may be better explained by changes in affect, rather than penetration of perception or perceptual belief. In contrast with the first positive thesis—which expands on the intra-perceptual and suggests that the source of certain candidate effects is not cognitive—the second positive thesis thus expands on the extra-perceptual and suggests that the locus of certain effects of interest is not perceptual. It is argued accordingly that although the phenomenon of suggested hallucination is not a sham, it still doesn't even abductively follow that subjects experience genuine perceptual hallucinations. Nor does it follow that they are genuinely deluded. A more plausible alternative is that it only feels to them as if the suggested state of affairs really pertained.

Now the idea that cognition may penetrate perception-related affect rather than perceptual experience proper or perception-related cognition is itself not new. However, in the cognitive-penetrability debate, the proposal has only been applied so far to apparent cases of perceptual learning and hence diachronic cognitive penetration (Dokic & Martin, 2015). On the other hand, in the hypnosis literature, the argumentation of contemporary proponents of the view seems more consistent with the causal priority of affect than the non-occurrence of positive perceptual hallucination (Woody & Szechtman, 2000a, 2007). In any case, in addition to an extended review and rebuttal of traditional skeptical, hallucination, and delusion accounts of suggested hallucination, an important contribution to this topic is an all-out defense of an affective interpretation of a possibly strongest candidate of synchronic cognitive penetration.

Here's how the discussion will proceed. *Chapter 2* elaborates in more detail on some key issues of cognitive penetrability. The reader is first to be acquainted with the development of the concept within a computational–functional framework. The discussion then turns to how theses of cognitive impenetrability within this framework bear on issues of cognitive penetrability of perceptual experience. It is argued that, contrary to a common exegetical mishap, all participants of the debate have a theoretical stake in whether perceptual experience of low-level features is cognitively penetrable. The next section inquires about which attitudes, states, or representations may qualify as proper cognitive penetrators, and how this question relates to the requirement that relevant effects should semantically (logically, rationally) cohere with the content of the source cognitive state. Having thus touched on issues relating to perceptual locus, cognitive source, and semantic coherence, the latter part of the chapter focuses on directness, that is, the requirement that relevant effects should not be mediated by changes in the stimuli, the sensory organs, or attention. It is concluded that although debates regarding the possible effects of attention are only likely to strengthen, the prima facie best candidate for an intermediary of cognitive penetrability is imagination. This shall lay the groundwork for subsequent chapters, which show why the candidate still fails.

Chapter 3 is an extended even if not exhaustive analysis of the face–race lightness illusion that has received perhaps the most attention from philosophers engaged in cognitive-penetrability debates in the past few years. After a brief summary of the initial findings, the discussion focuses on a prima facie most powerful argument according to which, in a critical experiment, changes in perceived lightness of racially ambiguous faces were a function of lexical labeling, and hence the effect was most plausibly mediated by categorization at the cognitive level. The chapter argues that this reasoning is flawed, and the assumptions on which it relies are questionable on both empirical and theoretical grounds. A low-level explanation of

the phenomena is accordingly proposed, which is arguably both empirically more plausible and abductively preferable to a cognitive-penetration account.

Chapter 4 engages in-depth with the phenomenon that may well be the strongest candidate for cognitive penetration: suggested hallucination. The first part of the chapter surveys evidence and arguments against traditional skeptical accounts of the phenomenon. On these accounts, suggestion effects can be explained without residue by factors such as simulation and compliance. Much of the latter part of the chapter dedicates to arguing that not only does it not deductively follow from debunking traditional skepticism, but nor is it even abductively plausible, that subjects experience relevant distortions in perceptual experience. Although there is merit to the idea that at least some effects may be mediated by cognitive delusions, it is ultimately granted to the mainstream view that not all phenomena afford a strictly cognitive interpretation. Thus the proposal that various relevant effects may be carried by penetration of affect, rather than penetration of perception or cognition.

Chapter 5 concludes the discussion with a short summary of the main conclusions and their theoretical and practical implications.

2. Cognitive Impenetrability

2.1 Background

The terms "cognitive penetration," "cognitive penetrability," and "cognitive impenetrability" were coined by Zenon W. Pylyshyn some three–four decades ago (1979a, 1979b, 1980, 1981, 1984). He introduced the notion of cognitive penetration in particular in the context of a famous debate in psychology about whether mental images are more akin to pictures or linguistic expressions; or, more specifically, whether the representational format of their medium is better construed as "analogue" (holistic) or "propositional" (structured). The analogue–propositional distinction, as Pylyshyn understood it, is as follows.

A representation is analogue if generalizations about certain determinants of the behavior of a system that the representation is supposed to explain can be sufficiently captured by appeal to intrinsic properties such as biological or physical properties of the system itself. This corresponds to the hardware level in computational models of the mind. In contrast, a representation is propositional if appropriate explanation of particular system regularities can be more successfully cast by appeal to explicit representations of, for example, grammatical rules, beliefs, and goals. This corresponds to the software level in computational frameworks.

With this distinction in place, Pylyshyn's aim was to show through a series of experiments that, despite its prima facie plausibility, mental images are unlikely analogue and hence more plausibly propositional. For, he argued,

> many of the phenomena that appear to be reasonable candidates for an analogue explanation show . . . "cognitive penetration"; that is, the phenomena can be critically influenced by cognitive factors such as the subjects' beliefs and interpretations. Whenever that is the case, we conclude that the whole

> phenomenon cannot be explained by a single analogue mechanism (although analogue subprocesses may still be involved if they resist cognitive penetration). (1979a, p. 21)

Similarly:

> time and again we find that how the [candidate] analogue behaves is determined primarily by factors extrinsic to the [candidate] analogue itself, notably by what subjects believe about what is being presented, how they interpret the stimulus, or what they intend by certain actions. This dependence of the behavior of a hypothetical operation upon cognitive factors is . . . cognitive penetration. (1979b, pp. 385–386)

The particularities of the imagery debate need not concern us here. What is relevant is that Pylyshyn ultimately construed the "cognitive impenetrability condition" as a methodological criterion for providing an empirical ground for a principled distinction between the fixed "functional architecture" of the mind on the one hand, and cognition on the other hand. Functions of fixed architecture are then cognitively impenetrable in that "they cannot be influenced by such purely cognitive factors as goals, beliefs, inferences, tacit knowledge and so on" (1980, p. 111). As such, they can be considered implicit or intrinsic, and may thus be appropriately described as merely instantiating biological or physical properties. In contrast, cognitive functions are penetrable by and hence susceptible to the influence of beliefs and goals and other purely cognitive factors. As such, they can be considered explicit or algorithmic, and thus require the positing of internal representations.

Hence, understood as a computational–functional demarcation criterion of cognition, if a function is cognitively penetrable, then it "must be explained, at least in part, by reference to computational cognitive processes whose behavior is governed by goals, beliefs, and tacit

knowledge" (1981, p. 16). As some have pointed out, though, it cannot suffice for cognitive penetrability that a function is susceptible to cognitive influences. If it did, then heart rate, blood pressure, and many other merely biological (/physical) functions would also qualify (cf. Stokes, 2013). Thus the criterion of semantic or logical or rational coherence. For of course the issue is not merely whether some or another function may be causally influenced by cognition. The crucial question is whether the regularities in causation are more successfully captured at an explanatory level that makes explicit reference to represented content. And for that to be the case, the receiver or target of any relevant cognitive influence must itself be susceptible to such content; indeed, it must be susceptible to the influence in virtue of the very content.

A more precise computational characterization of cognitive penetrability is then:

> if a certain behavior pattern (or input-output function) can be altered in a way that is rationally connected with the meaning of certain inputs (i.e., what they refer to, as opposed to their physical properties alone), then the explanation of that function must appeal to operations upon symbolic representations such as beliefs and goals: It must, in other words, contain rule-governed cognitive or computational processes. A function that is alterable in this particular way is said to be cognitively penetrable. (Pylyshyn, 1981, p. 21)

A somewhat simpler and more recent, most widely cited formulation is:

> if a system is cognitively penetrable then the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows. (1999a, p. 343)

It follows that if a function is not semantically (logically, rationally) sensitive to one's goals and beliefs, then it is cognitively impenetrable, and hence no appeal is needed to rule-governed cognitive or computational processes. Since fixed-architectural functions are not susceptible to such penetration by cognition, it thus also follows that "processes carried out in the functional architecture are processes whose behavior requires no explanation in terms of semantic regularities—that is, in terms of rules and representations" (1984, p. 131).

Now paralleling the distinctions between analogue and propositional, or holistic and structured, representational form; and fixed-architectural and cognitive, or merely-physical-or-biological and symbolic-algorithmic, function or process; Pylyshyn (1999a, 2003) later also proposed a distinction between "early vision" and "late vision"; which in turn closely parallels Fodor's (1983, 1985, 2001) distinction between modular and nonmodular perceptual processes. A common central idea is that early/modular perceptual processing is cognitively impenetrable, while late/nonmodular processing is cognitively penetrable. Of course, few if any would deny that at least a small portion of perceptual processing is relevantly encapsulated from cognition (cf. Noë & Thompson, 1999; Prinz, 2006). What makes the respective theses substantive is that on these accounts, a rather significant part of perceptual processing is cognitively impenetrable.

How significant is significant and in what sense? The following section discusses two relevant considerations: first, the level at which candidate processes deliver their outputs; and, second, whether such outputs are phenomenally conscious or at least determining of the phenomenal character of perceptual experience. Hierarchical level and phenomenal character are logically independent determinables of perceptual representations. So it is at least not inconceivable that perceptual experience may be cognitively penetrable despite relevant encapsulation of even relatively high-level perceptual processes. Contrary to a common

(mis)interpretation, though, arguably, no one actually holds such a view. So everyone has a theoretical stake in whether or which or how perceptual properties are penetrable.

Of course, insofar as perception is to be distinguished from perceptual belief and judgment, penetration of the latter does not qualify as cognitive penetration of perception. So the locus of penetration must be genuinely perceptual. The third section addresses in turn the requirement that the source of the effect must be genuinely cognitive, as well as the criterion of semantic coherence. The fourth section focuses on a final requirement, namely, that the effects should be direct at least in the sense that they are not mediated by changes in the stimuli, the state of the sensory organs, or the locus of attention.

Along the way, we shall distinguish, among other notions, between computational and experience-focused theses of cognitive impenetrability; low-level and high-level perceptual properties; top-down and side-to-side effects; intra-perceptual and extra-perceptual influences; attitudes and concepts and states; beliefs and aliefs; concepts and categories; synchronic and diachronic penetration; direct and indirect effects; and early and late attention. This will lay the conceptual and theoretical groundwork for addressing some purportedly most convincing cases of cognitive penetration in the following chapters.

2.2 Perception

2.2.1 Levels of output

Let us backtrack and refocus on the question: in what sense and to what extent are traditional computational–functional theses of cognitive impenetrability theoretically significant? This question relates to the requirement that the locus of cognitive penetration should be genuinely perceptual. Recall that according to traditional theses, early/modular perceptual processes are cognitively impenetrable, whereas late/nonmodular perceptual processes are cognitively

penetrable. A purported relevance of this distinction is that a significant portion of perceptual processing is supposedly of the first sort.

Now "significant portion" should not be understood in a merely quantitative sense: theses of cognitive impenetrability also bear on qualitative issues regarding the kinds or properties of representations that might constitute or be determined by the outputs of encapsulated systems. For traditional computational theories of perception posit a hierarchy of representations and processes, whereby each successive stage of processing encodes even higher-level properties or categories in terms of abstractness (or many-to-one mapping) relative to the initial inputs provided by sensory transducers.

A paradigm example is David Marr's (1982/2010) three-stage model of visual processing. On this model, early vision first extracts from perceived intensities a "primal sketch" of edges, bars, ends, curves, boundaries, etc. The second stage encodes a "2½-D sketch" of local surface orientations and depth and discontinuities along viewer-centered coordinates. The final stage then delivers a "3-D sketch" that describes a hierarchy of shapes and their organization in terms of surface and volumetric primitives in an object-centered way.

A crucial question for computational–functional theses of cognitive impenetrability, then, is at what processing stage or representational level encapsulated perceptual systems may deliver their outputs. Fodor (1983) and Pylyshyn (1999a, 2003) are clearly both committed to that, at a minimum, early visual processing of classic psychophysical properties such as color, surface shape, and three-dimensional depth—the latter of which correspond to Marr's 2½-D sketch—is relevantly encapsulated from cognition. Although encoding such features already requires quite extensive processing, in terms of the interlevels of representational hierarchies, they are still considered low-level properties (cf. Briscoe, 2015; Fodor, 1983; Helton, 2016).

2. Cognitive Impenetrability

Since Fodor wants to extract as much juice out of encapsulated processes as theoretically possible, he proposes accordingly that perceptual modules typically deliver representations at a much higher level: namely, at the level of basic categories such as *dog* and *table* (cf. Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In similar vein, Pylyshyn (1999a, 1999b) emphasizes that it is a real empirical possibility that early vision also encodes for example non-sensory properties such as causal relations, primitive affective properties like dangerousness, or certain functional properties such as Gibsonian affordances (e.g., pushability of a button or turnability of a knob). Some current defenders of cognitive impenetrability (e.g., Firestone & Scholl, 2016b; Mandelbaum, 2017) argue that such possibilities are indeed actualities, as the list of candidate high-level properties that might be extracted in modular fashion is ever further extended:

> Perception may be immune from cognitive influence, but it nevertheless traffics in a truly fascinating and impressively rich array of seemingly higher-level properties—including not just lower-level features such as color, motion, and orientation, but also causality (Scholl & Nakayama 2002), animacy (Gao et al. 2010), persistence (Scholl 2007), explanation (Firestone & Scholl 2014a), history (Chen & Scholl 2016), prediction (Turk-Browne et al. 2005), rationality (Gao & Scholl 2011), and even aesthetics (Chen & Scholl 2014). (Firestone & Scholl, 2016b, p. 55)

Hence, spelled out in terms of levels or hierarchies of perceptual representation and processing, computational–functional theses of cognitive impenetrability have far-reaching implications for theories of mental architecture. However, it is widely thought that such theories have little if any relevance for philosophical debates concerned with the nature of perceptual experience, the justificatory role of perception, or the theory-ladenness of

observation. For, it is argued, what concerns such debates is whether or to what extent or in what way the phenomenal content or character of perceptual experience is cognitively penetrable. Accordingly, it is often pointed out, even if early perceptual processing is cognitively impenetrable, insofar as later perceptual processing is not, perceptual experience may still be cognitively penetrated (e.g., Macpherson, 2012, 2015; Siegel, 2012, 2015; Stokes, 2013, 2015; Teufel & Nanay, 2017).

The next subsection presents three models and their possible combinations regarding the compatibility of process-impenetrability and conscious-penetrability. The subsection after that then argues that, contrary to a currently widely shared view, no actual thesis of process-impenetrability is theoretically neutral as to the conscious-penetrability of low-level perceptual features. Accordingly, all defenders run arguments for conscious-impenetrability on the assumption that such arguments provide abductive support for process-impenetrability. This makes sense given the assumed teleology of perception to provide cognition with quick and reliable information about the current state of affairs in one's immediate surroundings. But let us not run too much ahead. First things first: models of logical compatibility.

2.2.2 Process-impenetrability and conscious-penetrability: Logical compatibility

We have seen in the previous subsection that computational–functional theses of cognitive impenetrability propose that the extent and level of encapsulation of early/modular processing is both quantitatively and qualitatively significant. However, as many note, it is still a logical possibility that perceptual experience may be cognitively penetrated. A weak reading of this claim is that whatever features remain unprocessed or at least underprocessed by encapsulated systems may still be cognitively penetrated. For example, say early vision encodes at most only low-level properties such as color and surface shape. Assume further that late perceptual processes can encode kind properties such as being a pine tree. Assume still further that

perception of such properties can involve distinctive (high-level) perceptual phenomenology. This set of assumptions allows that perceptual experience may be cognitively penetrable even if the phenomenal character of low-level perceptual features is not (cf. Siegel, 2012).

So one possibility is that there is a tradeoff between process-impenetrability and conscious-penetrability. This is uncontroversial and theoretically unthreatening for defenders of cognitive impenetrability—what do they care if properties they have nothing to say about are penetrable? Yet there are alternative readings of the compatibility claim that may be more worrisome. For example, on a stronger reading, even if encapsulated perceptual processing is a determinant of the phenomenal character of all consciously perceivable properties, and even if only low-level properties can be consciously perceived, in at least some conditions and for at least some properties, other determinants may also relevantly impact perceptual phenomenology.

For example, say one is shown an orange–red colored heart shape. Since most heart shapes in our culture are typically red, perception of the shown heart shape may perhaps activate a belief with according content. In turn, this belief may give rise to imagery of a red heart shape. Now assume that this mental image is phenomenally mixed or fused with or superimposed on the percept of the orange–red heart shape. Some propose that this can result in color experience as of a hue somewhere between the actual orange–red and pure red, and so perceptual experience is at least indirectly penetrable by cognition (e.g., Macpherson, 2012). We shall discuss in subsequent chapters why this is empirically unlikely. The present point here is merely that one way conscious-penetrability and process-impenetrability may be compatible is if encapsulated processing only constrains the manner or extent of penetration (cf. also Macpherson, 2015).

On an even stronger reading of the compatibility claim, at least some if not all of the (potential) phenomenal consequences of encapsulated processing may in effect be trumped. For example, as we shall see in *Chapter 4*, some hold that, evidence for relatively intact early perceptual processing notwithstanding, at least some subjects may perceive even a blue shape as if red under hypnotic suggestion for color hallucination (e.g., Kallio & Koivisto, 2016). Even if such claims often seem suggestive of a most extreme possibility whereby process-impenetrability places no limits on conscious-penetrability whatsoever, it is at unclear that anyone actually defends this view. Accordingly, it bears emphasis that weaker versions of the trumping view are compatible with weaker versions of the tradeoff view and the mixing view. So for example a proponent may allow that some perceptual properties are not consciously penetrable (tradeoff), and that the manner or extent of penetrability is still constrained for some other conscious properties (mixing), as long as the phenomenal consequences of encapsulated processing of still further features may still be fully prevented or overridden (trumping).

The burden of the proceeding two chapters is to cast doubt on any version of the mixing view or the trumping view in relation to low-level perceptual properties. The next subsection argues accordingly that all actual theses of cognitive impenetrability—including those whose prime focus is on mental architecture—have a theoretical stake in whether low-level perceptual experience is penetrable. So even if any version of process-impenetrability is in principle compatible with any version of conscious-penetrability, for defenders of impenetrability, the abductive cutting line of the tradeoff is high enough as to include all relevant low-level features.

2.2.3 Process-impenetrability and conscious-penetrability: Theoretical compatibility

We have so far covered this: Computational–functional theses of cognitive impenetrability posit that a significant portion of perceptual processing is encapsulated from cognition. Although there is some variation among such theses as to just how high up the processing

ladder they are willing to bet respective outputs are delivered, what is definitely common ground is that classic psychophysical properties such as color and viewer-centered shape are included. Yet many philosophers are unmoved. Why? For starters, because cognitive penetrability of perceptual experience is at least logically compatible with the encapsulation from cognition of early perceptual processes.

We may now add that, even more importantly, many also hold that cognitive penetrability of perceptual experience is in fact theoretically consistent with at least Pylyshyn's view. For, it is variably noted, his thesis regarding early vision is "not about perceptual experience but about the subpersonal representational outputs of brain processing mechanisms" (Macpherson, 2012, p. 27). Note that it doesn't follow that Pylyshyn is neutral about conscious-penetrability. For example, he may (as he does) appeal to examples of conscious-impenetrability to strengthen his case for processing-impenetrability. We shall come back to this point later. For now, let it suffice that many hold that he at least allows for conscious-penetrability. Indeed, some take it that conscious-penetrability in effect follows from his posit that late vision and hence vision "as a whole" are cognitively penetrable. And still others read him as explicitly espousing that "conscious visual experience is cognitively saturated" (Noë & Thompson, 1999, p. 386).

The above interpretations thus feed into a dialectic an aim of which is to theoretically insulate debates about conscious-penetrability from debates about process-impenetrability. We shall presently challenge this dialectic, on account that there is in fact significant abductive tradeoff between theses of process-impenetrability and theses of conscious-penetrability. It shall be argued accordingly that no thesis of process-impenetrability actually assumes, or would even remain theoretically unthreatened by, cognitive penetrability of low-level

perceptual experience. On the defended understanding, this includes Pylyshyn's thesis. So interpretations such as the above are arguably exegetically mistaken.

2.2.3.1 Fodorian modularity and methodological pitfalls of empirical research

Let us start our inquiry of what defenders of process-impenetrability think (if anything) about conscious-penetrability with a quick survey of the views of the arguably most important defenders besides Pylyshyn. The most prominent and outspoken philosopher is of course Fodor (1983, 1984, 1985, 1988, 2001). The currently most vocal and discussed psychologists appear to be Firestone and Scholl (2014, 2015a, 2015b, 2015c, 2016a, 2016b, 2017). We shall briefly acquaint ourselves with their views in this order.

Fodor explicitly proposes that modular outputs are typically "phenomenologically accessible" (or "phenomenologically salient"). Perhaps a defender of conscious-penetrability may argue that this is good enough insofar as it allows that any output may be penetrated in at least some instance or other. But Fodor's point is not merely that most output tokens are conscious; it is that at least most output types are. Does it follow that some features of perceptual experience may still be penetrable—namely, those that correspond to the minority of output types that may be phenomenologically inaccessible? No, it doesn't. Some argue for example that visually guided action may be served by a cognitively impenetrable visuomotor system the outputs of which are always unconscious (Goodale & Milner, 1992; Pylyshyn, 1999a). It doesn't follow that there is any corresponding visuomotor feature of experience that is cognitively penetrable.

In any case, a most common interpretation of Fodor's view, including by defenders of cognitive penetrability of perceptual experience, is that "the output of the sensory modules determines, or is identical with, the content and phenomenal character of perceptual experience" (Macpherson, 2012, p. 29). Even if this characterization requires minor

qualification to accommodate "typically," it is not in question that the purported encapsulation of candidate perceptual modules has implications for perceptual experience. Hence, insofar as any understanding of cognitive impenetrability has philosophical relevance, the one presumed by Fodorian modularity certainly does.

Now consider Firestone and Scholl's view. As empirical psychologists, they "have the broader aim of evaluating the evidence for top-down effects on what we see as a whole—including visual processing and the conscious percepts it produces" (2016, p. 3). They claim that of the more than 175 papers published since 1995 that purport to provide positive evidence for relevant top-down effects of cognition on perception, not a single avoids all of an ultimately small number of methodological pitfalls.[1] Based on their own empirical demonstrations, they anticipate that once the pitfalls are avoided, all evidence for cognitive penetrability will turn out to be artefacts (cf. also Machery, 2015).

It bears emphasis in the present context that, like other defenders of cognitive impenetrability, Firestone and Scholl consider penetrability or lack thereof of conscious percepts to bear heavily on theories of mental architecture. They even allow in principle that a single methodologically sound empirical demonstration of conscious-penetration might warrant a true theoretical revolution in psychology. So, even if they do not dispute the conceptual independence of the respective notions of penetrability, they clearly hold that conscious-penetrability would at least seriously abductively undermine the case for process-impenetrability.

But say they did endorse the abductive compatibility of conscious-penetrability and process-impenetrability; they would still deny the abductive compatibility of conscious-impenetrability and process-penetrability. This meshes with the general view that, if true,

[1] A full list of the papers can be accessed from: http://perception.yale.edu/Brian/refGuides/TopDown.html

conscious-impenetrability provides strongest abductive support for process-impenetrability. Hence, not only are Firestone and Scholl in fact committed to cognitive impenetrability of both perceptual experience and early perceptual processing—their being so is also at least in part related to their theoretical presumption that these issues are not empirically unrelated.

2.2.3.2 Early vision

We have seen that two of the arguably three most crucial defenses of cognitive impenetrability are committed to both process- and conscious-impenetrability. And for no arbitrary reason, either: the parties agree that, on the one hand, a high reach of modular processing in terms of output level poses strong abductive pressure against cognitive penetrability of perceptual experience; and, on the other hand, a lack of methodologically sound empirical evidence for the latter is made best abductive sense of by encapsulation from cognition of relevant perceptual processes. So, for at least most prominent defenders of cognitive impenetrability, issues of processing and experience are not theoretically independent.

This brings us back to the question where Pylyshyn stands on these issues. Although few would deny that his thesis of cognitive impenetrability of early vision bears importantly on theories of mental architecture, recall that many still consider his views philosophically inconsequential. For, it is variably argued, his position is at least theoretically compatible if not entailing or even espousing of cognitive penetrability of perceptual experience. I disagree on both theoretical and exegetical grounds. Here are some considerations.

To start with, similarly to other defenders of cognitive impenetrability, Pylyshyn appeals to "a remarkable fact about the perceptual illusions that knowing about them does not make them disappear" (1999a, p. 344). This phenomenon is often referred to as the persistence of illusions. A common example is the Müller–Lyer illusion: even if one knows that the two horizontal lines are of equal length, one still sees the line with inverted arrowheads as longer

than the line with normal arrowheads. Pylyshyn emphasizes that "It is not only that the illusions are stubborn, in the way some people appear unwilling to change their minds in the face of contrary evidence" (ibid.). The crucial point is that "in these cases there is a very clear separation between what you see and what you know is actually there—what you believe"; so "it is simply impossible to make some things look to you the way you know they really are" (ibid.).

At a minimum, arguments from the persistence of illusions are intended to underscore that at least some of perception is relevantly encapsulated from at least some of cognition (cf. Fodor, 1983, 1985). Never mind whether they actually succeed in this (cf. Lupyan, 2015a). The upshot is that Pylyshyn's argumentation along such lines reveals a commitment to at least some form or extent of conscious-impenetrability. Macpherson (2012) takes notice, but she thinks this is paradoxical given Pylyshyn's insistence that his thesis is not about perceptual experience but "merely about subpersonal brain processing and architecture" (p. 29). But there is no paradox here—or at least the paradox is not that hard to resolve.

Firstly, strictly speaking, Pylyshyn's thesis is not about brain processing, even if he concurs that the processes of interest are realized by the brain. He explicitly emphasizes that

> early vision is defined functionally. The neuroanatomical locus of early vision, as we understand the term . . . , is not known with any precision. However, its functional (psychophysical) properties have been articulated with some degree of detail over the years . . . . As various people have pointed out (e.g., Blake 1995), such analysis is often a prerequisite to subsequent neuroanatomical mapping. (1999a, p. 344)

Thus, pace a common misconception, early visual processing as functionally specified should not be mistaken for early visual processing as neuroanatomically specified (e.g.,

primary visual cortex; V1). Two considerations highlight the theoretical significance of this point. One is that, in a strict sense, as Pylyshyn understands the term, it is analytic that early vision is cognitively impenetrable. So for him the question is not whether early vision is cognitively impenetrable, but whether it exists to start with, and if so, which processes are included. The purported significance of his thesis is thus that early vision includes much more than just basic perceptual sensors. This leaves it an open empirical question whether or to what extent early vision in this functional sense overlaps with early visual areas as neuroanatomically specified.

A related consideration is that, like any computationally independent system, early vision may contain its own proprietary, functionally local memory. This will bear importantly in the next chapter when we discuss whether changes in color experience mediated by stored color–shape associations imply that cognitive penetration has occurred. (Spoiler: no, it doesn't.) The present point is that just because proprietary storage is functionally local, it doesn't follow that it cannot be implemented as a subset of global memory, the exact neuroanatomic realization of which is yet another empirical question about which Pylyshyn has nothing to say (cf. also Fodor, 1983).

So is it paradoxical that Pylyshyn argues for cognitive impenetrability of perceptual experience (for example from the persistence of illusions) while defending a thesis about subpersonal brain processes? An upshot so far is that either his thesis is misconstrued, if it is taken to be about specific brain processes qua brain processes; or its being about brain processes is trivial, to the extent that all mental processes are presumed to be at least token identical with brain processes. So if there is still a paradox around the corner, it must be driven by an assumption that the processes of interest are subpersonal, whereby their being so is not derived from their being brain processes.

2. Cognitive Impenetrability

However, the assumption that the processes of interest are subpersonal involves an analogous dilemma to the above. For, on the one hand, there is a trivial sense in which all computationally independent or functionally local systems are subpersonal; it doesn't follow that the output of such systems must be unconscious. On the other hand, it is exegetically mistaken that Pylyshyn endorses or implies that all output of early vision are unconscious. So, in this respect as well, either his thesis is misconstrued (if "subpersonal" is understood as "unconscious"), or its being about subpersonal processes is computationally–functionally trivial and hence non-entailing of unconsciousness.

Why do some insist that, despite his arguments for at least limited conscious-impenetrability, Pylyshyn in effect implies that all output of early vision are unconscious and hence subpersonal in a non-trivial sense? A typical argument goes as follows. First, Pylyshyn is assumed to grant that "conscious visual experience is cognitively saturated" (Noë & Thompson, 1999, p. 386). It is then inferred that "If the output of early vision were conscious, it would be cognitively penetrable" (ibid.). However, the output of early vision is by posit cognitively impenetrable. So "it would seem that the output of early vision must be inaccessible to consciousness" (ibid.). Yet if this is so, it is concluded, then "the output of early vision is inaccessible to the subject, and therefore cannot be an object of evaluation" (p. 387). Which is another way of saying that the respective output is subpersonal.

The argument is possibly invalid, but most certainly unsound. Regarding its starting premise, we will soon see why it is exegetically imprecise to attribute to Pylyshyn the view that visual experience in a strict sense is "cognitively saturated." But even if he held this view, recall that there may still be a tradeoff between the features that can be processed by early vision and the features that might be consciously penetrated. Accordingly, Pylyshyn can in principle allow that the phenomenal character of certain high-level visual properties is

cognitively penetrable (and hence that respective high-level perceptual phenomenology exists), meanwhile upholding that all other conscious perceptual features are cognitively impenetrable.

So even if Pylyshyn maintained that conscious visual experience is cognitively penetrable—which is debatable, as argued below—it still wouldn't follow that the output of early vision must be subpersonal qua unconscious. This brings us to the second horn of the dilemma that either Pylyshyn's thesis is misconstrued, or its focus on the subpersonal is trivial and hence non-entailing of unconscious processing. For Pylyshyn of course agrees that the output of early vision is subpersonal in a computational sense. What he denies is that it follows that all such output are unconscious and hence inaccessible to the subject. In his words:

> To equate the distinction between conscious and unconscious with the distinction between personal and subpersonal is to invite such nonsensical conclusions as "the output of early vision is inaccessible to the subject" (Noë & Thompson), which by parity of reasoning would lead one to believe that *nothing* computed by the brain is accessible to the subject. My view is that some things are and some things are not . . . . (1999b, p. 407; emphasis in original)

Recall from our earlier discussion of Fodorian modularity that even if some modular output are unconscious, it doesn't follow that any feature of perceptual experience is cognitively penetrable. It is common ground in modern psychology that much of what goes on in the mind is inaccessible to consciousness. This seems true whether or not perceptual experience is cognitively penetrable. So by implying that some (but not all) output of early vision are inaccessible to the subject, Pylyshyn is not thereby suggesting that some features of perceptual experience are cognitively penetrable. Rather, a crucial upshot of his reasoning is that, insofar as the subpersonal cuts across the conscious–unconscious divide, cognitive impenetrability of early vision has implications for *both* conscious and unconscious perception.

2. Cognitive Impenetrability

Why, then, is Pylyshyn interpreted as if paradoxically also avowing "cognitive saturation" of visual experience? The prime sources of exegetical confusion appear to be the following claims:

> [W]hat is commonly referred to as "visual perception" is potentially determined by the entire cognitive system. (1999a, p. 342)

> [V]ision as a whole *is* cognitively penetrable. (p. 344; emphasis in original)

> [W]hat we see—the content of our phenomenological experience—is the world as we visually apprehend and know it; it is not the output of the visual system itself. (p. 362)

Taken out of context, these remarks may seem to suggest that Pylyshyn is a full-fledged, card-carrying, dyed-in-the-wool believer of cognitive penetrability of visual experience. However, to start with, notice the quotation marks around "visual perception" in the first claim. This should already raise suspicion of pragmatic implication that the referent is not visual perception in a strict sense. Or consider the claim that vision as a whole is cognitively penetrable. This calls to mind Fodor's reasoning regarding why, even if modular perceptual processes are informationally encapsulated,[2] perception as such is not:

> Nobody doubts that the information that input systems provide must somehow be reconciled with the subject's background knowledge. We sometimes know that the world can't really be the way that it looks, and such cases may legitimately be described as the correction of input analyses by top-down

[2] On some interpretations, informational encapsulation is stronger than cognitive impenetrability in that the former (but not the latter) also disallows or constrains side-to-side influences between (in addition to top-down influences from cognition on) perceptual modules. We may bypass whether this interpretation is correct, and whether or in what way Fodor's and Pylyshyn's views may differ with respect to side-to-side or inter-modular influences. Let it suffice that they both agree that though early/modular visual/perceptual processes are cognitively impenetrable, late/modular visual/perceptual processes and hence vision/perception as such are/is cognitively penetrable.

> information flow. (This, ultimately, is the reason for refusing to identify input analysis with perception. The point of perception is the fixation of belief, and the fixation of belief is a *conservative* process—one that is sensitive, in a variety of ways, to what the perceiver already knows. Input analysis may be informationally encapsulated, but perception surely is not.) (1983, p. 73; emphasis in original)

Accordingly, when Pylyshyn argues that vision as a whole is cognitively penetrable, he is not predicating about the phenomenal character of visual experience. What he more plausibly has in mind is that the fixation of visual belief is cognitively penetrable. Consider his examples of recognizing paintings as genuine Rembrandts and identifying tumors in medical X-rays. Such recognition and identification surely play a role in the fixation of respective beliefs about what is seen. They also most plausibly require accessing background knowledge or beliefs involving cognitive concepts or categories such as GENUINE, REMBRANDT, INDICATION, and TUMOR. Thus, if vision as a whole is understood to include all such processes underserving the fixation of perceptual belief, then of course vision in this sense is cognitively penetrable.

So there is a loose and general sense in which one may "see" a painting as a genuine Rembrandt, or "see" a tumor in a medical X-ray. It doesn't follow, however, that properties such as *genuine*, *Rembrandt*, *indication*, and *tumor* can be encoded by any visual system, or that such properties can be associated with distinctive high-level perceptual phenomenology. Nor does it follow that the cognitive encoding of such features can lead to any relevant change in the perceptual encoding or experience of any other visual feature. Which is to say, cognitive penetrability of vision "as a whole" has no implication for either visual processing or visual experience of either high-level or low-level visual properties. What it has implications for is perceptual belief.

Yet, in the sense that it matters for related debates, penetration of perceptual belief does not qualify as penetration of perception. Thus the reason for having defined cognitive penetration in the introductory chapter as requiring (among other conditions) that the locus of the effect must be genuinely perceptual. For any defender of cognitive impenetrability, this is a most crucial point. So there is no paradox in acknowledging cognitive penetrability of perception in a loose and general sense (which includes perceptual belief), meanwhile denying cognitive penetrability of perception in a strict and specific sense (which excludes perceptual belief). For Pylyshyn, as for other skeptics of cognitive penetrability, arguably, this denial extends to perceptual experience of at least standard low-level properties.

So how to explain (away) the earlier quoted claim that "what we see—the content of our phenomenological experience—is the world as we visually apprehend and know it; it is not the output of the visual system itself" (1999a, p. 362)? Firstly, "content of phenomenological experience" should not be understood as "phenomenal content" or "phenomenal character of conscious experience." A more plausible reading is "representational content of phenomenal experience." So the claim concerns not the what-it-is-likeness of visual experience characterized in terms of phenomenal properties, but the features or categories that may be represented by such experience.

Second, for Pylyshyn, "visual apprehension" involves all processes involved in understanding the world through the visual modality, including attention, motor control of the eyes and head, problem solving, and inference. An end result of these processes is the fixation of perceptual belief. (Another is visually guided action.) Thus, in this context, the claim that "what we see . . . is not the output of the visual system itself" should be understood in the loose and general sense in which vision "as a whole" includes more than just the output of the visual system.

2. Cognitive Impenetrability

Recall the examples of recognizing a painting as a genuine Rembrandt, or identifying a blob in a medical X-ray as indicating a tumor. We already discussed that, insofar as such cases involve fixation of visual belief, (for Pylyshyn) they have no logical–theoretical bearing on the phenomenal character of visual experience. However, and this is the present point, just as there is a loose and general sense in which one may "see" that a painting is a genuine Rembrandt or that a blob in a medical X-ray indicates a tumor, so there is a loose and general sense in which such cognitive content relating to recognition and identification figures into the representational content of visual experience. So if the respective perceptual beliefs differed (e.g., the painting was considered a fake, or the blob an artefact of the X-ray machine), then so would the content of visual experience. In this sense, of course the content of visual experience is not the output of early vision itself. Yet, as such, this has nothing to do with phenomenology.[3]

With that said, two last points bear mention. First, despite all of the above, strictly speaking, it may indeed be true that the output of early vision our unconscious. One possible explanation of this may be that their content still require enrichment, for example by some higher-level perceptual content. Another (compatible) possibility is that their content may need to be encoded into a different representational format to allow for conscious phenomenology. In either case, it doesn't follow—and Pylyshyn does not seem to allow—that the phenomenal character of relevant visual features is cognitively penetrable.

Finally, a general theoretical consideration is why Pylyshyn (or anyone else, for that matter) even cares whether early vision is cognitively impenetrable. An apparent answer is that he is essentially interested in how vision may solve the evolutionary computational problem of

[3] Perhaps content such as *that this painting is a genuine Rembrandt* or *that this blob in the X-ray indicates a tumor* may constitute or determine cognitive phenomenology? This would suggest a more straightforward sense in which respective content might figure in phenomenological experience. However, insofar as cognitive phenomenology is construed as non-reducible to perceptual phenomenology by definition (cf. Prinz, 2011), such constitution or determination of conscious experience would not bear relevantly on perceptual phenomenology. Accordingly, issues regarding cognitive phenomenology are at least oblique to issues of cognitive penetrability.

delivering relatively accurate representations of one's current immediate surroundings in a relatively quick fashion. Accuracy is an issue insofar as behavior relies heavily on cognition, which in turn relies heavily on the information provided by conscious percepts. If perceptual experience were cognitively penetrable, the information it provided might thus be unreliable, which may in turn have negative evolutionary consequences.

Of course, this is not to deny that cognition might in certain cases positively contribute to the accuracy of representations. The persistence of illusions in the face of correct contrary beliefs demonstrates just such a case. However, there is a plausible tradeoff between the amount of information vision may access on the one hand, and the speed with which it may deliver its output on the other hand. A computational optimization of this speed–accuracy tradeoff may thus also require functional encapsulation of vision from cognition. If perceptual experience were penetrable, then this advantage in speed would also be undone. Thus, for theoreticians like Pylyshyn, computational considerations regarding functional decomposition of tasks to secure fast and reliable perceptual processing only provide further abductive reason to deny cognitive penetrability of conscious percepts (Pylyshyn, 1980, 1984, 1999a, 2003; cf. also Fodor, 1983; Stokes, 2013).

Hence, to conclude, by defending cognitive impenetrability of perceptual experience, Pylyshyn is not contradicting himself. For even if all output of early vision are inaccessible to consciousness qua vehicles of representation, the content they encode may still determine or constrain relevant features of phenomenology. Proponents of the mixing model of conscious-penetrability interpret Pylyshyn to only adhere to partial constraint (e.g., Macpherson, 2012, 2015). However, if the present exegesis and theoretical construal are correct, all things considered, the thesis is at least strongly suggestive of relevant determination of low-level perceptual experience. No theoretical neutrality, no paradox.

2.2.3.3 Summary

Let us take stock: It is widely held that, in contrast to the views of other prominent defenders of cognitive impenetrability, Pylyshyn's view is inconsequential for philosophical debates of conscious-penetrability. For, it is variably argued, his thesis is not only logically but also theoretically compatible, if not entailing or even espousing of all-out cognitive penetrability of visual experience. However, as we have seen, Pylyshyn's arguments such as related to the persistence of illusions and the subpersonal–unconscious distinction reveal a commitment on his part to at least partial conscious impenetrability. This commitment arguably covers at least all low-level visual properties such as viewer-centered shape and color.

Indeed, Pylyshyn may even allow for complete conscious-impenetrability. This is paradoxical only insofar as early vision is misinterpreted as non-determining of visual phenomenology. What appears to drive such misinterpretation is that functional–representational terms such as "vision as a whole" and "content of phenomenological experience" are exegetically mistaken for phenomenal notions such as "visual experience" and "phenomenal character of experience" (respectively). The upshot is that while Pylyshyn has a real stake in whether low-level visual features are consciously penetrable, he is at best non-committal about—rather than espousing or entailing or even suggesting of—cognitive penetrability of high-level perceptual experience.

Hence, contrary to a common but arguably mistaken view, there is in fact a theoretical tradeoff between theses of process-impenetrability and theses of conscious-penetrability. On the one hand, the greater the extent and the higher the level of impenetrable processing, the more plausible it starts to seem that at least some—indeed, that an even larger portion and an even higher level of—respective experience is also impenetrable. On the other hand, the greater the extent and the lower the level of consciously penetrable features, the more plausible it starts

to seem that at least some—indeed, that an even larger portion and an even lower level of—candidate processing is itself penetrable. So process-impenetrability puts at least some abductive pressure on theses of conscious-penetrability; and, vice versa, conscious-penetrability puts at least some abductive pressure on theses of process-impenetrability.

The theoretical tradeoff between conscious-impenetrability and process-penetrability is even more significant. For even if it is a logical possibility that certain candidate processes are penetrable but perceptual experience is not, no one actually holds (or is considered to hold) such a view. It is indeed implausible that whenever some candidate process is penetrated, perceptual experience is underserved by some other, impenetrable process; or the output of the penetrated process is itself subsequently penetrated so as to reverse the original effect of penetration; etc. Accordingly, even those who downplay the relative abductive pressure that conscious-penetrability puts on theses of process-impenetrability, do not deny that conscious-impenetrability constitutes great abductive pressure against theses of process-penetrability. Little wonder that defenders of process-impenetrability themselves so often appeal to conscious-impenetrability.

The general upshot is that irrespective of theoretical focus or framework, all theses of cognitive impenetrability have a stake in whether or how or which features of perceptual experience are cognitively penetrable. The discussion in subsequent chapters focuses on issues of conscious-penetrability accordingly. Thus also the reason that cognitive penetrability was defined in the introductory chapter as involving influences on perceptual experience. Since it is argued in subsequent chapters that even some strongest candidates of cognitive penetration of low-level experience are not real cases, if the argumentation is correct, it should thus cast corresponding doubt on process-penetrability as well. So, on the defended view, perception of low-level properties is cognitively impenetrable in both the functional and experiential sense.

This concludes the discussion on why a prime locus of interest of candidate penetration effects is perceptual experience, rather than merely preparatory processes or perceptual belief. In the remaining two sections, we shall cover some conceptual and theoretical ground relating to other key requirements of cognitive penetration. We shall inquire in particular about what may qualify as a genuinely cognitive penetrating attitude or concept or state, and how this question relates to the criterion of semantic–logical–rational coherence (*Section 2.3*); as well as in what sense the effects should be direct, and why imagery may be the best candidate intermediary of cognitive penetration (*Section 2.4*).

2.3 Cognition

In the previous section, we discussed some issues related to the requirement that the target of cognitive penetration of interest must be genuinely perceptual. As we saw, computational–functional theses specify the locus of interest in terms of processing types and levels, whereas experiential theses characterize the locus of interest in terms of phenomenal properties. On both understandings, a most important distinction is that between perception and perceptual belief (or judgment).[4] While there are lively debates about whether or to what extent or which features of perception are cognitively penetrable (cf., e.g., Raftopoulos & Zeimbekis, 2015), no one doubts that, at least in paradigm cases, perceptual belief is relevantly sensitive to the influence of other beliefs. This is not to suggest that there is agreement about the philosophical significance of cognitive penetrability of perceptual belief (cf., e.g., Lyons, 2011, 2015; Siegel, 2012, 2015). The common ground is merely that, qua beliefs, perceptual beliefs are cognitive; and, qua cognitive, they are paradigmatically cognitively penetrable.

[4] Some construe belief as a standing state or attitude, whereas judgment an occurrent act or event. We shall disregard such conceptual fineries. So perceptual belief may be understood in the following as either standing or occurrent.

2.3.1 Attitudes

Picking up the thread from the above, let us now focus on a second crucial requirement of cognitive penetration, namely, that the source of the effect must be genuinely cognitive. Recall that on a traditional computational–functional account, any process or representation that is sensitive in a semantically–logically–rationally coherent way to what a person believes is thereby cognitive (Pylyshyn, 1981, 1984, 1999a). The sufficiency of such sensitivity is generally understood to hold with respect to desires and hopes and other propositional attitudes as well. Accordingly, even if debates about cognitive penetrability most typically focus on beliefs as candidate penetrators, any relevantly cohering influence on perception by any propositional attitude would qualify as a case of cognitive penetration.

Recall for example (from the introduction) the thirsty desert nomad whose desire for water purportedly leads to visual experience as of a sheet of water. If desires could really influence experience in such content-relevant ways, that would provide strong evidence for cognitive penetrability of perception.

However, the proposed sufficiency condition of cognition presents a serious dilemma. To see its first horn, assume (arguably contrary to fact) that perception is cognitively penetrable. It follows on traditional accounts that it is relevantly sensitive to the influence of propositional attitudes. If such sensitivity is sufficient for cognition, then perception is itself cognitive. This may make theoretical sense if the contrast of interest is with fixed functional architecture, but it muddles for example the distinction between perception and perceptual belief. Although some would cherish such a consequence (e.g., Lupyan, 2015a), it is not unreasonable to demand that, for the purposes of debating cognitive penetrability, cognition should be characterized in a way that discounts perception (cf. Macpherson, 2015, 2017).

This leads to the second horn of the dilemma. For say it is rendered analytic that perception is that part of sensory-input processing that is not relevantly sensitive to cognitive content. This is in part how perceptual modules and early vision are defined. Accordingly, insensitivity to semantic content by definition would secure that perception is discounted as cognition. However, whereas it makes sense to ask whether perceptual modules or early vision exist, this is an unintelligible question about perception in the context of debating cognitive penetrability. Further, and most importantly, insofar as semantic sensitivity to cognitive content is a necessary condition of cognitive penetrability, it would analytically follow that perception is cognitively impenetrable. Which is, needless to say, quite the obstacle to any relevant debate.

Hence, a dilemma with treating semantic sensitivity to propositional content as sufficient for cognition is that the condition is too weak if it allows perception to qualify as cognitive, but too strong if it disallows perception from being cognitively penetrable. Should the condition be considered necessary, though? Some argue that the criterion thus construed is still too strong (e.g., Stokes, 2013), or at least they (re)interpret it to allow for effects whereby the source of cognitive penetration is merely some possessed or primed concept (e.g., Macpherson, 2012). A quick note on the latter view is in order.

2.3.2 Concepts

In the context of cognitive penetrability, a concept is understood to be primed if it is likely to be triggered or activated or involved in the formation of certain beliefs or desires (Macpherson, 2012). For example, it is argued that priming the concept RED makes it likely that one forms the belief about a shown orange–red heart shape *that this heart shape is red*.

Notice, however, that priming RED may also make it likely that one forms the belief *that this heart shape is not red*, or *that this heart shape is the complementary color of red*. Individual concepts and unstructured (or improperly structured) concept bundles are insensitive

to such semantic–logical differences as between positive vs. negative predication ("is red" vs. "is not red") and contradictories vs. contraries (red/not-red vs. red/green). In this respect, merely primed concepts cannot sustain semantic coherence.

We shall discuss in more detail some related issues and purportedly best examples of conceptual penetration of perceptual experience in the next chapter. The current upshot is that if all conceptual influences are to qualify as cognitive by default, it must be non-trivially specified what it is that distinguishes concepts qua cognitive categories from non-conceptual qua non-cognitive categories. It is important that various perceptual processes themselves involve categorization and respective top-down categorical influences within perception. So it doesn't in itself follow from some categorical influence on perception that cognitive penetration has occurred. For that to be the case, the condition that the source of the effect must be cognitive and hence extra-perceptual (and hence not intra-perceptual) must also be satisfied.

2.3.3 States

On one view, then, only propositional attitudes would qualify as genuine sources of cognitive penetration. On another view, concepts might also pass the bar. These views differ in at least two important respects: whether they require that the vehicles of penetration be appropriately structured, and whether the vehicles must themselves be sensitive to logical–rational properties. Propositional attitudes satisfy both conditions; concepts satisfy neither.

Could there be any hybrid candidate that satisfies one but not the other condition? As we saw in relation to concept priming, a lack of appropriate structure precludes sensitivity to logical–rational properties. However, it is at least conceptually compatible that a candidate may be appropriately structured yet still insensitive to logical–rational properties. Indeed, as Katalin Farkas and I took notice some years ago (Bitter & Farkas, 2012), such a candidate may be but

one subtype of a kind of mental state proposed not much earlier under the neologism “alief” (Gendler, 2008a, 2008b).

A paradigmatic alief is a mental state with associatively linked content that is representational, affective, and behavioral. A relevant feature of such purported states is that whether or not they are propositionally structured (they may come in either or both versions), they are insensitive to rational–logical properties of belief. For example, many people seem reluctant to put a piece of fake vomit to their mouth, even if they readily endorse that the fake vomit is made of rubber, and they unhesitatingly put a sink stopper of similar size and material to their mouth (Rozin, Millman, & Nemeroff, 1986). A possible explanation of this discrepancy is that although no one believes that the fake vomit is real or dangerous or to be avoided, due to certain associations based on prior experience, many still “alieve” so.

The crucial point is not that alief may be belief-discordant—many attitudes may also be so. For example, one can imagine Superman flying through the sky despite the belief that Superman does not exist. However, whereas belief is paradigmatically responsive to the call of rationality, alief is paradigmatically not. This is not to say that alief is irrational; it is simply arational. For it involves the activation of an associative chain regardless of what attitude one bears to the content activating these associations. Alief is thus a state and not an attitude.

Here’s a more elaborate example. Say Lois Lane comes to believe that Clark Kent is Superman. She then imagines flying through the sky with Superman, while focusing on this belief and its implication regarding with whom she imagines flying through the sky. Barring special cases, Lois Lane will thereby be imagining flying through the sky with Clark Kent. Such attitudinal contexts thus generally allow substitution of co-referential or coextensive expressions *salva veritate*. Yet not alief contexts.

2. Cognitive Impenetrability

An approximate example of the referential–extensional hyperopacity of aliefs is that subjects are more disinclined to consume a sugary drink if the container from which the sugar is added is labeled "not poison" rather than "safe to consume" (Rozin et al., 1986). Never mind whether these expressions are in fact coextensive. The crucial point is that even if a subject believes that they are, aliefs involving one expression will still be logically–rationally inconsequential for aliefs regarding the other (cf. Gendler, 2008a).

Hence, the moral is still, even if some mental states may resemble attitudes in that they involve the activation of propositionally structured representations, they may still differ from attitudes in that they are insensitive to logical–rational properties. Which is to say, it doesn't follow from a representation's being appropriately structured that it is cognitively penetrable in a traditional sense. Accordingly, if it is considered a necessary condition that all of cognition should at least in principle be cognitively penetrable—which allows at most individual anomalies and special cases—such states as aliefs would not qualify as cognitive proper.

Then again, note that it doesn't follow that alief-like states could not even in principle influence perception in a semantically coherent way. It is one thing to say that A is cognitively impenetrable by B (and hence A is not cognitive), and quite another that A cannot influence C in a relevantly coherent way. If cognitive impenetrability of some state or process precluded the possibility of its exerting any such influence on any other state or process, then cognitive impenetrability of perception would itself bar perception from informing cognition in a relevantly meaningful way. Yet perception most certainly does so.

Some argue accordingly that perception must be conceptual (e.g., McDowell, 1994/1996, 2006/2009). Assume that this is correct; it still doesn't follow that perception is cognitively penetrable. So perception may relevantly influence cognition even if cognition cannot relevantly influence perception. Mutatis mutandis, even if alief-like states do not heed

to semantic–logical–rational pressure, they may in principle still exert such pressure on perceptual states and processes.

Thus, even if they are not strictly speaking cognitive, there would seem to be at least some prima facie reason not to exclude alief-like states as potentially interesting vehicles of cognitive penetration. However, note that just as it doesn't follow from a top-down categorical influence on perception that the influence is cognitive, so it doesn't follow from penetration of perception by some alief-like state that the respective state is non-perceptual. This is not a merely logical point: as originally construed, alief need not be a sui generis mental kind for it to have explanatory utility in providing a unified account of seemingly disparate mental phenomena. So at least some aliefs may well be intra-perceptual.

Indeed, the very first example used to introduce the idea of alief involves a scenario in which stepping onto a visibly high transparent surface may induce feelings of vertigo despite one's belief that the surface is completely safe to step on (Gendler, 2008a). Yet recall that on some accounts it is a real empirical possibility if not demonstrated actuality that early visual processing may encode such primitive affective properties as dangerousness (e.g., Firestone & Scholl, 2016b; Pylyshyn, 1999a). So it is at least not out of the question that, upon stepping on a high transparent surface, encapsulated perceptual systems might output aliefs such as "Really high up, long way down. Not a safe place to be! Get off!!" (Gendler, 2008a, p. 635) If such output may in turn be fed back to lower levels of visual processing and/or relevantly influence visual experience, this would suggest that perception is intra-perceptually penetrable by alief. Surely these would be interesting effects—but they would not be cases of cognitive penetration.

So penetration by alief does not entail penetration by cognition, even if we drop the requirement that genuine sources of cognitive penetration must themselves be cognitively penetrable. This underscores that even if propositional structure and a potential to exert relevant

influences are both assumed to be necessary, they are still insufficient for candidate sources or vehicles to qualify as cognitive. Analogously to defenders of conceptual penetrability, then—who must still convince why candidate effects could not be mediated by intra-perceptual categories—possible defenders of penetration by alief would still have to argue on independent grounds that the effects of interest cannot be mediated by intra-perceptual aliefs. So there is no easy way (if any) of bypassing the issue of what it is about the content of some apparent penetrator that makes it plausibly cognitive.

Little wonder, then, that the debate is so fierce about whether any truly cognitive penetrator may in fact produce any relevant effect. The next chapter (*Chapter 3*) argues through a specific example accordingly that even a prima facie best candidate of cognitive penetration of color experience is more plausibly explained by intra-perceptual mechanisms. The chapter after that (*Chapter 4*) then explains through another example why it still doesn't follow from establishing that the source of a relevant effect is cognitive that either perceptual experience or perceptual belief must have been penetrated. In a nutshell, in at least some cases, a further plausibility is that cognition penetrates affect. So many considerations must be met in order to qualify either the source or the locus of a candidate effect as genuinely cognitive.

Before proceeding to the discussions of intra-perceptual penetration and penetration of affect, though, some comments are still due concerning the relation between cognition and perception that must be sustained for cognitive penetration of perception. We have already touched the semantic-coherence criterion. Another relational requirement is that the effect must be direct at least in that it is not mediated by changes in the stimuli, the sensory organs, or the allocation of attention. This is the main topic of the next section, which concludes the general theoretical discussion of cognitive penetrability.

2.4 Directness

Let us summarize: In the first section, we discussed how the notion of cognitive penetrability developed, starting from issues concerning the representational format of imagery (analogue vs. propositional), through issues of demarcating between functional architecture and cognition (fixed vs. rule-based functions), to issues of visual processing in particular (early vs. late) and perceptual processing in general (modular vs. nonmodular). In the second section, we then addressed the relation between computational–functional and experiential theses of cognitive (im)penetrability. We concluded that all actual theses have a theoretical stake at least in whether or to what extent or which aspects of low-level perceptual experience are cognitively (im)penetrable. A possible source of misunderstanding in this regard is that when defenders of process-impenetrability avow that perception as a whole is cognitively penetrable, they have in mind not perceptual experience but the totality of processes underserving the fixation belief. Since everyone agrees that perceptual belief is cognitively penetrable, the issue is thus whether perception in a more restricted sense is amenable to such influence.

The flip side of the requirement that the locus of an effect must be perceptual is that the source of the effect must be cognitive. The third section inquired accordingly about which mental kinds may qualify as cognitive. It is uncontroversial that paradigmatic propositional attitudes such as beliefs and desires do. Some also allow for example that primed concepts may function as the sources or vehicles of cognitive penetration. They drop or (re)interpret the criterion of semantic or logical or rational coherence accordingly. This leaves logical room for alief-like penetrators as well, which may be propositionally structured yet still insensitive to rational–logical properties. However, if there is no prima facie reason to exclude that perception may be cognitively penetrable, then certainly there is no prima facie reason to exclude that it may be conceptual or alief-like. So just as debates about cognitive penetrability

are closely linked with debates about which processes and features of experience qualify as genuinely perceptual, so they are closely linked with debates about which features should count as at least paradigmatic markers of cognition.

With that said, let us address some respects in which an effect of cognition on perception must be direct, and how this relates to issues of content-specificity and semantic coherence, perceptual learning and diachronic cognitive penetrability, and possible intermediaries of relevant effects such as attention and imagination.

2.4.1 Content-specificity and semantic coherence

Can cognition influence perception in a goal-directed way? Of course it can. A most trivial way it does so frequently is by influencing whether or how or what stimuli impinge on the sensory organs. Accordingly, we can intentionally change our perceptual experience for example by shutting off the lights in a room (*whether*), or putting on a pair of sunglasses (*how*), or changing the direction of our gaze by turning our eyes or head (*what*).

Now it may be argued that even though these cases are trivially indirect and uninteresting, they should be discounted as cases of cognitive penetration not on account of their indirectness, but on account that the respective effects are not specific to the content of source cognitive states. For example, say I turn around to see my friend Mary, who I believe is standing behind me. If, after turning around, I in fact see Mary, the resulting experience would seem to cohere with my intention. However, say my intention was to see George; I would plausibly still see Mary. Thus, changes in visual experience as a result of turning around do not seem content-specific, insofar as any cause or reason for turning around elicits the same effect.

Some argue accordingly that it is insufficient for content-specificity that the source content and the resulting content actually cohere. In their view, it should also hold

counterfactually that if the source content had relevantly differed, then the resulting content would also have relevantly differed (e.g., Stokes, 2013; but cf., e.g., Gross, 2017).

However, such specificity may still not suffice for logical–rational coherence. For say my visual experience as of Mary was specific to my intention to see her. My experience would then relevantly differ if my intention were instead to see George. Still, it is at least a logical possibility that while my intention to see Mary results in Mary-like visual experience on one occasion, the same intention leads to George-like visual experience on another occasion, yet Peter-like visual experience on a still further occasion, and so on. Such multiple realizability of cognitive-content-specificity would not seem capable of sustaining logical–rational coherence. So, provided that all other parameters are fixed, perhaps counterfactual determination should run both ways between cognitive and perceptual content.

All the same, the condition of directness does not reduce to the conditions of specificity and coherence. Consider another example. Say I desire the taste of chocolate. I thus eat some chocolate, which leads to gustatory experience as of chocolate. If various unrelated desires would also have led to some or other action that resulted in chocolate experience, then my current experience is not specific to my desire. But even if it is, if the same desire led to relevantly different action or at least relevantly different experience on each occasion, then my current chocolate experience does not counterfactually cohere with my desire.

Let us entertain no such funny business, though. Assume, on the contrary, that my chocolate-related desire and experience are mutually counterfactually dependent. (So, yes, assume such luck that my desire for chocolate is always satisfied.) Specificity and coherence with bells on. Still no one is (or would be) clapping for cognitive penetration. For the effect is trivially mediated by external stimuli. It is an uninterestingly indirect way for cognition to influence perception in this way.

2.4.2 Perceptual plasticity and diachronic cognitive penetrability

The point that externally mediated influences of cognition on perception are trivially indirect and uninteresting generalizes from synchronic to diachronic candidates of cognitive penetration as well. Synchronic effects are temporally direct responses to some token stimulus or content that may have no lasting effect on perceptual capacities. In contrast, diachronic effects are temporally indirect responses to repeated exposure to some type of stimulus or content that leads to long-term changes in perceptual functioning (cf. Fodor, 1983, 1984, 1988; Pylyshyn, 1999a, 2003).

Sure enough we often engage in strategic training of our perceptual capacities that results in perceptual learning consistent with our needs. Musicians, painters, radiologists, wine tasters, and various other expert perceivers could hardly achieve what they do if the relative plasticity of our perceptual systems could not be exploited in various indirect ways. Still, even some leading researchers of perceptual learning acknowledge that indirect diachronic penetration of perception does not amount to cognitive penetration:

> If we understand cognitive impenetrability to mean that our perceptual processes are structured such that we have access to their outputs but no ability to adjust their internal workings, then these examples of expert performance are imperfect cases of cognitive penetrability to be sure. . . . However, we routinely and strategically modify human perceptual systems by giving ourselves and our students targeted training. Professors of dermatology strategically present their medical students with paired examples of Lichen planus and Psoriasis to help them distinguish these confusable diseases. Very different training is required of music students to master discriminations between absolute pitches (e.g. A vs A#) vs relative intervals (e.g. minor vs major thirds) (Aruffo, Goldstone, &

> Earn, in press; Hannon & Trainor, 2007), and students regularly avail themselves of training methods suited to their musical goals. Dr. Susan Barry lacked binocular stereoscopic depth perception, but was able to strategically train herself to have this ability by presenting to herself colored beads at varying distances and forcing her eyes to jointly fixate on them (Sacks, 2006). It is worth noting for this last example that binocular depth perception is one of the human perceptual abilities with the strongest empirical claims for having status as a neurophysiologically and functionally genuine module (Nakayama, 2005). (Goldstone, Leeuw, & Handy, 2015, p. 27)

Two relevant conclusions that the authors draw are that, one, modules can be revamped to fit one's needs; and, two, even if such effects are indirect, they can still rationally cohere with one's beliefs, expectations, or values. The moral for our discussion may be put thus: even if perception is diachronically penetrable; and even if diachronic penetration can sustain rational coherence; insofar as the effect is indirect for stimulus mediated, it is not a case of cognitive penetration.

So far, so good. The not so good part is that, like many others, the above authors misinterpret diachronic penetrability as incompatible with Fodorian modularity, and the sense in which it may sustain rational coherence as in friction with Pylyshyn's view. It may serve to touch upon these points for the sake of uncovering some underlying issues. So let us consider, first, whether guided perceptual learning is compatible with Fodorian modularity. The authors deny this on account that "it is the very nature of a module that it cannot be strategically revamped to fit one's needs" (ibid.). This echoes Churchland's (1988) (ill-)famed attack on Fodor's thesis on account of perceptual plasticity. Here's a relevant passage:

> Now I will grant that, its cognitive origins aside, the Müller-Lyer illusion cannot be overridden by any casual, fleeting, "voluntary" attempt to modify the character of one's visual experience. By itself, however, this means relatively little, for the issue is not whether visual processing is in general very *easily* or *quickly* penetrated by novel or contrary information: the issue is whether in general it is penetrable at all, where the acceptable means of penetration include long regimes of determined training, practice, or conditioning. If the Müller-Lyer illusion is an incidental consequence of a long period of perceptual training on certain typical kinds of perceptual problems, then presumably a long period of training in an environment of a quite different perceptual character would produce a subject from that particular illusion. Fodor, it seems to me, is in no position to insist otherwise, especially given examples of the following kind, which are not speculative, but real. (p. 174; emphases in original)

The key premises of Churchland's argumentation seem to be: (i) It is not a requirement of cognitive penetrability that the influence of cognition on perception must be *temporally* direct. (Hence, even if perception is synchronically impenetrable by cognition, it doesn't follow that it is diachronically impenetrable.) (ii) Perception is most plausibly diachronically penetrable by *some* means (e.g., by training, practice, conditioning). He thus concludes that Fodor's "views on impenetrability of perceptual processing are almost certainly false" (p. 167).

However, it is clear both from Fodor's (1988) reply to Churchland, as well as from his earlier writing on these matters (1984), that he agrees with both premises. Consider:

> To deny diachronic penetrability [by any means] would be to claim, in effect, that *all* the background information that is accessible to modular perceptual systems is endogenously specified, and that is viewed as implausible even by

> mad dog nativists like me. For example, parsing may be modular, but children must learn *something* about their language from the language that they hear; why else would children living in China so often grow up speaking Chinese? The point about the diachronic penetrability of perception is, however, just like the point about its synchronic penetrability: it offers an argument for the continuity of perception with cognition only if just any old learning or experience can affect the way you see, and there is no reason at all to suppose that that is so. Perhaps, on the contrary, perception is diachronically penetrable only within strictly—maybe endogenously—defined limits. Not only do your current Copernican prejudices fail to much dispel the apparent motion of the Sun, it may be that there is *no* educational program that would do the trick; because it may be that the inaccessibility of astronomical background to the processes of visual perceptual integration is a consequence of innate and unalterable architectural features of our mental structure. (1984, pp. 39–40; emphases in original)

So Fodor argues against cognitive penetrability of both the synchronic and diachronic kind, yes. But, pace Goldstone et al. (2015) and Churchland (1988), his skepticism regarding diachronic penetrability is not rooted in skepticism about perceptual plasticity or perceptual learning (so he does not deny [ii]). Further, pace Churchland, his reason for discounting evidence for diachronic penetrability as evidence for cognitive penetrability is not based on temporal indirectness (so he does not deny [i]). Rather, his main consideration is that all candidate evidence involve external mediation by stimuli. So Fodor doubts *direct* diachronic penetrability of perception by cognition, not indirect diachronic penetrability via stimuli. To argue for the former based on the latter is clearly invalid (indirect $\vdash$ direct), even if only informally when concealed by ambiguity (diachronic $\vdash$ cognitive).

2. Cognitive Impenetrability

A most crucial moral of the above is that, some unfounded assumptions to the contrary notwithstanding, no one considers temporal directness a criterion of cognitive penetration. Skeptics of diachronic cognitive penetrability have no issue with lack of synchronicity per se. Changing some beliefs can also take time. However, whereas changing beliefs may involve persuasion and dissuasion and inference, perceptual learning might involve no more than the shaping of basic sensors and reinforcement and extinction of associations, whereby "tens to thousands of repetitions" of stimulus exposure are needed "to allow an organism to better make discriminations and categorizations" (Goldstone et al., 2015, p. 25). Insofar as time scale is of any issue, it is this dependence on multiply repeated stimulation that renders diachronic candidates of cognitive penetration suspect even if not by default discounted.

So much for Fodorian modularity and perceptual plasticity and temporal directness. Let us now turn to whether Goldstone et al. (2015) are right in attributing to Pylyshyn the view that the reason indirect diachronic penetration does not qualify as cognitive penetration is that the respective effects do not sustain a rational relation to cognition. The authors emphasize, contra, the example that

> as Susan Barry trained herself to have depth perception, she *wanted* to have depth perception, she *believed* that her training regime would serve to give her this ability, and the training itself was systematically related to her developing the ability. If this does not count as a rational process, then many canonical cases of human rational inference will probably fail to count as well. (p. 27, emphases in original)

It is unclear what such argumentation is supposed to show. Surely, Pylyshyn agrees that desires and beliefs are sensitive to rational properties. He also agrees that propositional attitudes can influence behavior in relevantly cohering ways. So there is no reason to assume

that he would object to counterfactuals of this sort: in at least some perceptual-learning cases, had cognitive content relevantly differed, perceptual training via stimulus exposure would also have relevantly differed, in which case the resultant changes in perceptual capacities would also have relevantly differed. If this is what is meant by that perception is rationally sensitive to cognitive influences, then indirect diachronic effects may well sustain semantic coherence.

Why does Pylyshyn still claim, then, that "changes produced by shaping basic sensors, say by attenuating or enhancing the output of certain feature detectors . . . do not count as cognitive penetration because they do not alter the contents of perceptions in a way that is logically connected to the contents of beliefs, expectations, values, and so on" (1999a, p. 343)? What he has in mind is that such modulation may only alter for example the amplitude of their response, rather than the features to which the detectors are sensitive (cf. also Fodor & Pylyshyn, 1981).

This relates to two key background assumptions. First, recall that as Pylyshyn understands the term, functional architecture is merely physically (or biologically) specifiable. Insofar as basic feature detectors form part of functional architecture, the properties to which they are sensitive must therefore also be merely physically specifiable. Second, on this view, the reason generalizations about cognition are more successfully cast by appeal to explicit content is that cognitive categories cross-cut and hence are non-reducible to physical categories (cf. also Fodor, 1974; Pylyshyn, 1984). If these assumptions are correct, then perception simply cannot form certain categories that would be required in a strict sense for content-specific responsiveness to cognitive content.

Here's an example that connects closely to the discussion in the next chapter. Say one detector encodes lightness, whereas another simple form features. Consider relatedly that there is some correlation between the lightness and form features of faces of different races. For

example, white (Caucasian) faces are typically brighter in luminance than black (African American) faces. It is thus not implausible that repeated exposure to an appropriate sample of faces may lead to differential amplitudinal response of the lightness detector in the context of typical white vs. typical black facial-form features, even if the mean luminance of the faces is objectively similar (Baker & Levin, 2016; Bitter, 2014; Firestone & Scholl, 2015a; Levin & Banaji, 2006).

In such cases, we might say that lightness experience is diachronically penetrated. But not by cognition: as we shall soon see, it is unlikely that perception of such features is sensitive to such cognitive categories of race as CAUCASIAN and AFRICAN AMERICAN. Accordingly, once the intra-perceptual form–lightness associations are in place, they produce similar effects for non-faces as well (Lupyan, 2015a). So even if such effects appear to semantically cohere with cognitive content on first thought (white face → light; black face → dark), on second thought, even strategic perceptual training may involve no more than content-general effects of "a cognitively mediated directing of the visual system . . . toward certain physical properties" (Pylyshyn, 1999a, p. 342).

Hence, pace Goldstone et al. (2015), Pylyshyn does not deny the possible surface rationality of indirect diachronic effects of cognition on perception. Everyone agrees that "the cycle of interactions between moment-to-moment goals, strategic training, perceptual learning, regular monitoring of performance, and revision of goals . . . can often achieve functional hacks to our perceptual systems even without our being able to directly manipulate our neural networks" (Goldstone et al., 2015, p. 27).

Further, a crucial reason Pylyshyn discounts indirect diachronic effects as sustaining rational coherence in a strict sense is not that they are externally mediated. His consideration is rather that—whether externally or internally mediated—mere amplitude changes in stimulus

response do not alter the set of properties to which early vision is sensitive. To the extent that cognitive concepts may cross-categorize such categories, content-specificity in a strong sense thus seems off the table. This view may be false; but to argue so on account of strategic functional hacks to the perceptual system is to argue against a straw man.

2.4.3 Attention

Nutshell: Directness is at least oblique to semantic–logical–rational coherence and temporality (synchronicity vs. diachronicity). However, it is not oblique to the locus of mediation (internal vs. external) and the psychological properties of mediators (mental vs. non-mental). Regarding locus, it is common ground that cognitive penetration should be carried by internal states or processes. Internal is generally understood as within the body (excluding the periphery and the sense organs), and more specifically as neural (or a neural correlate). So effects mediated by external stimuli, the sense organs, or the heart and kidneys are excluded. Those who adhere to the semantic-coherence criterion also expect that the mediators should be representational qua intentional (cf. Pylyshyn, 1984). But even those who discard or reinterpret the criterion agree that all mediating states and processes should be mental (cf. Macpherson, 2015; Stokes, 2013).

Might cognition directly penetrate perception without the involvement of any internal–mental intermediary? Few propagate such a view (e.g., Lupyan, 2015a). Most defenders of cognitive penetrability accordingly lax the requirement of directness. Those who still commit to the criterion of semantic coherence might then argue for example that the source cognitive content is delivered to perception "as is" by proxy, or at least all intermediaries have the content that they do in virtue of the source cognitive content (Macpherson, 2015; cf. also Gross, 2017).

A currently most widely debated such candidate intermediary is attention (e.g., Block, 2016; Fazekas & Nanay, 2017; Firestone & Scholl, 2016a; Gross, 2017; Lupyan, 2015a; Mole, 2015; Raftopoulos, 2017b; Stokes, 2017; Wu, 2017). As is the case regarding virtually all

issues of cognitive penetrability from a computational–functional perspective, some most influential skeptical arguments against attention-mediated effects were initially put forth by Pylyshyn (1984, 1999a, 2003, 2007) and Fodor (1988); and from a cognitive-neuroscientific perspective, by Raftopoulos (2001, 2009). The topic is wide-ranging and complicated, but the details need not concern us here. Of relevance are three general issues: the locus of processing at which attention might intervene; the objects of attention that may be selected or filtered; and possible ways in which such functions might influence perceptual experience.

Regarding the locus of intervention, a relatively longstanding and highly influential proposal is that attention can be allocated to certain locations or features or objects only before or after early visual processing. Which is to say, attention cannot directly influence early vision. What early or pre-perceptual attentional processes may do is to select or filter the input that get fed into the system. Late or post-early-perceptual mechanisms may then facilitate decision processes regarding the interpretation of the output by facilitating the retrieval from or encoding into long-term memory of corresponding categories (and/or inhibiting similar retrieval or encoding of other categories) (Pylyshyn, 1999a, 2003).

It is nowadays increasingly claimed that this is an outdated model of attention, for example on account that attentional effects are detectable at all levels of visual processing (e.g., Carrasco, 2011; Fazekas & Nanay, 2017; Lupyan, 2015a). However, much of the evidence for this claim comes from neurophysiology, the theoretical implications of which for functional theses of attention are at least debatable. For example, it would be naïve to (mis)interpret the idea of pre-perceptual attentional selection as requiring that it must be implemented outside of early visual areas as neuroanatomically defined. A possible alternative, among others, is that such effects merely temporally precede early visual processing, whereby they may for example alter the weighting of certain functions (cf. Raftopoulos, 2017b; also Gross, 2017).

2. Cognitive Impenetrability

One may recall relatedly the exchange between Churchland (1988) and Fodor (1988) regarding the large number of neural pathways that descend from higher cognitive to higher perceptual centers. Churchland argued that such evidence militates against cognitive impenetrability. Fodor noted in reply that while our knowledge about such brain structures and functions is seriously limited, "One thing *is* clear: if there is no cognitive penetration of perception, then at least 'descending pathways' aren't for *that*" (p. 194; emphases in original). As they say, one's modus ponens is another's modus tollens. Neuroscience might inform psychology, but it hardly trumps it by default (cf. Coltheart, 2006, 2010, 2013).

Yet one may just as well ask: what, if any, theoretical relevance do functional distinctions such as between direct and indirect attentional effects, or early and late selection, have for issues of cognitive penetrability of perceptual experience? One consideration is that attention may be the only internal–mental means by which cognition may influence perception. A second consideration is that architectural constraints on when or at which processing stages attention may exert an influence may have implications for how or which perceptual features may be selected or filtered. Unless one assumes a mental architecture in which early perceptual processing may be completely redundant or inconsequential for perceptual experience (which, recall, no thesis actually allows), constraints on attention may thus plausibly translate to constraints on penetrability of phenomenal content or character.

Consider for example that if early attention is pre-perceptual, its targets can only be relatively low-level, physically specifiable features such as spatial location or color. However, recall that early vision may even encode basic objects such as *table* and *dog*, non-sensory properties such as causal relations, primitive affective properties such as dangerousness, and functional properties such as affordances. Once such output are delivered, there may be little

tinkering room left for late attention, beyond the facilitation or inhibition of retrieval from or encoding into memory of corresponding categories, response selection, and the like.

Consider relatedly the famous case of expert "chicken sexers." Apparently, it takes years of training to master the visual skill to reliably determine the sex of one-day chicks. More precisely, it used to take years, because experts could not quite tell what it was that visually informed their discriminations. After studying the phenomenon, a pair of psychologists found that it was the convexity or lack thereof of a particular area of a cloacal structure. When naïve subjects were instructed to focus on the respective location and feature, their performance suddenly neared that of expert chicken sexers (Biederman & Shiffrar, 1987).

Now, before perceptual learning of chicken sexing was understood, the phenomenon may have appeared a prime candidate for penetration by cognitive categories such as FEMALE and MALE. But the finding that a simple contrast of shape (convex vs. concave or flat) does the trick suggests otherwise. All that seems required is early (pre-perceptual) allocation of attention to the locus of interest, and/or late (post-early-visual) attentional facilitation of the perceptual category CONVEX.

That experts could not before report the exact location or feature they were focusing on is irrelevant. The issue is not whether their acquisition of the perceptual skill was explicit in terms of conscious cognitive access to any relevant category or belief. The moral is that insofar as acquisition of the perceptual skill was mediated by learning where or on which feature to allocate attention, possible differences in their visual experience as compared with novices may have been constrained in ways that preclude cognitive penetration (cf. Pylyshyn, 2003).

Indeed, it seems that the possible perceptual effects of attention are highly constrained. For example, true, perceived contrast and spatial resolution appear affected by where spatial attention is allocated (Carrasco, 2011). But, then, it is not for example the case that if one

focuses more or less intensively on some location, then either of these parameters increases or decreases respectively.

Or consider the switches of perspective of ambiguous figures such as the Necker cube. Apparently, this can be voluntarily facilitated by switching attention from one location to the other. However, whether or not a phenomenal switch of perspective actually occurs is ultimately a function of a change in perceptual coupling, whereby visual interpretation of one region of the figure affects visual interpretation of other regions of the figure. Yet the principles of such interpretation and coupling seem paradigmatically hardwired into the visual system and hence independent of cognition (Rock, 1983; cf. also Pylyshyn, 1999a, 2003).

Consider relatedly that feature-based attention—which is generally distinguished from both spatial and object-based attention—can apparently influence phenomenal brightness (Tse, 2005) and saturation (Fuller & Carrasco, 2006). Still, no evidence has been found that attending to color might influence perceptual experience of hue (ibid.). Or, more precisely, there is no evidence that attention might alter the qualitative appearance of hue. At best, late attentional facilitation may enable visual pop-out, whereby a target hue is detected more easily and hence searched faster among distractor hues. Although such effects have been cited as evidence for cognitive penetration (e.g., Prinz, 2006), the overwhelming consensus is that mere attention grabbing or speeded detection are non-qualifying insofar as not (relevantly) content altering.

This is not to suggest that any other attention-mediated effect qualifies as cognitive penetration. Recall the example that spatial attention affects perceived contrast. In certain respects, the effect resembles a contrast filter: whatever location or region is viewed through the filter appears sharper in contrast by a gradient pre-determined by the filter. However, call to mind also that changes in perception mediated for example by putting on glasses are non-cases of cognitive penetration, even if the resulting experience in fact coheres with the reason

for putting on the glasses. Like contrast filters, once you put them on, they achieve what they do no matter what.

The problem with attentional effects, then, is not just that their perceptual consequences are too rigid and too narrow in range. A further consideration is that their potential causes are too flexible and too wide in range, involving various cognitive and non-cognitive sources as well (cf. Firestone & Scholl, 2016a).

Finally, relating to the latter point, note that most of the evidence on attentional effects comes from studies of exogenous rather than endogenous attention. Exogenous attention is passive, reflexive, stimulus-driven, bottom-up, and transient. In contrast, endogenous attention is active, voluntary, goal-directed, top-down, and sustained. Surely, it is the latter that has more relevance vis-à-vis cognitive penetrability.

Hence, while the fast-growing research on attention is truly fascinating, and the debate about its candidacy as a prime intermediary of cognitive penetration is only anticipated to strengthen, as of the writing of this chapter, we may conclude that its effects on perception are not relevantly direct or content-specific. Thus a reason for defining cognitive penetration in the introductory chapter as a semantically–logically–rationally cohering effect of cognition on perceptual experience that is not mediated by changes in the stimuli, the sensory organs, or the allocation of attention.

2.4.4 Imagery

So is perception cognitively penetrable? The discussion of this chapter started with mention of a famous debate in psychology concerning the representational format of mental images. As we saw, the argument from cognitive penetrability was intended to underscore that mental images cannot be holistic–analogue. We left that issue as is, to discuss related distinctions

between cognitively penetrable vs. cognitively impenetrable functions; how function-focused theses relate to experience-focused theses; why demarcating cognition from perception can be challenging whether or not one commits to propositionally structured representations and/or semantic sensitivity to logical–rational properties; and the various respects in which mediators of penetrating effects might (not) be expected to be direct. We are now ready to make full circle and ask: might perception be penetrable by cognition via the proxy of mental imagery?

Imagery seems the perfect candidate: it is mental; it is at least in some cases cognitively penetrable and hence relevantly sensitive to cognitive content; it is not mediated by exogenous attention; and its phenomenal character most closely resembles that of perceptual experience. Little wonder, then, that it has been variably proposed as a most plausible intermediary of cognitive penetration (e.g., Arstila, 2017; Block, 2016; Fazekas & Nanay, 2017; Lupyan, 2015a; Macpherson, 2012; Newen & Vetter, 2017; Vetter & Newen, 2014; Stokes, 2013).

The following two chapters address some purportedly strongest and most discussed candidates of cognitive penetration by imagery accordingly. *Chapter 3* discusses in particular an apparent lightness illusion that has been widely interpreted as supporting a mixing view on which the phenomenal content or character of perception and imagery fuse in a way so as to relevantly alter for example color experience. *Chapter* 4 takes on a most exciting phenomenon called suggested hallucination, a current mainstream interpretation of which implies not only frequent extreme phenomenal mixing, but also occasional full phenomenal trumping of perception by imagery. As wild as the interpretation sounds, there is hardly a more plausible or better-researched candidate of cognitive penetration. Thus the dialectic significance if the pending arguments succeed in showing that, appearances to the contrary notwithstanding, none of these cases plausibly involve cognitive penetration.

Proponents of cognitive penetration should not despair, though. The conclusion is not that all non-attentionally mediated mental effects can be successfully explained away either as intra-perceptual penetration, or as penetration of perceptual belief. A defended third possibility is that some effects involve a penetration of perceptual affect. This extension of the list of potential targets of penetration hopefully provides food for thought and contributes positively to future fruitful debates.

3. The Face–Race Lightness Illusion

Is there any evidence for cognitive penetrability of perceptual experience? The previous chapter discussed some crucial constraints that such evidence would have to meet. For example, it is insufficient to show that some candidate effect involves a change in perceptual belief. The question is whether perceptual phenomenology may itself be relevantly infiltrated. By what and how? First, all mediating states and processes must be mental. So, for example, stimulus-mediated changes or hormonal effects do not count. Second, the source of penetration must be cognitive. So intra-perceptual top-down effects are also non-cases. Third, the effects should not be mediated by attention. For example, flips of perspective of ambiguous figures facilitated by shifts in spatial focus are paradigmatically discounted. Once all such factors are fixed: are there plausible cases of content-relevant influences of cognition on perception?

Many philosophers (e.g., Brössel, 2017; Macpherson, 2012; Marchi & Newen, 2015; Nanay, 2015; Stokes, 2013) and psychologists (e.g., Collins & Olson, 2014; Hugenberg & Sacco, 2008; Hugenberg, Young, Sacco, & Bernstein, 2011; Vetter & Newen,[5] 2014) alike have argued that a particular series of psychological experiments provides most convincing evidence for such an influence. In these experiments, subjects judged images of prototypical white (Caucasian) faces as lighter than objectively equiluminant images of prototypical black (African American) faces. Interestingly and importantly, they also judged the same image of a racially ambiguous face as lighter when it was labeled 'White' as opposed to 'Black' (Levin & Banaji, 2006). The latter result is deemed especially indicative of cognitive penetration on account that the distortions in judgment most plausibly reflected distortions in perceptual experience, which were in turn most plausibly lexically and hence cognitively mediated.

[5] The first author, Petra Vetter, is a psychologists. The second author, Albert Newen, is a philosopher.

The apparent face–race lightness illusion is thus widely interpreted as demonstrating that even high-level beliefs or concepts such as concerning race can influence perceptual experience of such low-level features as color and lightness. If this claim is true, then quite possibly any or at least most conscious perceptual features are similarly cognitively penetrable. In turn, this has potentially important implications for issues regarding mental architecture and the continuity of perception with cognition, the justificatory role of perception in the formation and updating of beliefs, and the theory-ladenness of scientific observation. So it warrants careful examination whether or to what extent the above interpretation is correct.

The first section (*Section 3.1*) provides a summary of the psychological experiments in question accordingly. The second section (*Section 3.2*) dissects a prima facie most powerful argument for cognitive penetration that draws on data from these experiments. The analysis suggests that none of the premises of this argument are warranted, and hence there is plenty of room to resist its conclusion that visual experience is cognitively penetrable. The third section (*Section 3.3*) proposes a purely perceptual account of the evidence, which is arguably both empirically and abductively more plausible than a cognitive-penetration account. The upshot is that cognitively impenetrable perceptual systems may be psychologically more plastic and hence philosophically more significant than is commonly assumed. Some implications are in the final section (*Section 3.4*) accordingly touched upon.

It bears emphasis that even if the proposed explanation is ultimately off, a non-cognitive account of the phenomena may still be more warranted. This point will be elaborated in the next chapter. The moral of the present chapter, then, is that it is at least a real empirical possibility that intra-perceptual processes may themselves carry significant explanatory burden. Here's why.

3.1 The Experimental Evidence

In a series of four experiments, Levin and Banaji (2006) tested whether images of faces racially categorized as white are judged differently for lightness than objectively equiluminant images of faces racially categorized as black. Here's a quick rundown of the experiments.

The first experiment used computer-morphed grayscale images of a prototypical white face and a prototypical black face. The face images were presented in pairs of all combinations (white–white, black–black, white–black, black–white). In each trial, the initial luminance of the images was offset. The task was to adjust the luminance of one image to match the luminance of the other image. As expected, subjects adjusted images to objectively lighter/darker levels when matching for white/black faces. The effect was significant for all combinations of faces, but it was especially marked for pairs including faces of different races.

The second experiment added to the stimuli a computer-morphed grayscale image of a racially ambiguous face. During instructions, one group saw the image of the ambiguous face under the label 'White,' next to an equiluminant image of the prototypical black face, which was labeled 'Black' (ambiguous–black). A second group was presented the ambiguous-face image under the label 'Black,' next to an equiluminant image of the prototypical white face, which was labeled 'White' (ambiguous–white). In the trials, the images were shown each on its own without any label, next to a gray patch the initial luminance of which was offset. The task was to adjust the luminance of the patch to match the luminance of the presented face image. Consonant with the first experiment, the gray patch was adjusted to an objectively lighter level for the white face than for the black face. In addition, and most importantly, the ambiguous face was also matched to a lighter gray if it was initially labeled 'White' as opposed to 'Black.' The magnitude of the effect was even greater than in the previous experiment.

3. The Face–Race Lightness Illusion

The third experiment changed the stimuli to evenly gray-filled line drawings of an unambiguous white face, an unambiguous black face, and a racially ambiguous face. In some trials the lines were dark, whereas in other trials the lines were light. The procedure was similar to that of the second experiment, although the between-subjects design was changed back to a within-subjects design. So all subjects completed trials for both when the ambiguous face was initially paired with the white face (ambiguous–white), as well as for when the ambiguous face initially appeared next to the black face (ambiguous–black).[6] Cohering with previous findings, subjects adjusted a gray patch to an objectively lighter level for the white as opposed to the black face, as well as for the ambiguous face when it was initially paired with the black as opposed to the white face. Though the magnitude of the effect was smaller than for computer-morphed images, both the magnitude and the direction of the effect were similar for dark- and light-line drawings.

The fourth experiment used the same computer-morphed grayscale images of a prototypical white face and a prototypical black face as the first experiment. Subjects were told that the pairs of faces would vary in luminance, but they were instructed to ignore these differences. The task was to indicate by pressing a button as quickly and accurately as possible whether the presented faces were of the same race or of different races. The results indicated that the more similar in luminance the pair of faces were, the longer it took to discriminate the faces by race. Yet, importantly, reaction times were slowest not when the faces were objectively equiluminant, but when the white (black) face was a bit darker (lighter) than the black (white) face.

Hence, to summarize, it seems that differences in facial form (white vs. black vs. ambiguous), lexical label (‘White’ vs. ‘Black’), and/or visual contrast (ambiguous–white vs.

[6] Unfortunately, the authors do not mention whether they used the labels ‘White’ and ‘Black’ in this experiment.

ambiguous–black) influence the judged lightness of faces. Do such differences in judgment reflect differences in perceptual experience?

One reason for skepticism relates to the results of the first experiment. Recall that in that experiment, subjects judged white/black faces as lighter/darker in all trials, by means of adjusting the luminance of one face to match that of another face. Yet recall also that in some pairings of faces (white–white, black–black), the reference face and the matching face differed only in their initial luminance. So why did subjects not adjust the luminance of the matching face to at least roughly the same objective level as that of the reference face? The faces would then have been token replicas of each other. That subjects nevertheless over- or under-adjusted the luminance of the matching face thus suggests that, in at least these cases, their task performance was driven not by distortions in perceptual experience, but by distortions in judgment and/or some other response bias(es).[7]

Now Levin and Banaji attempt to explain away the above finding in purely perceptual terms. But we need not concern ourselves with their account. For even if their proposal is wrong, recall that distortions in judgment were greater in the different-race trials (white–black, black–white) than in the same-race trials (white–white, black–black). Consider further that the effect was even greater in the second experiment, in which the matching stimulus was not a face but a mere gray patch. Recall still further that in the fourth experiment, subjects were slowest in discriminating a pair of faces by race not when the faces were equiluminant, but when the white (black) face was slightly darker (lighter) than the black (white) face (fourth experiment). The latter result seems especially hard to explain on cognitive grounds.

[7] For more on the error of inferring distortions in perceptual experience on the basis of distorted judgments when the matching and the reference stimuli are similar, cf. discussion of the "El Greco fallacy" in perception research (Firestone & Scholl, 2014, 2016a).

Yet perhaps the results reduce to stimulus or attentional artefacts? Some take note for example that the eyes of the prototypical white face were objectively darker than the eyes of the prototypical black face (e.g., Machery, 2015; cf. also Lupyan, 2015a). This was needed to assure that the average luminance of the faces were identical. But, then, recall that the third experiment found a similar effect for gray-filled line-drawing faces, whether the lines were light or dark. Although it cannot be fully ruled out that the latter result was itself driven by response bias, it at least lends prima facie support to the assumption that some portion of responding was carried by relevant distortions in perceptual experience.

Hence, even if another portion of responding was likely contaminated by response bias, all things considered, it is at least not implausible that "the relative associations between lightness and White faces and between darkness and Black faces . . . make White and Black faces appear lighter and darker, respectively, than they actually are" (Levin & Banaji, 2006, p. 501). We shall grant this conclusion in the remainder of this chapter.

Note, though, that nothing ultimately hangs on doing so. For the crux of the ensuing argument is that even if the face–race lightness illusion is a genuine perceptual phenomenon that is uncontaminated by artefacts, the effect still doesn't constitute a convincing case of cognitive penetration. If it turns out that the effect is not even genuinely perceptual, or that it is a complete artefact, then so much the worse for theses of cognitive penetrability.

3.2 The Argument from Lexical Labeling

What makes Levin and Banaji's (2006) findings so special? Note that it has long been known that color judgments are distorted for color-diagnostic objects. For example, in a classic psychological study that has received much attention in the recent debate about cognitive penetrability, subjects judged orange–red cutout figures as more red for red-associated figures

(heart, apple, lips) than for other-color-associated (horse, bell, mushroom) and non-color-associated figures (oval, circle, ellipse) (Delk & Fillenbaum, 1965; cf. Arstila, 2016; Brogaard, 2014; Brogaard & Gatzia, 2017a, 2017b; Burnston, 2017; Deroy, 2013; Gatzia, 2017; Gross et al., 2014; Lupyan, 2015b; Macpherson, 2012; Nanay, 2015; Newen & Vetter, 2017; Raftopoulos, 2017a; Varga, 2017; Vetter & Newen, 2014; Zeimbekis, 2013).

Such results cohere with more recent evidence according to which subjects set the subjective gray point of color-diagnostic objects such as a banana or a Nivea tin slightly toward the opponent color of the typical color of the object (e.g., toward blue for a banana, and toward yellow for a Nivea tin). The alleged explanation of these findings is that achromatic images of color-diagnostic objects still appear slightly in their diagnostic color (e.g., an image of a gray banana still looks a bit yellowish), and thus sensation of the opponent color is required for gray perceptual experience (e.g., Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Olkkonen, Hansen, & Gegenfurtner, 2008; Witzel, Valkova, Hansen, & Gegenfurtner, 2011).

Some recent studies cast doubt on whether these findings reflect genuine perceptual phenomena (e.g., Gross et al., 2014). But assume that they do. Still, many defenders of cognitive penetration themselves acknowledge that such cases of perceptual learning may ultimately be explained by non-cognitive factors. For example, as Macpherson (2012) notes,

> It might be that the early visual system, autonomously from belief or other cognitive states, alters the colour experiences of characteristically red shapes, to make them appear more red than they really are. This might happen in accordance with associationist principles, so that it is past exposure to a certain shape having a certain colour that has altered the way the visual system processes that shape's colour. (p. 46)

3. The Face–Race Lightness Illusion

Now distortions in the judged lightness of racially prototypical white/black faces seem to afford a similar explanation in terms of intra-visual shape–color (or form–lightness) associations. Since such intra-visual associations are plausibly cognitively impenetrable, this would explain why Levin and Banaji's (2006) fourth experiment found face discrimination by race to be sensitive to differences in perceived luminance of faces despite explicit instructions to disregard such differences. So non-cognitively mediated perceptual learning may well account for why the face–race lightness illusion persists for racially prototypical faces despite one's contrary beliefs and desires. Accordingly, in and of themselves, distortions in the judged lightness of such faces warrant no special attention.

Yet the finding that subjects judged the very same image of a racially ambiguous face as relatively lighter when it was beforehand labeled 'White' as opposed to 'Black' strike some as particularly revealing of cognitive penetration. The apparent basis for this view is an unargued contention that "categorization was clearly done at the cognitive level as it was the labelling of the face that was responsible for the effect" (Macpherson, 2012, p. 48). A rather straightforward argument for cognitive penetration would thus seem to run as follows:

Argument from lexical labeling

(1) Distortions in the perceived lightness of racially ambiguous faces were a function of *lexical labeling*.

(2) If distortions in the perceived lightness of racially ambiguous faces were a function of lexical labeling, then the effect was mediated by *cognitive categorization*.

(3) If the effect was mediated by cognitive categorization, then low-level visual experience was *cognitively penetrated*.

Hence,

(4) Low-level visual experience was cognitively penetrated.

The following subsections argue that none of the assumptions of this argument are warranted, and hence there is little reason to assume that low-level visual experience is cognitively penetrable.

3.2.1 Re (2): Lexical vs. conceptual vs. cognitive effects

Let us first consider premise (2), according to which a lexical effect entails a cognitive effect. Why assume this? The premise seems to rely on the following (hidden) assumptions:

(2a) If the effect was lexically mediated, then it was *conceptually* mediated.

(2b) If the effect was conceptually mediated, then it was *cognitively* mediated.

Yet these assumptions are hardly a matter of course. On the contrary, there may be various reasons to question both. Here are some considerations.

3.2.1.1 Re (2a): Lexical–visual associations

In a classic experiment on the effects of prior semantic context on lexical access during sentence comprehension, subjects would hear for example the following sentence: "Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other bugs in the corner of his room" (Swinney, 1979, p. 650). Simultaneously with the word 'bugs'—which in itself is ambiguous between 'insects' and 'spyware'—subjects were visually presented with a letter string. The task was to decide as quickly as possible whether the letter string was a word or a non-word. As expected, lexical decisions were facilitated for words related to 'bugs,' but not for unrelated words or non-words. However, importantly, this facilitation occurred irrespective of whether a word was contextually appropriate ('ant') or contextually inappropriate ('spy').

Such findings suggest that, at least in some cases, the effects of sentence context on lexical access have not so much to do with semantic processing as such. Rather, the effects may be mediated by certain intra-modular associations between mere lexical forms stored in a proprietary database of a language module (Fodor, 1983). This coheres with Collins and Loftus's (1975) spreading-activation theory of semantic processing, according to which the names of concepts (words) are stored in a lexical network (dictionary) that is separate from—even if connected with—a semantic network of concepts.

Now the hypothesis of intra-modular/inter-lexical associations can be relatively easily extended to inter-modular/cross-modal associations. Accordingly, it is at least an empirical possibility that the lightness distortion effect for ambiguous faces was mediated by certain infra-cognitive associations between lexical items such as 'black' and 'white' stored in a language module, and some corresponding color representations of (e.g.) black and white stored in a visual module. Associations of such a kind might suggest a fair degree of cross-talk between modular systems. But, importantly, these systems may still be encapsulated from cognition at large. This coheres with evidence that cross-modal effects such as the "McGurk effect"—whereby visual exposure to the lip movements of a speaker influences the perception of concurrent speech sounds (McGurk & MacDonald, 1976)—are themselves undergirded by cognitively impenetrable processes (Fodor, 1983, 1988; Pylyshyn, 2003).

Note that the moral is not that the above proposal in particular is correct. Indeed, it shall be argued ultimately that the ambiguous-race face lightness illusion was most likely not even lexically mediated. Nor is the point that the perceptual distortions were non-conceptually mediated. We have yet to address that issue. The present concern is specifically the conditional according to which *if* the respective illusion was lexically mediated, *then* it follows that the effect must have been conceptually mediated.

Hence, the reason for referencing the above proposals is mainly to underscore that, contra (2a), there are various empirical and theoretical considerations on the basis of which one may wish to be cautious with inferring a conceptual effect from a lexical effect. This is a dialectically important point insofar as the burden of the present argumentation is to show that none of the (hidden) premises of the argument from lexical labeling are as firmly grounded as proponents of the argument would have us believe. Examples like that of Swinney, Fodor, Pylyshyn, Collins and Loftus, and McGurk and MacDonald thus provide at least some reason for resisting (2a). But not much ultimately hangs on this. For, as we shall presently see, the other premises of the argument from lexical labeling are themselves debatable.

3.2.1.2 Re (2b): Non-cognitive concepts

Let us quickly recap: According to the argument from lexical labeling, the ambiguous-race face illusion was lexically mediated (1), from which it follows that it was cognitively mediated (3) and hence penetrated (4), on account that lexical mediation of the effect implies its conceptual mediation (2a), which in turn entails its cognitive mediation (2b). In the previous sub-subsection, we saw that the inference from lexical mediation to conceptual mediation (2a) is far from trivial. In the present sub-subsection, let us address the conditional on which *if* the effect was conceptually mediated, *then* it follows that it was cognitively mediated (2b). If this is analytic, then so be it. But it may strike many as far from clear that the premise is justified on a substantive reading.

For example, consider a central dialectic of early theses of cognitive impenetrability. The cognitive revolution in psychology led to a widespread belief that, as opposed to earlier theorizing, even relatively low-level perceptual processes are by and large inferentially mediated (within the usual limitations of the metaphor). Inferences need premises; and at the

time it seemed natural to many that the requisite premises must be represented in the mind by sentence-like representations the constituents of which are concepts (Fodor, 1975, 2008).

Traditional proponents of cognitive impenetrability accepted this view. What they vehemently denied is that it follows from the inferential nature of some mental process that it is thereby informationally promiscuous and hence cognitively penetrable (Fodor 1985; Pylyshyn 1984; cf. also Stich, 1978). Accordingly, proponents of impenetrability may in principle allow that perceptual processes are by and large conceptually mediated. What they would still deny is that such processes have relevant access to cognitive content.

With that said, it is important to keep in mind that the issue of whether conceptual mediation entails cognitive mediation is orthogonal to the issue of whether perception is cognitively penetrable. For example, although a staunch defender of cognitive impenetrability, Raftopoulos (2014a, 2014b, 2017a) argues that there is a mutual entailment relation between concepthood and cognitive penetrability. On the other hand, while Toribio (2014) seems neutral on the issue of cognitive penetrability, she counters that the mutual entailment thesis is either trivial, or it completely fails to engage with the contemporary literature on perceptual nonconceptualism. This issue will certainly not be resolved here. The takeaway message is merely that, given such considerations and controversies, it is hardly a matter of course that one may infer a cognitive effect on the basis of a conceptual effect.

All the same, perhaps a defender of cognitive penetration may still argue that such an inference is warranted in the particular case of the ambiguous-race face illusion. For example, it might be pressed that even if we allow for the existence of non-cognitive concepts, surely, concepts of race are still cognitive. Perhaps so; but nothing follows from this. For notice that the lexical labels used in the relevant experiment were ambiguous between color terms and race terms (‘Black’ and ‘White’). So even if distortions in the perceived lightness of racially

ambiguous faces were lexically and conceptually mediated, it would still be an open empirical possibility that the effect might have been mediated by non-cognitive concepts.

Indeed, "psychophysical" concepts such as of color are prime candidates of non-cognitive concepts that could be available for tokening by cognitively impenetrable perceptual systems (Fodor, 1987). So a skeptic of cognitive penetration may well grant that the relevant effect was mediated by low-level color concepts such as WHITE and BLACK. Contra (2b), it is unclear why it would follow that the effects of such color concepts were in turn mediated by high-level cognitive-race concepts such as CAUCASIAN and AFRICAN AMERICAN, respectively.

It has been objected earlier that even if the above scenario is an empirically real possibility, all things considered, it is still abductively more plausible that the ambiguous-race face illusion was cognitively mediated. *Section 3.3* argues to the contrary. For now, the relevant reply is that our concern is still not whether the consequent of (2b)—the claim that there was a cognitive effect—is abductively plausible. The target of this subsection was merely the conditional according to which a conceptual effect entails a cognitive effect. It is this claim, it was argued, that there is plenty of room to resist.

3.2.2 Re (3): Cognitive priming vs. cognitive penetration

So far, we have provided some reasons to resist taking it for granted that the alleged labeling effect was conceptually (2a) and hence cognitively mediated (2b). The second premise (2) of the argument from lexical labeling thus seems at least more questionable than many defenders of cognitive penetration suggest. Of course, this still leaves open the possibility that the ambiguous-race face illusion was in fact cognitively mediated. Premise (3) states that it would then follow that low-level visual experience was cognitively penetrated. Is there reason to question this claim? The below two sub-subsections argue that indeed there is.

3.2.2.1 Semantic–logical–rational coherence

It bears emphasis that the face lightness illusion for racially prototypical faces persists even if one knows that the faces are objectively equiluminant. Both Levin and Banaji (2006), and Macpherson (2012) emphasize this. Which is somewhat ironical given that the persistence of an illusion in the face of contrary beliefs is traditionally considered strong prima facie evidence that the very states or processes underserving the illusion are cognitively impenetrable.[8]

Beliefs and similar attitudes are paradigmatically cognitively penetrable. So the persistence of the face lightness illusion for racially prototypical faces strongly suggests that whatever states or processes underlie the face lightness illusion, beliefs and similar attitudes are not among them. Macpherson (2012) suggests accordingly that perhaps the illusion is not mediated by beliefs, but by merely primed cognitive concepts. But recall (from *Chapter 2*) that, in and of themselves, concepts—and unstructured or non-appropriately structured concept bundles—can hardly satisfy the condition of sustaining some logical–rational coherence between source cognitive content and penetrated perceptual content.

The proposal that the cognitive concepts are merely primed hardly helps. For priming is a reflex-like, non-inferential process that is notoriously insensitive to the compositional semantics of structured representations. Consider an experiment in which subjects first watched as an experimenter poured sugar from a single source into two separate clean bottles. The experimenter then asked the subjects to place a label reading "not sodium cyanide, not poison" on a bottle of their choice, and a label reading "sucrose, table sugar" on the other bottle. Amazingly, subjects still rated drinks sweetened with sugar from the bottle labeled "not sodium

[8] Recall from *Chapter 2* that it is at least an oblique issue with respect to cognitive penetrability whether an effect is synchronic rather than diachronic. So the point about the persistence of the face–race lightness illusion is not (merely) that it doesn't immediately succumb to the content of beliefs. More importantly, it would seem that the phenomenon lacks sensitivity to such content altogether.

cyanide, not poison" as less desirable, and they were also more reluctant to take a sip from these drinks (Rozin, Markwith, & Ross, 1990).

This coheres with apparent paradox effects of negative suggestions in hypnosis and advertising, whereby a person who is given a suggestion not to think or feel or do something may as a result be more likely to think or feel or do that thing, respectively (cf., e.g., Cyna & Lang, 2011).[9] A common feature of such phenomena is that they are insensitive, among other syntactic factors, to semantic modifiers such as negative logical operators. Accordingly, priming simply cannot track such differences in content as between the beliefs *that Oswald killed Kennedy* and *that Oswald did not kill Kennedy*.

So, then, say one subject is shown a racially ambiguous face labeled 'white,' whereas another subject is shown the same face labeled 'not white.' Will these subjects perceive the face differently? The priming account predicts that the answer is no. For the labels are expected to exert an identical effect (if any) on perception, in virtue of an identical priming of some concept corresponding to WHITE.

Yet even on Macpherson's (2012) account, "perceptual experience is cognitively impenetrable if it is not possible for two subjects . . . to have two different experiences on account of a difference in their cognitive systems" (p. 29). In our example, one subject believes that the ambiguous face is white, and hence only tokens WHITE. The other subject believes that the face is non-white, and hence tokens NOT WHITE. These are logically contradictory beliefs, with correspondingly different concepts employed. If the priming account predicts that such differences have no bearing on perceptual experience, then the account would not seem to satisfy the condition of semantic–logical–rational coherence.

[9] We shall discuss the hypnotic phenomenon of suggested hallucination and why it is arguably still the best candidate of cognitive penetration in the next chapter.

3. The Face–Race Lightness Illusion

Still, perhaps penetration of perceptual experience by primed cognitive concepts may meet a watered-down criterion of semantic coherence? After all, it would seem, if the priming of WHITE led to white color experience, whereas the priming of BLACK to black color experience, then the effects would be semantically non-arbitrary. But such seeming is misleading: the effects might indeed be carried by arbitrary associations that are inherently insensitive to semantic properties.

Consider the results of some ingenious experiments on unconscious semantic priming by hybrid primes composed of subword fragments. For example, 'angel' and 'warm' both act as evaluatively positive primes. If these words are repeatedly primed, the non-word 'anrm' (composed of *an–* from 'angel' and *–rm* from 'warm') also acts as an evaluatively positive prime. What happens if 'smut' and 'bile' are repeatedly primed? Since both are evaluatively negative, the otherwise positively evaluated word 'smile' (composed of *sm–* from 'smut' and *–ile* from 'bile') itself turns into a negative prime (Abrams & Greenwald, 2000). Semantic insensitivity with a vengeance—or, perhaps even better put, semantic ignorance with a smile.

Here is then a fun experiment which has yet to materialize. Say a subject is repeatedly primed with the color words 'wheat' (light yellow) and 'amazonite' (bright green). Since these words provide the fragments to the hybrid prime 'white' (*wh– + –ite*), an interesting possibility in line with the above is that, upon subsequent priming of the concept WHITE, color experience would be influenced not towards white but towards wheat–amazonite (yellow–green). Is this preposterous? The joke may be on the conceptual-priming view of cognitive penetration.

In any case, lacking appropriate empirical evidence, the issue of whether conceptual priming of cognitive categories may lead to semantically non-arbitrary changes in perceptual experience that are not carried by non-semantic associations is at least moot. Yet semantic non-arbitrariness is widely held to be a requirement of cognitive penetrability on either side of the

debate. So it is hardly trivial that conceptual penetration of perceptual experience would amount to cognitive penetration proper. Hence, contra (3), even if the ambiguous-race face lightness illusion were mediated by cognitive categories, it still wouldn't follow that the illusion constituted a genuine case of cognitive penetration.

3.2.2.2 Weak cognitive penetration

A crucial moral of the previous sub-subsection is that penetration by merely primed concepts may be incapable of sustaining semantic coherence even in a watered-down sense. But say it could. The kind of weak semantic coherence we are assuming, then, is an arational–alogical one which allows for example that CAUCASIAN and NON-CAUCASIAN might both influence color experience toward white. Even if this may not suffice for cognitive penetration on traditional accounts, granted, it would be very interesting if cognitive concepts could influence perceptual experience in such ways. But there is little reason to assume that low-level perceptual experience is thus penetrable. Some made-up examples should help to underscore why.

For example, say I believe that a grayscale image of a face I am presented depicts a Chinese person. As it happens, I associate China with communism, and communism with the color red. Thus, when my concept CHINESE is primed, RED is correspondingly primed. Will I then perceive the face as somewhat red?

That the very question seems absurd suggests that whatever processes underlie the relevant face lightness distortion effect, those processes are stimulus-triggered and modality-specific. In turn, this suggests that the effect is not mediated by relatively abstract concepts such as those of race. It will soon be argued accordingly that the effect is more plausibly mediated by relatively low-level perceptual representations of shape (facial form) and color (lightness).

3. The Face–Race Lightness Illusion

Yet perhaps it may be objected that the above example is misleading because, as opposed to color associates of race, color associates of political–socioeconomic systems or ideologies are irrelevant to the perceived lightness of a face. Indeed they are. But how could the processes that mediate priming take this into account? In particular, how could reflex-like, non-inferential processes that are insensitive to syntactic structure and semantic composition and logical–rational properties assess for and selectively prime concepts based on relevance?

Alternatively, a proponent may object that no one actually holds that it is primed concepts as such that penetrate perceptual experience. Perhaps the assumption is that priming can influence the level or threshold of activation of concepts that uncontroversially contribute to experience when activated. For example, perhaps priming the race concept AFRICAN AMERICAN influences visual experience by enhancing the activation level of the color concept BLACK. Yet, the objection might go, whereas it is independently plausible that BLACK is activated by a grayscale image of a racially ambiguous face, there is little corresponding reason to assume that a grayscale image of a Chinese face leads to stimulus-dependent activation of RED. So even if RED is primed by COMMUNISM upon exposure to a Chinese face, to the extent that RED is only primed but not activated, of course the face is not perceived as somewhat red.

However, say the Chinese face is depicted not in grayscale but in orange–red color, like the cutout figures of Delk and Fillenbaum (1965) mentioned earlier (in *Section 3.2*). If it is plausible that grayscale images can elicit stimulus-dependent activation of color concepts such as BLACK and WHITE, then it is just as plausible that orange–red images can elicit stimulus-dependent activation of color concepts such as ORANGE and RED. And if it is plausible that the priming of AFRICAN AMERICAN can influence visual experience via a modification of the actual activation level of BLACK, then it should be just as plausible that the priming of COMMUNISM can influence visual experience via a modification of the activation level of RED.

There may be no empirical study to date that rules out the above possibility. But hardly anyone would hold their breath. If perceptual experience were really so thoroughly penetrable as the priming account suggests, then it would be a mystery how, following seven decades of research starting with Bruner and Goodman (1947), defenders of cognitive penetrability still have not provided more convincing evidence than the face–race lightness illusion (cf. Firestone & Scholl, 2016a; Machery, 2015).[10]

Hence, even if not conclusive, considerations of the above kind at least raise serious doubts about the empirical plausibility of weak cognitive penetration. In turn, this provides further reason to question the antecedent of premise (3), according to which the ambiguous-race face lightness illusion was mediated by cognitive categories. We already discussed earlier that the illusion is unlikely undergirded by belief-like cognitive states. As we have now seen, the idea of conceptual penetration via cognitive priming is not more plausible. So the worry by now is not merely that we may lack good reason to assume that the relevant effect was cognitive. If the preceding analysis is at least roughly correct, there is also considerable reason to assume outright that this was in fact not the case.

How does this bear on the very conditional that if the relevant illusion was mediated by cognitive categorization, then it constitutes a case of cognitive penetration (3)? As opposed to the truth-functional account of material implication, there is broad consensus that it doesn't follow from a false antecedent that an indicative conditional regarding some empirical matter is true. So reasons to doubt the antecedent of (3) hardly translate to reasons to assume that the conditional as such is true. Indeed, as we saw in the previous subsection, the conditional is

[10] In line with my own earlier advocacy of the effect (e.g., Bitter, 2015a, 2015b, 2017), two recent defenders have proposed that (hypnotically) suggested hallucination is probably the best candidate of cognitive penetration (Arstila, 2017; Newen & Vetter, 2017). But the particular examples provided by these authors are not very convincing (cf. *Subsection 4.6.3*). Further, discussion of the phenomenon in the hypnosis literature has not been explicitly connected to the issue of cognitive penetrability. The next chapter discusses relevant examples and provides a theoretical bridge accordingly.

most plausibly false insofar as it is at least a possibility that the purported type of priming might be carried by non-semantic associations. Others might argue that the conditional lacks truth value if its antecedent is in fact false. Thus, however one looks at it, the third premise of the argument from lexical labeling is hardly beyond question.

3.2.3 Re (1): Lexical labeling vs. visual contrast

The argument so far has been this: Even if the relevant face–race lightness illusion was lexically mediated (1), contra (2a), it still doesn't follow that it was conceptually mediated; and even if it was conceptually mediated, contra (2b), it still doesn't follow that it was cognitively mediated; and even if it was cognitively mediated, contra (3), it still doesn't follow that it involved cognitive penetration. All the same, many defenders of cognitive penetration contend on account of the first premise that it is still abductively more plausible that the illusion involved cognitive penetration (4). Yet is the first premise warranted? Is it really true that whether the racially ambiguous "face was labelled . . . as the face of a white person or the face of a black person . . . alone determined what shade of grey the subjects chose as a match for the lightness of the face" (Macpherson, 2012, p. 48)? Careful (re)consideration of the experimental design provides plenty of space for doubt.

Recall that in the relevant experiment, faces were only labeled during the instruction phase, at which point an ambiguous face always appeared next to an unambiguous black or white face. After instructions, all faces were presented one at a time without label. So whenever a face was labeled, it appeared in the visual context of another face (lexical label → visual context); and whenever a face was unlabeled, it was presented on its own (no lexical label → no visual context). In technical terms, this means that the factor of lexical labeling was confounded with the factor of visual context. This would constitute quite a methodological flaw if the authors' intention had been to test for the effects of lexical labeling.

3. The Face–Race Lightness Illusion

Now Levin and Banaji (2006) advertise in their abstract that judgments of lightness were distorted "even for racially ambiguous faces that were disambiguated by labels" (p. 501). This suggests (although it doesn't imply) that the crucial effect was indeed lexically mediated. Yet the authors themselves formulate more cautiously in a later section of their paper, where they claim only that ambiguous faces were differentiated "on the basis of their context and/or a label" (p. 510). Indeed, the first author even noted in personal communication that "I would be surprised if the absence of a label had much of an effect" (D. T. Levin, July 3, 2012). This suggests that the effect might as well have been mediated solely by visual context.

Hence, for all we know, the relevant perceptual effect might be preserved if an ambiguous face were presented without any label next to an unambiguous white or black face (no label / context). On the other hand, it is a real empirical possibility that the effect would disappear if a labeled ambiguous face were presented on its own (label / no context). These possibilities suggest not merely that the effect may not have been mediated by labeling alone. More importantly, they suggest that the labels may have had nothing to do with the effect.

So, then, not even the very starting premise (1) of the argument from lexical labeling enjoys the support that one would have expected. That the mentioned design flaw went completely ignored until a few years ago (e.g., Bitter, 2013a, 2013b, 2013c, 2014) is especially surprising given that the labeling effect was widely considered to provide most convincing evidence for cognitive penetrability of perceptual experience. In any case, in consideration of the noted methodological oversight, it is unclear that there is any wind left in the argument from lexical labeling.

3.3 A Tentative Account of the Phenomenon

In the preceding section, we discussed various reasons to doubt that the ambiguous-race face lightness illusion is an example of cognitive penetration. In the last subsection, we saw in particular that given the design of the relevant experiment, it is a real empirical possibility that whereas a lexical label is neither necessary nor sufficient, presentation of the racially ambiguous face in the visual context of a racially unambiguous face is both necessary and sufficient for the effect to occur. How could this be so? This section provides a brief sketch of a positive account.

3.3.1 Form–lightness associations

A primary visual diagnostic feature of race is facial form. For example, black faces tend to have thicker lips and a wider nose than white faces, with such features typically somewhere in between for racially ambiguous faces. Another diagnostic feature is of course color: black faces tend to have darker skin tone than white faces, with the skin tone of racially ambiguous faces typically somewhere in between. So why assume that the face lightness illusion is mediated by extra-perceptual categories of race? A much simpler and more plausible assumption is that the effect is mediated by intra-perceptual associations between facial form and color or lightness.

Their earlier remarks and suggestions to the contrary notwithstanding, Levin and Banaji (2006) turn out to argue exactly along these lines in their conclusion:

> This distortion might be considered a case where a set of correlated features mutually facilitate each other such that the presence of most members of the set causes activation of representations of the missing members. So, the correlation between form and shading causes shading features to be activated in the presence of form features. . . . Thus, Black faces might appear relatively dark

> . . . as the result of feature activations resulting from a *perceptual* [emphasis added] classification. (p. 511)

This assumption coheres well with evidence on the so-called memory color effect, of which the face–race lightness illusion is arguably just a special case. The effect involves a modulation by memory colors of the perceived color of subjectively color-diagnostic objects. For example, recall that subjects seem to perceive an orange–red heart shape as if more red (Delk & Fillenbaum 1965), a yellow banana shape as if more yellow (Hansen et al., 2006; Olkkonen et al., 2008), and a blue Nivea tin as if more blue (Witzel et al., 2011) than similarly colored objects of which the respective colors are not diagnostic. Although these effects seem to be object-sensitive, they also appear to be stimulus-driven, modality-specific, and thought-insensitive. The latter features strongly militate against the idea that the memory color effect is cognitively mediated (cf. Deroy, 2013).

Witzel et al.'s (2011) findings are especially telling in this regard. The study only used images of objects that met preset criteria of high subjective color-diagnosticity, as measured by reaction time and accuracy of identifying the typical color of an object. Yet of the 14 images used, only 10 (71%) elicited an observable effect congruent with predictions; of which only seven (50%) were statistically significant; and even among these images there was great variation in the magnitude of the effect.

Still further, the direction of the effect was effectively reversed for some images. The memory color hypothesis predicts that the subjective gray point of an object shifts toward the opposite color of the typical color associated with the object (e.g., toward yellow for a typically blue object such as a Nivea tin). It would only make sense, then, if the subjective gray point of typically red objects shifted toward green. Yet this was not the case. For red-associated objects such as a heart shape, the subjective gray point shifted toward the associated color red.

3. The Face–Race Lightness Illusion

It is very hard to think of any plausible explanation why cognition would penetrate perception in such diverse and unexpected ways. On the other hand, a still further finding that shifts in subjective gray point were particularly large for objects the typical colors of which are close to the daylight axis—which passes from the bluish to the yellowish part of the color space—only underscores that, rather than the influence of cognition, these effects reflect built-in constraints of perception.

So there is good reason to assume that the associations between shape and color or lightness on which the memory color effect depends are intra-perceptually mediated. That the face–race lightness illusion is itself undergirded by intra-perceptual associations is further supported by two recent studies in which images of the prototypical white and black faces were blurred to an extent that most subjects could not tell the race of the faces (Firestone & Scholl, 2015a) or even that the images depicted faces (Lupyan & Lang, as cited in Lupyan, 2015a). Still most subjects judged the respective images differently for lightness.

Granted, such findings do not rule out that race concepts were still unconsciously tokened or primed. Alternatively, as the first author of the original experiments suggests on the basis of some follow-up experiments conducted with a new co-author, the measures of race perception may have been insufficiently sensitive (Baker & Levin, 2016). However, although the phrasing of the authors is at places ambiguous, their overall argumentation is in line with the hypothesis of intra-perceptual form–lightness associations. Consider a key passage from their conclusion:

> the effect of racial categories on lightness judgments does not necessarily reflect the impact of transitory, abstract, online cognitions shaping perception independent of visual form-specifying information. More likely, these effects stem from aggregation of *perceptual links* [emphasis added] between form and

> lightness that can be swayed by *context* [emphasis added]—for instance, by interpreting a racially ambiguous face as Black or White in comparison to a different stimulus. (p. 1995)

Defenders of cognitive penetration may want to read "context" as short for "lexical context," and "interpreting" as short for "cognitive interpreting." But it is clear from the end of the passage that by context the authors have in mind stimulus context. As for interpreting, a related remark is revealing according to which "Repeated experience linking the form of faces to their reflectance could lead to *associations contained within midlevel vision*" (ibid.; emphasis added). This strongly suggests that the alluded-to interpretation is intra-perceptual.

Thus, possible appearances to the contrary notwithstanding, we fully agree: as a result of intra-perceptual form–lightness associations, a visual context of a racially prototypical face may intra-perceptually bias lightness perception of a racially ambiguous face. Intriguing evidence for perceptual plasticity? Sure. Cognitive penetration by race concepts? Hardly.

3.3.2 Lightness contrast

How exactly might differential visual context explain (away) differential lightness perception of racially ambiguous faces? The hypothesis of intra-perceptual form–lightness associations predicts a memory color effect whereby white faces are perceived as lighter than equiluminant black faces, and the lightness of ambiguous faces is perceived as being somewhere in between. If this hypothesis is correct, then presenting an ambiguous face next to a black or white face is like presenting a shade of gray next to a shade of black or white, respectively. Yet it is a classic textbook example of lightness contrast that the very same shade of gray looks lighter/darker in the context of a black/white shade. So given a memory color effect for faces of different races, it is not that surprising if the lightness of an ambiguous face is perceived differently in the context of a black face or a white face.

Note that perceptual contrast effects are a function of the perceived (subjective) and not the actual (objective) reflective properties of the stimuli. So it is a non-issue that the ambiguous and unambiguous faces were objectively identical in luminance. Nor is it an issue that, after instructions, in the trials in which subjects actually judged for lightness, the faces were presented one at a time. For anchoring effects due to simultaneous (within-trial) contrast in the instruction phase may have easily carried over to the first trial. After that, perceptual anchoring and adaptation might both explain successive (between-trial) contrast effects.

Hence, on the present proposal, differences in the judged lightness of ambiguous faces had not so much to do with lexical labeling or racial disambiguation, as with lightness contrast mediated by intra-perceptual form–lightness associations. This hypothesis makes sense of various findings that would be very hard to explain on cognitive terms. For example, it is universally missed that, in the relevant experiment, lightness judgments were distorted only for the apparently lighter face among a racially ambiguous–unambiguous face pair, and hence all distortions tended toward over-lightening. Which is to say, if the ambiguous face was paired with the black face, the ambiguous face was judged as relatively lighter, whereas the black face was correctly judged (*ambiguous*–black). On the other hand, if the ambiguous face was paired with the white face, it was the latter that was judged as relatively lighter, and the ambiguous face was correctly judged (*white*–ambiguous).

The present proposal can easily account for this apparent oddity. For example, lightness contrast may have led to an application by the perceptual system of a highest-luminance rule (Adelson, 2000; Gilchrist et al., 1999), whereby the perceived luminance of the apparently lighter face was anchored to white. Of course, not much hangs on whether this was the actual rule or mechanism underlying the effect. The important point is that insofar as the relevant

distortions were biased in such eccentric yet consistent ways, an intra-perceptual account of the phenomena is indeed much simpler and more plausible than a cognitive account.

3.4 Conclusion

The conclusion of this chapter is that the face–race lightness illusion is an unconvincing case of cognitive penetration. A prima facie powerful argument suggests the contrary based on the contention that differential lightness judgments for racially ambiguous faces were a function of differential lexical labeling. Yet, on closer scrutiny, the premises of the argument are neither empirically, nor theoretically well supported. Accordingly, the phenomena is proposedly more easily and plausibly explained by intra-perceptual than cognitive mechanisms.

The proposed intra-perceptual account is clearly more parsimonious than a cognitive account, for it does not assume the involvement of any lexical or cognitive representations. It is also empirically more plausible insofar as it does not posit any as-of-yet unestablished or controversial psychological mechanisms. Finally, it is abductively preferable insofar as it can explain everything that a cognitive account can, but even more. For example, the proposal can easily account for why the face–race lightness illusion persists despite contrary beliefs, or why the effect was selective and unidirectional in the ambiguous-race face experiment.

The upshot is that memory color effects like the face–race lightness distortion effect might be sufficiently explained by the internal workings of cognitively impenetrable perceptual systems. This suggests that perceptual systems are more plastic and hence psychologically richer than is commonly assumed. Importantly, and contrary to a currently widespread view in philosophy (cf. *Section 2.2*), it also suggests that the significance of cognitively impenetrable systems is not confined to an unconscious level.

For example, consider the apparent epistemic threat posed by cognitive penetration with respect to the role of perception in the justification of beliefs. If our beliefs can penetrate our perceptual experience, which we in turn consider as evidence for our beliefs, then the structure of belief formation would seem to be circular (Siegel, 2012). By securing the cognitive impenetrability of low-level perceptual experience, a much-feared epistemic menace may thus be warded off (cf. also Fodor, 1984, 1985, 1988).

It bears emphasis that not all accounts of cognitive impenetrability circumvent the threat of circularity. For example, recall that a common strategy of explaining away alleged effects of cognitive penetration is in terms of perceptual misjudgment (cf., e.g., Firestone & Scholl, 2016a). One might thus maintain that perceptual experience remains intact in the face of cognitive biases in judgment. But if a subject is genuinely deluded about his or her perceptual experience, then she or he might just as well consider her or his mistaken judgments as evidence for beliefs that were the source of the bias in the first place.

A further advantage of the proposed intra-perceptual account, then, is that it can circumvent the epistemic threat of circularity without the undesirable implication that we are possibly way more often deluded about our first-person phenomenal experience than epistemologists and psychologists would have it (not to mention regular folk). Of course, the proposal still suggests that we may often be victims of perceptual illusions. So certain epistemic worries remain. But at least these are not the worries of circular belief formation and introspective bias.

Finally, one may wonder whether or how the present discussion extends beyond for example those of Deroy (2013) and Firestone and Scholl (2015a). Deroy provides a thorough analysis of the memory color effect, but she doesn't address the face–race lightness illusion.

Firestone and Scholl address the effect head on for racially prototypical faces, but they have nothing to say about the effect for racially ambiguous faces.

As noted earlier in the chapter, most defenders of cognitive penetrability are in principle prepared to grant that standard memory color effects such as that orange–red heart shapes are perceived as more red than similarly colored non-red-associated shapes are intra-perceptually mediated. These defenders are correspondingly ready to grant that the face–race lightness illusion for prototypical white and black faces affords a non-cognitive explanation. Yet many defenders still argue that the illusion for ambiguous faces cannot be so explained. Thus the importance of addressing the relative weaknesses of the latter argumentation, and of showing how fairly simple visual mechanisms may account for the target phenomena.

Of course, debunking evidence and arguments that defenders of cognitive penetrability do not find most crucial is still important progress. But it is less likely to make serious headway in the overall debate. So it gives special weight to the present analysis that it specifically addresses what many defenders themselves claim (or at least have claimed until most recently) to be most convincing evidence for cognitive penetration. Given this dialectic, the present challenge cannot be swept aside on account that it merely casts doubt on a single case.

Indeed, the tides may already be changing. For example, whereas Vetter and Newen (2014) argued at length that the face–race lightness illusion constitutes most convincing evidence for cognitive penetrability, in their most recent defense of cognitive penetrability, they do not even mention the phenomenon (Newen & Vetter, 2017). It is also encouraging that Gross et al. (2014), Machery (2015), Macpherson (2015), and Varga (2017) have all taken note of some or another of the discussed arguments (Bitter, 2013a, 2014). Hopefully, the present analysis thus only further facilitates a reassessment of the face–race lightness illusion as a case not of cognitive penetration but intra-perceptual learning and perceptual contrast.

4. Suggested Hallucination

The previous chapter concluded that even a most widely proposed candidate is ultimately implausibly a case of cognitive penetration of perceptual experience. The empirical likelihood of any actual occurrence of cognitive penetration would thus seem correspondingly low. Yet a prima facie most astonishing phenomenon called suggested hallucination has gone virtually unnoticed in the dedicated debate, with discussion of the effect mostly limited to the literature on hypnosis.

The first part of this chapter argues that if there ever was a strong candidate of cognitive penetration, it is this phenomenon. Following a brief introduction of hypnosis and suggested hallucination (*Section 4.1*), it is argued in particular that contrary to widespread skepticism, the phenomenon is not a sham (*Section 4.2*); as opposed to behaviorist theorizing, the source of the effect is in relevant ways at least in part cognitive (*Section 4.3*); and even if it has an indubitable role, attention does not fully mediate all effects (*Section 4.4*). Thus, as opposed to most candidates of cognitive penetration, the phenomenon cannot be sufficiently explained away by demand compliance, response bias, intra-perceptual processes, or selective attention.

Still, the central claim is, suggested hallucination is not a case of cognitive penetration. The latter part of the chapter surveys evidence accordingly that the effect does not involve relevant changes in early perceptual processing (*Section 4.5*). Nor, it is argued, is it plausible that subjects experience relevant distortions in perceptual experience (*Section 4.6*). Further, and contrary to even a deflationary account of the phenomenon, there is reason to doubt that subjects entertain genuine misbeliefs about the actual stimulus state of affairs (*Section 4.7*). Hence, despite the robustness of what is truly a jaw-dropping psychological phenomenon, neither perception nor perceptual belief is cognitively penetrated.

So what explains the effect? The defended view is that a core feature of relevant cases is a distortion in perception-related affect (*Section 4.8*). For example, some people may on occasion experience the sensory features of pain without the usual concomitant feelings of unpleasantness. Suggested analgesia may thus alleviate the affective–motivational component of pain even if the sensory–perceptual component of pain is cognitively impenetrable.

Similarly, some subjects may on occasion perceive the color of a blue image accurately but still feel as if the image were red. This may resemble what is called displacement in psychodynamic theories, whereby the emotion associated with some thought or desire (etc.) is detached from the original content and attached to the content of some other thought or desire (etc.). If this view of suggested hallucination is correct, it helps to explain away a large number of alleged cases of cognitive penetration without implying that a relatively significant portion of subjects in various perception experiments are deluded or merely complying with demands.

An important upshot is that various debates related to cognitive penetrability rest on some or another mistaken assumption (*Section 4.9*). For example, some argue for the existence and penetrability of high-level perceptual experience. Others contend that high-level perceptual experience does not exist, or that alleged cases of its penetration are not cognitively mediated, or that they in fact involve penetration of cognitive phenomenology. An apparent implicit assumption in these debates is that if some experience in addition to or as opposed to low-level perceptual experience is cognitively penetrable, then either high-level perceptual experience or cognitive phenomenology or both exist and is/are cognitively penetrable. The affective interpretation of suggested hallucination underscores that even if both the antecedent and the consequent of this conditional are true, the conditional as such is still false. For at least in some relevant cases, it is plausibly affect rather than perception or cognition that is cognitively or otherwise penetrated.

4.1 Hypnosis and Suggested Hallucination

When using hypnosis, typically, one person (the subject) is guided by another (the hypnotist) to respond to suggestions for changes in subjective experience, alterations in perception, sensation, emotion, thought, or behavior (Green, Barabasz, Barrett, & Montgomery, 2005). For example, it may be suggested that a subject's hand feels so light that it will move spontaneously upward (positive motor suggestion for hand levitation); or that the arms become so rigid that they cannot be bent (negative motor suggestion for arm catalepsy); or that there is a buzzing mosquito in the vicinity (positive cognitive suggestion for auditory hallucination); or that one will lose his or her sense of smell (negative cognitive suggestion for anosmia hallucination).

Such invitations to experience the world differently can also be delivered without hypnosis. However, many subjects manifest increased susceptibility to suggestion following a hypnotic-induction procedure (Elkins, Barabasz, Council, & Spiegel, 2015; Wagstaff, 2014a, 2014b). These procedures basically involve an extended initial suggestion to encourage responses to further suggestions (Green et al., 2005), and aim to induce a state of consciousness involving focused attention and reduced peripheral awareness (Elkins et al., 2015).

For example, a standard procedure is to ask a subject to fixate on a small bright object, as the hypnotist talks about paying close attention to his or her words, relaxing and just letting go, and entering a state resembling sleepwalking whereby the subject is not quite awake but nevertheless still capable of doing many of the things that people do while awake. By the end of this brief introduction, some people spontaneously close their eyes. Others are given further suggestion that their eyes are getting tired, that they will strongly wish their eyes were closed, that they should let their eyes close by themselves, and so on (Stanford Hypnotic Susceptibility Scale, Form C [SHSS:C], Item 0: Induction by Eye Closure; Weitzenhoffer & Hilgard, 1962).

4. Suggested Hallucination

Now as opposed to intrinsic research in hypnosis, which focuses on a proprietary domain of phenomena and their theoretical underpinning, instrumental research uses hypnosis as a tool to investigate phenomena outside its immediate domain (Barnier, Dienes, & Mitchell, 2008; Cox & Bryant, 2008). Of much relevance to the cognitive-penetrability debate, research of the latter kind inquires for example whether color perception can be modulated in a top-down fashion by suggestions. Consider a path-breaking neuroimaging study that used a technique called positron emission tomography (PET) to measure regional cerebral blood flow (rCBF) in eight highly hypnotizable subjects both in and out of hypnosis (Kosslyn, Thompson, Costantini-Ferrando, Alpert, & Spiegel, 2000).

In the hypnosis condition, subjects were shown either a multi-colored image or a grayscale image, which they were instructed to perceive either veridically or as altered. In the latter case, subjects were asked to "drain" color from the colored image and perceive it as if grayscale (color draining), and to "add" color to the grayscale image and see it as if colored (color adding). In the no-hypnosis condition, subjects received similar instructions. However, instead of asking them to see the color image as if grayscale and the grayscale image as if colored, they were only asked to "remember and visualize" the images in their other form. The no-hypnosis condition thus served as a mental-imagery control for suggested hallucination in the hypnosis condition.

The results indicated that in the hypnosis condition, irrespective of which image was shown (colored vs. grayscale), activity in a classic color area in the fusiform region of the brain increased in both hemispheres if subjects were asked to perceive color. Correspondingly, the same area showed a bilateral decrease in activation when subjects were asked to perceive an image in grayscale. In contrast, in the no-hypnosis condition, a similar pattern of increased and decreased activity in the respective area was only observed in the right hemisphere.

4. Suggested Hallucination

These results have been multiply conceptually replicated, using alternative neuroimaging (McGeown et al., 2012), electrophysiological (Kallio & Koivisto, 2016), and subjective measures (Mazzoni et al., 2009). The first of these studies even extended the findings insofar as highly suggestible subjects were almost equally responsive to hallucination (vs. imagery) suggestions in and out of hypnosis. Though the second study did not confirm this, it was consistent with earlier studies according to which "hypnotic virtuosos" (who are extremely suggestible under hypnosis) can perceive a blue image as if red and vice versa (Kallio & Koivisto, 2013; Koivisto, Kirjanen, Revonsuo, & Kallio, 2013).

Recall that defenders of cognitive penetration only posit for example that, as a function of one's beliefs regarding the typical skin tone of black (African American) vs. white (Caucasian) faces, one may perceive a grayscale image of a black face as if somewhat darker than an equiluminant grayscale image of a white face. Or, similarly, because one believes that heart shapes but not circle shapes are typically red, one may see an orange–red heart shape as somewhat more red than a similarly colored circle shape.

It is not part of such assumptions—to the contrary, it is generally denied—that a subject might perceive a grayscale image in color, or a colored image in grayscale or another color, if some suggestion to that end is delivered. However, many studies of hypnosis and suggestion hint that at least highly suggestible subjects may achieve exactly such feats. Perhaps cognitive penetrability is not only a still open empirical possibility, then, but it may strike with vengeance in a most extreme form imaginable.

Of course, even if some people may experience such dramatic distortions in perception in some situations, it doesn't follow that all or even most people could ever even come close to having similar experiences. Indeed, as is the case with intelligence and personality traits, there appears to be huge inter-individual variation in the level of suggestibility and

hypnotizability in the general population. Depending, among other factors, on the context (hypnotic, non-hypnotic), demands (task-motivational, honest reporting, etc.), and type (positive, negative; etc.) and modality of hallucination (visual, auditory, etc.), only about 2.5% to 5% of subjects meet stringent criteria for suggested hallucination (Barber, 1970; Hilgard, 1965). Still, if every twentieth to fortieth person can really experience perceptual effects of this kind and magnitude, the phenomenon's potential for providing mind-blowing existential proof of cognitive penetrability justifies careful examination in its own right.

Finally, even if suggestions for hallucination do not lead to gross perceptual distortions in most cases, a relatively large number of cases may still involve relevant changes in experience. This point bears special emphasis in relation to that skeptics are generally even more suspicious of extravagant as opposed to more modest effects (Spanos, 1991; Wagstaff & Cole, 2005; Wagstaff, Toner, & Cole, 2002). So, even from a dialectic perspective, a relative rarity of jaw-dropping cases hardly lessens the case for suggested hallucination as a prime candidate of cognitive penetration.

4.2 Psychological Reality of the Phenomenon

How did such an exciting phenomenon evade the attention of most participants of the cognitive-penetrability debate?[11] After all, there is hardly a lack of literature on the topic. As an illustration of the magnitude, already some four decades ago, in the short few years between the first and second volumes of a most important edited book on hypnotic phenomena (Fromm & Shor, 1972, 1979), more than 1,000 publications appeared in the field of hypnosis in total (as cited in Barnier & Nash, 2008).

[11] As mentioned in the previous chapter, in the recent debate, aside from the present author (e.g., Bitter, 2015a, 2015b, 2017), only Arstila (2017) and Newen and Vetter (2017) seem to have taken notice of suggested hallucination as a prima facie prime candidate of cognitive penetration. The latter authors argue that the phenomenon is indeed a real case of cognitive penetration. The present chapter ultimately argues to the contrary.

Nor is it the case that the implications of suggested hallucination for more general issues regarding perception and imagination and cognition are immaterial to researchers of hypnosis. On the contrary, there is wide agreement that "any satisfactory theory of hypnosis should also be a theory bearing on psychology at large" (Hilgard, 1991, p. 101), with corresponding and increasing interest in using hypnosis as a tool for investigating phenomena outside of the domain of hypnosis (Cox & Bryant, 2008; McConkey, 2008).

Accordingly, some recent empirical papers explicitly emphasize for example that "hypnosis can alter the functioning of automatic visual processes" (Kallio & Koivisto, 2016, p. 261), and that "Conscious color experience . . . can be modulated, at least in some individuals, by top-down factors such as hypnotic suggestions" (Koivisto et al., 2013, p. 1).

Beyond sheer ignorance, a further possible factor in explaining widespread disregard of such research is general skepticism about the very reality of hypnosis.[12] In turn, such skepticism may stem from common cultural associations of hypnotism with occultism and stage hypnosis; a relatively slow historical track record of separating myths and facts; various misconceptions and misrepresentations of theoretical issues and the state of hypnosis science; and general worries regarding the prospect of obtaining sound objective evidence of relevant subjective phenomena.

Subsection 4.2.1 briefly addresses the first two of these issues, namely, the sociological issue concerning general cultural associations, and the historical issue of separating myths and facts. *Subsection 4.2.2* focuses on the third batch of issues, in particular on why it is conceptually and theoretically important to distinguish skepticism about the psychological

[12] Of course, skeptics of hypnosis may allow that hypnotic subjects engage in vivid mental imagery. However, they doubt that such activity ever leads to relevant changes in perceptual or cognitive or affective experience. By implication, those who hold that perceptual experience is indirectly penetrable by cognition via mental imagery arguably qualify as non-skeptics of suggested hallucination.

reality of suggested hallucination (real–fake debate) from skepticism about whether the phenomenon involves relevant distortion in perceptual experience (hallucination–delusion debate), and whether the effect is explained by some special psycho-neurophysiological state (state–nonstate debate). *Subsection 4.2.3* concludes the discussion of the psychological reality of suggested hallucination by addressing the empirical–methodological issue as to whether or how simulating subjects may be objectively distinguished from "real" hallucinators for whom the effects of relevant suggestions have a subjectively compelling quality.

4.2.1 Common cultural associations and the separation of myths and facts

The association between hypnosis and occultism dates back at least to the 18th century, when Anton Mesmer claimed to have discovered a new natural force in the form of a universally diffused fluid which he dubbed animal magnetism. This force or fluid was alleged to serve as "the vehicle of a mutual influence between the celestial bodies, the earth and the bodies of animate organisms" (Franklin et al., 1784/1785, p. 20). Most importantly, if properly channeled with a metal rod, it would supposedly give rise to drastic effects such as convulsions and other phenomena later more commonly associated with hypnosis.

Whether Mesmer really believed any of this is unclear, but many will know that a commission charged by the King of France and led by Benjamin Franklin (Franklin et al., 1784/1785) famously debunked animal magnetism as non-existent. What seems often lost in this exemplary story of prevailing science, though, is that the Commission did not question the reality of the elicited phenomena. Rather, by using the method of differences, it eloquently demonstrated that animal magnetism could not account for the observed effects, whereas other factors such as touching, imitation, and especially imagination plausibly could. So, contrary to a common misunderstanding, mesmerism qua hypnotism was not shown to be a hoax by the

Commission. The domain was merely demystified, meanwhile an important precedent was set regarding the application of proper scientific methods within the field.

As opposed to mesmerism, where a supposedly real phenomenon was given a false explanation, stage hypnosis is essentially a form of entertainment that is known to have included at least on past occasions the use of tricks and stooges among other feats to ensure success of the show (Meeker & Barber, 1971). Relatedly, it is not implausible that some naïve participants on stage conceal their failure to experience relevant effects due to perceived social pressure from the hypnotist or the audience or both, or perhaps because of a simple desire to show off (Wagstaff, 2004). However, it is important that much supporting evidence for such claims comes from credulous rather than skeptical research on hypnosis. Accordingly, separating the wheat from the chaff should not be mistaken for casting doubt on the very existence of relevant phenomena.

This is not to suggest that the study of hypnosis has been historically quick to separate either scientific or lay myths from facts. Consider for example the very term "hypnosis," coined in the 19th century from *hypno-* "sleep" and *-osis* "condition." Though hardly anyone ever thought that hypnotized subjects were literally asleep, it took until the last quarter of the 20th century to empirically falsify the scientific view that hypnosis essentially involves a passive, drowsy, sleep-like state (Banyai & Hilgard, 1976).

A foremost figure of hypnosis research in the latter half of the 20th century relatedly emphasized not long ago that "Historical accounts make note of the fact that at different times, hypnosis was construed as animal magnetism, nervous sleep, conditioned responses, prestige suggestion, dramaturgic skills, role-taking, absorption, expectancies, dissociation, and entrancement" (Sarbin, 2005, p. 129). This may raise suspicion that experts in the field even today have little if any inkling as to the exact nature of the phenomena. However, once again,

debating some proposed explanation of relevant effects is not equivalent to debating the reality of those effects.

There is great disagreement about what mediates hypnotic and suggested phenomena, sure. But if by the reality of some relevant effect we understand that at least not all subjects are faking or simulating or merely complying with demands, then there is in fact virtually unanimous consensus in the field that at least some phenomena are real (e.g., Kihlstrom, 2008; Lynn, Kirsch, & Hallquist, 2008; McConkey, 2008; Sheehan & Perry, 1976/2015; Spanos, 1986; Spanos & Barber, 1974; Wagstaff, 2014b; Wagstaff & Cole, 2005). Let us now turn to this issue and how it differs from some theoretically distinct even if often insufficiently demarcated issues.

4.2.2 Varieties of skepticism

We have been uncovering some of the reasons why, despite its being a prima facie most striking example of cognitive penetration, suggested hallucination has received so little attention outside the field of hypnosis so far. Beyond sheer ignorance about relevant research, a possible reply is that many still view hypnosis and related phenomena with suspicion. This may be partly due to general cultural associations of hypnosis with the occult, and its historical association with mesmerism and poor science. However, as noted in the end of the previous subsection, there is near-universal if not universal agreement in contemporary hypnosis research that at least some hypnotic phenomena are genuine insofar as not faked.

Why might it still appear as if it were controversial even in current hypnosis science whether suggested hallucination is real or sham? A possible reply is that the real–fake issue is still often conflated with issues such as whether the phenomenon is better construed as a distortion of perceptual belief than a distortion of perceptual experience (hallucination–delusion debate), and whether such effects involve some special mental state (state–nonstate

debate). The present subsection demarcates these issues to underscore that skepticism about one or another specific feature of suggested hallucination does not entail general skepticism about the psychological reality of the phenomenon.

4.2.2.1 Real vs. fake

If hypnosis research had never seriously addressed the issue of whether suggestion effects are real or fake, or at least the issue were still a matter of fervent debate, skepticism about relevant phenomena would be warranted. Yet the question was effectively posed by the first generation of modern researchers, and the answer seems to be that at least some hypnotic responding is real. Accordingly, since many decades, "the question of sham behavior no longer constitutes an issue in the hypnosis literature" (Sheehan & Perry, 1976/2015, p. 271), as "all thoughtful theorists would agree that hypnotic behavior is not simply faking" (McConkey, 2008, p. 68).

Perhaps only one participant of the debate dissented some decades ago that, at the time, there was still considerable controversy regarding the importance of compliance or sham behavior (Wagstaff, 1978, 1991). However, even this theoretician has increasingly emphasized that good hypnotic subjects at least try to experience what is suggested or expected of them, by applying various cognitive strategies to make the experiences veridical or believable (Wagstaff, 2004). Most recently, the same researcher also granted that "a growing volume of studies, especially in the neuropsychological literature, have found results consistent with the view that responses to hypnotic suggestions, including those for hallucinations and analgesia, may be experienced veridically" (Wagstaff, 2014b, p. 103).

Traditional proponents of a social-psychological or socio-cognitive interpretation of hypnosis also emphasize that "Of course suggestions can produce alterations in subjective experience" (Spanos, 1986, p. 489). So, for example when role theorists posit that hypnotic responding involves role-taking (e.g., Sarbin, 1950; Sarbin & Coe, 1977) or role enactment

(e.g., Spanos, 1986), they are not thereby suggesting that subjects are merely playing a role. Consider the analogue of taking a parent role or presenting oneself as a parent: it doesn't follow that one is merely acting or pretending to be a parent (Lynn et al., 2008). Accordingly, on a social-psychological view, "in most hypnotic situations the experiencing of the suggested effects is the major feature of adopting the hypnotic role and thereby in presenting oneself as hypnotized" (Spanos, 1986, p. 491).

Granted, the use of dramaturgical metaphors, allusions to misattributed experience, and a general attitude of debunking has not helped to dispel the misimpression that role theorists deny the reality of hypnotic phenomena (Kihlstrom, 2008). A further possible source of confusion is a frequent conflation of distinct—even if related—empirical and theoretical issues even by prominent researchers both in and out of the field. To clarify what is exactly at issue regarding the reality of the phenomena, a quick survey of two such issues may serve helpful.

4.2.2.2 Hallucination vs. delusion

In the early years of modern hypnosis research, a paper on suggested hallucination opened with the question: "Does the hypnotic subject 'actually experience' the suggested 'hallucinatory' object or does he just state that he sees it in order to please the hypnotist?" (Barber, 1959a, p. 136). In line with two highly influential papers that appeared soon after, we may construe the actual-experience view as credulous, and the pleasing-view as skeptical (Sutcliffe, 1960, 1961).

However, as the latter papers rightly pointed out, these positions do not cover the full theoretical landscape. For example, what if hypnotic subjects entertain the reality of suggested phenomena without undergoing corresponding hallucinatory experiences? We might construe this hybrid option as the delusion view (cf. also Barnier & Nash, 2008; Cox & Bryant, 2008; Kihlstrom, 2008; Sidis, 1906; White, 1941; Woody & Barnier, 2008).

4. Suggested Hallucination

So say a subject passes a suggestion to perceive a blue image as if red. On the credulous view, this is explained in part by the elicitation of some hallucinatory experience as of red. On the skeptical view, the subject merely behaves and reports as if perceiving red. The delusion view accords with the skeptical view in denying any quasi-perceptual experience as of red. However, it conforms with the credulous view in at least granting to the subject an "emotional conviction" (Sutcliffe, 1961, p. 200) that the blue image appears red.

Interestingly, Sutcliffe ultimately construed the delusion view as a version of the skeptical view (the other version being the simulation view). So on his understanding, suggested hallucination qualifies as real only if it involves perception-like experience. This view seems to contrast with both the contemporary and earlier mainstream position, on which "It is the subject's conviction that the suggested event is really happening that distinguishes a genuine hypnotic experience from overt behavioral compliance" (Kihlstrom, 2008, p. 32).

Although the latter formulation may not do full justice for example to pseudo-hallucinations (the non-veridical nature of which subjects are aware, yet which are not cases of mere behavioral compliance), the general point is clear. To the extent that a subject genuinely entertains some suggested state of affairs to be real, corresponding responding can also be considered real, irrespective of whether any quasi-perceptual experience accompanies the effect (cf. also Johnson, Maher, & Barber, 1972; Tellegen, 1978; White, 1941).

It bears emphasis relatedly that, similarly to the credulous view, on the delusion view, it doesn't follow from the inaccuracy of some verbal report that the reporting subject is faking or lying or merely complying with demands. As an analogue, consider a hypothetical debate about who could be said to have a real Napoleon identity. One side may hold that only the actual historical figure, meanwhile opponents might argue that any person who honestly identifies as such. In either case, it doesn't follow from an assertion that one is Napoleon that

the asserting person is in fact Napoleon, or else an impostor or actor etc. The person may also be deluded, and hence factually inaccurate despite his or her honest reporting.

Hence, it is a mistake to conflate the issue of whether suggested hallucinations are real or fake (and so whether subjective reports are honest or dishonest) with the issue of whether real effects involve actual perceptual hallucination or only some form of delusion (and hence whether subjective reports are accurate or inaccurate). Of course, if all subjects are in all cases dishonest (faking), then the issue of accurate reporting (hallucination vs. delusion) does not arise. But, it is argued, they are not, and so it does. Thus, even if credulous views are ultimately mistaken, it would be a mistake to disregard the reality of the phenomena on the basis of traditional skeptical considerations of simulation and compliance.

4.2.2.3 State vs. nonstate

Writing on the state of hypnosis science for a general audience in a pre-eminent newspaper thirty years ago, a prominent psychologist from outside of the field opened with this statement: "A fierce scientific debate is focusing on whether hypnosis involves a special state of awareness or, as challengers claim, is simply one person seeking to please another by cooperating with suggestions made to him" (Goleman, 1987, para. 1). About a decade and a half later, writing for a narrower audience in a leading psychiatric journal, a mixed group of experts and non-expert psychologists started their highly influential paper similarly:

> Hypnosis has a long and checkered history; at times it has been accepted as a genuine psychological state with unique consequences, and at other times it has been treated as nothing more than stage-show gimmickry. In contemporary psychology and psychiatry, these two views have crystallized in a debate. On one side are those who claim that hypnosis is a distinct psychological state . . . that allows one . . . [among other things] . . . to marshal one's resources in

> unusual ways. On the other side are those who equate hypnosis with acting, claiming that it is simply a role that people can adopt in which they cooperate with the wishes of the hypnotist. (Kosslyn et al., 2000, p. 1279)

Such misrepresentations of the so-called state–nonstate debate are incredibly confused even if not uncommon. To start with, recall that proponents of role theory do not equate hypnosis with acting or complying—no contemporary theorist does. Further, no one takes issue with talk of a hypnotic state if state is understood in a neutral sense, such as when referring to the state of the economy (Kallio & Revonsuo, 2003), or if used as a shorthand for denoting the effects of certain suggestions in a hypnotic context (Kirsch et al., 2011). Accordingly, as an editorial commentary noted long ago in relation to a social-psychological interpretation of hypnotic phenomena:

> it would seem that objective evidence—if it were possible to have it—that subjects actually *experience* analgesia or amnesia (as opposed to feigning it) would be evidence of a "special state" irrespective of whether the experience was induced by social expectations, imaging ability or hypnotic suggestion. (Harnad, 1986, p. 489; emphasis in original)

What is really at issue in the state–nonstate debate then? The primary question is whether "the hypothetical construct hypnotic state or trance state is a useful conceptual tool for organizing the available data and for generating fruitful research in this area" (Spanos & Barber, 1974, p. 500). This is not a descriptive question about experience or the product of hypnotic procedures. It is an explanatory question about the underlying factors that mediate the effects. Accordingly, when nonstate theorists express doubt about the existence of some special state, they are not thereby denying the reality of the area of inquiry. What they variably argue is that the data can be more parsimoniously explained "in terms of processes familiar to most

psychologists such as suggestibility, placebo effects, motivation, imagination, absorption, shifts in attention, and compliance with instructions" (Wagstaff, 2004, p. 85; cf. also Hilgard, 1992; Kihlstrom, 2008).

Historically, the association of hypnosis with a special state was closely related to the view that hypnotic behaviors involve highly unusual or transcendent phenomena (Spanos & Barber, 1974). This is not the current view. Non-hypnotic subjects can produce virtually any hypnotic effect if properly motivated (Barber, 1970; Hilgard, 1973; McGeown et al., 2012; Meeker & Barber, 1971; Miller & Bowers, 1993; Orne, 1959; Raz, Kirsch, Pollard, & Nitkin-Kaner, 2006; Spanos, 1986). So what are current defenders of the state view arguing for? Their main claim is that real hypnotic subjects still achieve what they do in a qualitatively different way (e.g., Miller & Bowers, 1993).

Of course, insofar as a hypnotic response is faked or simulated, nonstate theorists may agree with the qualitative-difference claim. More to the point, some even allow that genuine hypnotic effects may differ in magnitude or extent of response. So they have no in-principle issue with the idea of quantitative boost. What they deny is that any such quantitative difference is better explained by positing some qualitatively special state (Sarbin, 1950; Spanos & Barber, 1974; Wagstaff, 2014b; White, 1941).

Hence, in order to appreciate the state–nonstate debate, it serves to keep in mind distinctions such as between state as explanandum vs. explanans; hypnosis as product vs. process; and quantitative vs. qualitative differences. Note, further, that there is in fact a variety of state views, with a corresponding variety of nonstate views. For example, some state theorists posit that a hypnotic state is necessary for at least difficult suggestion effects like hallucination (e.g., Kallio & Revonsuo, 2003), whereas others allow that hypnotic and non-hypnotic states may both enable such effects (e.g., McGeown et al., 2012).

Other examples concern whether a formal induction is necessary and/or sufficient for hypnosis. Many state theorists deny the necessity of induction on account that some people occasionally spontaneously slip into a hypnotic state (e.g., Barabasz & Barabasz, 2015; Gruzelier, 2005). Spontaneous hypnosis is not identical with neutral hypnosis, which is conceptualized as an induced hypnotic state without any suggestion effect. Those who deny the sufficiency of induction thereby deny occurrences of the latter kind (cf. Kihlstrom, 2008; Kirsch et al., 2011).

Finally, the issue of whether hypnotic phenomena involve some special state is at least logically distinct from the issue of whether the domain of hypnosis constitutes a natural kind. The latter holds only if all and only hypnotic phenomena share some common core component. However, say the phenomena indeed share some common core, as many state theorists avow (e.g., Gruzelier, 1998). This component may still be shared by phenomena associated with other special states, such as meditation, sleepwalking, or drug trance. On the other hand, say nonstate theorists are correct in that there is no single, generally shared component. Smaller clusters of phenomena that correspond to subdomains of hypnosis may still each exhibit its own unique component that is unshared by other phenomena. Surely this would still please many state theorists (e.g., Kihlstrom, 2008).

The upshot is that, although confusions abound, it really doesn't matter which side of each variant of the state–nonstate debate one is on. For all such issues are orthogonal to the real–fake question. Accordingly, akin to proponents of the delusion view, who already presume the falsity of a simulation or compliance account, nonstate theorists themselves explicitly emphasize that "all contemporary nonstate theorists agree with state theorists that hypnotic phenomena involve genuine changes in the subject's experience that cannot be explained away in terms of faking or sham behavior" (Spanos & Barber, 1974, p. 508).

Hence, "No researcher with any experience would deny that many hypnotic subjects experience ASCs [altered states of consciousness]" (Wagstaff & Cole, 2005, p. 16), if such states are understood to constitute an area of inquiry rather than the explanatory constructs of some theory. Indeed, it is hard to see why else proponents of a nonstate view would engage in relevant research and debates for decades on. And if nonstate theorists are ultimately right in that hypnotic phenomena are qualitatively no different from ordinary, non-hypnotic phenomena, then that only provides further reason to take suggested hallucination seriously as a foremost candidate of cognitive penetration.

4.2.3 Empirical dissolution of traditional skepticism

Let us recap: On the face of it, the phenomenon of suggested hallucination is an astonishingly strong candidate of cognitive penetration. For example, it would seem that at least some people in at least some conditions may even perceive a blue image as if red. How has research on such an effect not penetrated debates about cognitive penetrability? A plausible even if only partial answer is that hypnosis and related phenomena are still not taken seriously. However, despite its checkered past and general cultural portrayal, as well as long-standing misconceptions and conflation of issues even within the field, there is unanimous consensus among current experts that at least some suggestion effects are not faked.

Yet perhaps current experts are themselves mistaken? Admittedly this wouldn't be the first time in a slow and arduous process of self-correction. A reason for suspicion in this regard among scientists outside the field relates to the issue that "at its heart, hypnosis is a private experience" (Barnier & Nash, 2008, p. 3). Needless to say, private experience is not directly accessible from a third-person perspective. Recall, however, that contrary to an earlier view, hypnotic subjects do not seem to transcend their capacities beyond levels that highly motivated non-hypnotic subjects can also reach. Accordingly, and most importantly, at least with proper

training, virtually any effect might be faked or simulated (e.g., Meeker & Barber, 1971; Orne, 1959).

A scientific-minded skeptic may then argue as follows: Insofar as there is no relevant objective difference between "real" and simulated suggestion effects, it is abductively most plausible that there is no relevant subjective difference, either. Yet to say that it is abductively most plausible that there is no relevant subjective difference between "real" and simulated suggestion effects is tantamount to saying that it is abductively most plausible that "real" suggestion effects are themselves simulated and hence lacking psychological reality. Hence, insofar as there is no relevant objective difference between "real" and simulated effects, it is abductively most plausible that there is no psychological reality to the phenomena.

The skeptical challenge, then, is to show that even if any effect may be faked in principle, there are still relevant objective differences between hypnotic "reals" and simulators. The remainder of this section provides an extensive (though non-exhaustive) survey of how this challenge has been met. *Sub-subsection 4.2.3.1* introduces the real–simulator paradigm of experimental hypnosis research. *Sub-subsection 4.2.3.2* discusses some real–simulator differences in retrospective first-person accounts of hypnotic experiences. *Sub-subsection 4.2.3.3* presents three behavioral distinguishers of simulators: overplaying, performance interference effects, and incongruent responding. *Sub-subsection 4.2.3.4* in turn presents some somatic–physiological and neurophysiological differences between "reals" and simulators.

4.2.3.1 The real–simulator paradigm

The real–simulator paradigm was introduced into hypnosis research more than half a century ago to provide a quasi-control of the demand characteristics of the hypnotic situation (Orne, 1959; Orne, Dinges, & Orne, 1986). Demand characteristics are the totality of cues which convey an experimental hypothesis to a subject, and which thereby constitute significant

determinants of a subject's behavior. These cues include, but are not limited to, "the rumors or campus scuttlebutt about the research, the information conveyed during the original solicitation, the person of the experimenter, and the setting of the laboratory, as well as all explicit and implicit communications during the experiment proper" (Orne, 1962b, p. 779).

Experimental subjects are generally both good at picking up on demand characteristics, as well as motivated to respond in ways that support the assumed experimental hypotheses. The simulator paradigm thus accepts that all suggestion-related responding is suspect of compliance or faking unless relevant observable differences can be shown between "reals" and simulators. However, it also builds on the assumption that if all "reals" are ultimately just faking, then any person should be capable of successfully picking up on and complying with relevant demands if properly motivated to do so. If it turns out that this is not the case, then there would seem to be reason to assume that at least some "real" responding is in fact real in that it is not simulated.

Operationally, hypnotic "reals" are subjects who are deemed highly (hypnotically) suggestible based on pre-testing. They are instructed before a critical experiment to try to undergo the suggested experiences. In contrast, experimental simulators are considered insusceptible or at most minimally susceptible to (hypnotic) suggestions based on pre-testing. They are instructed before a critical experiment to refrain from experiencing relevant effects, but to nevertheless act as if they were, so as to appear as if they were genuine "reals."

The real–simulator paradigm is thus only a quasi-control experimental design because the compared groups differ both in terms of subject population ("reals" pass suggestions but simulators don't during pre-testing) and pre-experimental treatment ("reals" are instructed to experience effects which simulators are instructed to fake). Why not compare similar subjects

to similar subjects, or provide the same instruction to all subjects? Consider first the issue of differential instruction.

Surely a lack of real–simulator differences would be uninteresting if both groups were asked to merely simulate. On the other hand, if both groups were asked to undergo relevant experiences, then the experiment would at best be a retest of which subjects appear highly responsive to hypnotic suggestions. Perhaps one might argue that low test–retest reliability of respective categorization would cast doubt on the genuineness of any hypnotic responding. Yet the test–retest reliability of relevant individual differences in standardized conditions is relatively well established (Hilgard, 1973; Kihlstrom, 2008). So providing the exact same instructions to low- and high-hypnotizables would be unrevealing with respect to the genuineness of "real" responding.

Now consider the issue of differential subject population. Why not compare simulating and non-simulating low-hypnotizables, or simulating and non-simulating high-hypnotizables? The first alternative is clearly a non-starter. If all subjects were low hypnotizables, then by definition no subject would be more than minimally responsive to hypnotic suggestions in standard conditions. Since no one doubts that the relative non-responsiveness of low-hypnotizables in standard conditions reflects a relative incapacity to experience suggested effects, a comparison of simulating and non-simulating low-hypnotizables would of course be irrelevant as to whether hypnotic "reals" may experience relevant effects.

What about the alternative of comparing simulating and non-simulating high-hypnotizables? One worry regarding the use of high-hypnotizables as simulators is that they may unintendedly slip into hypnosis or become responsive to suggestions for some other reason. Another possible concern is that highly suggestible experimental simulators might have already undergone some relevant experiences during pre-screening. However, it is crucial to

the real–simulator paradigm that the responses of simulators should be informed by whatever demand characteristics they can detect, rather than by any actual or prior relevant experience. Thus the reason that, beyond providing "reals" and simulators with different pre-experimental instructions, the respective groups are also drawn from different subject pools.

With that said, it bears emphasis that "reals" and simulators are similar in this respect: neither group is trained in hypnosis or given any instruction as to how to behave in particular. For the question is not whether simulators might successfully execute certain behaviors if trained or explicitly asked to do so. There is good independent reason to assume that they could. The question is whether simulators might figure out on their own how a "real" would respond in certain situations, and then act accordingly to remain undetected by an expert hypnotist.

Hence, simulators are merely instructed to try to fool the hypnotist into thinking that they are genuine "reals." To ensure proper motivation, they are also told that although the hypnotist is experimentally blind to who is a "real" and who is not, he or she knows the experimental setup and will try to identify simulators accordingly. If a simulator is caught, then it is game over and the experiment is terminated for that subject. This experimental design suggests that the more simulators and the more situations in which they are caught, the stronger the evidence that "real" suggestion effects are objectively distinguishable. Which is to say, the greater the number and kind of observable real–simulator differences, the more convincingly the empirical–methodological challenge of a traditional skeptic of suggested hallucination can be met. The following sub-subsections provide a survey of some such observed differences.

4.2.3.2 First-person evidence

Experimental simulators turn out to be very good at feigning hypnotic effects. So most studies do not find strong differences between "reals" and simulators (McConkey, 2008). Nonetheless, they do differ in various important ways. A most evident contrast is in the retrospective reports

of subjects: as opposed to "reals" who typically maintain during post-experimental questioning that they underwent relevant experiences, simulators admit after the critical phase of the experiment that they were only faking (Hilgard, 1986).

Recall however that the simulator paradigm only provides a quasi-control experimental design. For, on the one hand, "reals" are pre-selected from among highly hypnotizable subjects, whereas simulators are pre-screened for low hypnotizability. On the other hand, whereas "reals" are instructed to at least try to experience relevant effects, simulators are asked to avoid undergoing suggested experiences. So there are both pre-existing differences in trait, as well as situational differences in demand between the groups.

A skeptic may posit accordingly that "reals" initially only comply with demands—or, alternatively, that they initially experience some easily achievable effects, but none later when harder suggestions are given—and they subsequently lie about their experiences to sustain their image as responsive subjects (cf. Spanos, 1986; Wagstaff, 2004). In contrast, the skeptic may continue, simulators are initially non-compliant, and so they generally admit to a lack of relevant experiences. Granted, they go along with the simulation game if asked. Yet they are aware from the outset that at least the experimenter who asks them to do so knows that they are low-hypnotizable simulators. So, lest they be seen as genuinely dishonest deceivers, they have good reason to eventually (re)-reveal themselves as experientially unresponsive subjects.

But say simulators are instructed to try to retain their incognito even during post-experimental questioning. In a study of suggested visual hallucination, subjects who passed the suggestion were accordingly asked later what they meant when they said that they "saw" the suggested object. Almost all high-hypnotizable and task-motivated subjects indicated that they had imagined the object but had not believed it was actually present. As opposed to this, simulators maintained that the suggested object had been "really there." They were also more

likely to provide life-like descriptions of the suggested object (e.g., solid, colored, highly vivid) (Spanos, Bridgeman, Stam, Gwynn, & Saad, 1983).

A skeptic may object that, if anything, such results suggest that "reals" are not really undergoing relevant experiences—they are merely engaging in vivid imagery. What about the few "reals" who insist even retrospectively that they actually saw the suggested object? It may be proposed that they were deluded earlier, or they misremember their experiences, or they were in fact just complying.

Fair enough. But notice that we started from the question whether there is any objective difference between "reals" and simulators. We provided some initial evidence that there is. The skeptic then challenged that perhaps "reals" are the ultimate deceivers. By now, it would seem to be granted that at least an overwhelming majority of "reals" are honest, and even the minority who are suspect of faking may have undergone some genuine cognitive distortion (e.g., delusion or memory bias). This is definitely progress.

4.2.3.3 Behavioral evidence

We have been discussing how the skeptical challenge that no objective difference can be shown between "real" and simulated effects of suggestion has been met by the real–simulator paradigm in experimental hypnosis research. In the previous sub-subsection, we reviewed some first-person evidence that indicates that simulators report alleged experiences differently from "reals," whether they are asked to provide honest reports (in which case they admit to not having had appropriate experience) or feigned reports (in which case they insist on having had perceptual-hallucinatory as opposed to imagery-like experience). In the present sub-subsection, let us address some important behavioral findings. We shall start with evidence of overplaying; then move on to performance interference effects; and finally touch on findings of incongruent responding.

4.2.3.3.1 Overplaying

Recall that simulators tend to provide more life-like descriptions of suggested images than "reals." This relates to a general tendency to overplay and overemphasize the role of hypnotized subject (e.g., Hilgard, 1977, 1986; Orne, 1959; Spanos, 1986; Wagstaff, 2004). For example, a relatively large number of "reals" provide spontaneous reports of transparent hallucinations that they can allegedly "see through." In contrast, even after prompting, simulators paradigmatically report complete or "ideal" hallucinations that involve experiencing vivid, solid, sharply outlined images (Orne, 1959; Spanos, 1986; Spanos et al., 1983).

On a credulous interpretation, transparent hallucination is a prime example of "trance logic." The term refers to an ability of a subject "to mix freely his perceptions derived from reality with those that stem from his imagination and are perceived as hallucinations. These perceptions are fused in a manner that ignores everyday logic" (Orne, 1959, p. 295). However, many argue that transparent hallucination is merely a case of incomplete responding (e.g., Spanos, 1986).

Another alleged case of incomplete responding is the partial elimination of pain under (hypnotically) suggested analgesia, which is conceptualized as a negative hallucination. Indeed, even highly hypnotizable subjects still typically report a little (even if significantly diminished) pain when given suggestions for analgesia. Yet this only further differentiates them from simulators, who in similar conditions tend to behave and report as if they experienced no pain whatsoever (Hilgard, 1977, 1986).

How might a skeptic respond? It may be proposed that "reals" intentionally produce incomplete responses, for fear that total compliance would give them away as a fraud. This would suggest that "reals" are most sophisticatedly conning the experimenters. In turn, a skeptic may further propose, simulators intentionally allow themselves to be caught to avoid

being seen as too good at deception. Or perhaps they just exercise less caution, as being found out is less of a concern to them, given that simulation is their prescribed role anyway (cf. Wagstaff, 1981, 2004).

Consider again the case of suggested analgesia, though. Arguably, even after more than three quarters of a century of modern hypnosis research, "There is still no more convincing way to persuade a skeptic that hypnosis is 'real' than by showing that ordinarily painful stimuli can be endured without signs of pain" (White, 1941, p. 479). Appropriate suggestions for analgesia have been demonstrated to significantly reduce both chronic pains such as headaches, low back pain, arthritis, disability- and cancer-related pain, irritable bowel syndrome, non-cardiac chest pain, fibromyalgia, multiple sclerosis, and mixed chronic pain syndromes; as well as acute pains such as related to bone marrow aspiration, burn wound dressing change and debridement, labor pain, gynecologic or women's health procedures, invasive surgical procedures such as angioplasty or arteriograms, chemotherapy, elective plastic surgery, invasive vascular or renal procedures, burn wound care and treatment, and pain associated with healing of a wound or injury (Stoelb, Molton, Jensen, & Patterson, 2009).

A skeptic may grant the above but point out that "Whilst certain suggestions may effectively induce analgesia, it seems equally possible that some subjects may feel obliged to report a diminished sensation of pain irrespective of their actual subjective experience" (Wagstaff, 1977, p. 393). This may well be the case in experimental settings. However, first, insofar as some subjects do experience diminished pain, yet the reported level of reduction differs from that reported by simulators, this only provides further evidence for a crucial difference between "real" and simulated responding.

Second, few if any would deny that, in clinical settings,

> It seems implausible to assume that a person who has hypnosis as the only anesthetic for major surgery, childbirth, extensive dental work, burn debridement, or bone marrow aspiration reports little or no pain merely to meet the demands of the situation, that is, to fool or to please the hypnotist. Yet this same person again asks for hypnosis when it is necessary to undergo another painful treatment. (Orne et al., 1986, p. 477)

Finally, from a dialectic perspective, it is ultimately orthogonal whether some subjects might fake analgesia even in clinical settings. So long as it is granted—as it clearly is—that some subjects in some settings do experience an attenuation of pain as a function of relevant suggestions, traditional skepticism about at least suggested analgesia is already off the table.

4.2.3.3.2 Interference

We have been focusing—and shall continue to focus for a while—on whether there is any objective difference between "real" and simulated responses to suggestions. For starters, some crucial differences were noted in relation to first-person retrospective reports. An important moral of that discussion is that at least most hypnotic "reals" seem to provide honest reports when placed under high demands to do so. We then moved on to third-person evidence relating to overplaying on the part of simulators, as contrasted with frequent incomplete responding on the part of "reals." A crucial moral of the latter brief survey is that at least some cases of suggested analgesia involve genuine changes in pain experience.

Recall, however, that when specifically queried after an experiment what they really meant by that they saw the suggested object, most highly hypnotizable subjects replied that they had only imagined the object and had not believed that the object was there (Spanos et al., 1983). So, yes, a skeptic may grant, most "reals" are indeed honest when specifically required to be so. Yet, it may be argued, it doesn't follow that any subject ever experiences perceptual

hallucination or cognitive delusion or any other effect that would render traditional candidates of suggested hallucination such as visual or auditory hallucinations psychologically real.

Relatedly, a skeptic is likely to emphasize that even if suggested analgesia is experientially real, it doesn't follow that apparent hallucinations such as perceiving a blue image as if red are also real (cf. Wagstaff, 1991). For example, it might be pointed out, suggestions for analgesia may reduce the experience of pain by diverting attention away from the noxious stimuli, or by producing fantasies or counter-pain imagery which are inconsistent with the experience of pain (e.g., fantasizing of painful ice-cold water as soothing lukewarm water on a hot summer beach). However, the argumentation may continue, it is implausible that such factors could fully mediate the effects of standard perceptual hallucinations. Hence, it may be concluded, the reality of suggested analgesia creates little if any pressure for the skeptic to grant the reality of more traditional modes of suggested perceptual hallucination.

We shall later discuss why non-believed-in imagery experience—as opposed to real perceptual hallucination and cognitive delusion—is compatible with the genuineness of suggested hallucination. (The reader already knows the nutshell answer: because perception-related affect may itself be distorted.) In the remainder of this sub-sub-subsection, let us quickly address the argument from attentionally or imaginatively mediated changes in pain experience.

Proponents of a phenomenal-mixing or phenomenal-trumping view of cognitive penetration may object as follows. Insofar as counter-pain imagery may reduce or eliminate pain experience by phenomenally mixing with or trumping perceptual experience of pain, so it is plausible that for example counter-color imagery may reduce or eliminate or change in some other relevant way perceptual experience of color. But let us leave this objection at that, as we shall ultimately argue that, despite its psychological reality, suggested hallucination does not involve relevant changes in the sensory features of perceptual experience.

4. Suggested Hallucination

A dialectically more relevant and empirically more founded response to the skeptical argument from attentionally or imaginatively mediated analgesia is this. True, the mentioned kinds of cognitive strategies are quite effective in controlling short-term acute pain such as associated with various dental interventions (Hilgard, 1986; Miller & Bowers, 1993; Spanos, 1986). However, even highly hypnotizable subjects may find it difficult and mentally fatiguing to sustain attention and/or other cognitive effort for longer periods in case of more protracted pain such as associated with serious burns.

To test this hypothesis, a study exposed subjects to cold-pressor pain meanwhile they performed a cognitively demanding vocabulary task. (The task involved defining test words and choosing their synonym from among five one-word alternates.) As expected, high-hypnotizables were generally better at reducing pain than low-hypnotizables. However, whereas procedures such as attention diversion and counter-pain fantasies interfered with and impaired performance on the cognitive task in both groups, crucially, this was not the case for high-hypnotizables under hypnotic analgesia (Miller & Bowers, 1993).

Though the mentioned study did not instruct low-hypnotizables to simulate, still the results bear relevantly on real–simulator differences. For assume that simulators were included in the study. Assume further that, rather than overplaying their role, the simulators reported similar levels of analgesia as "reals" under similar conditions. Insofar as the simulators did not experience genuine hypnotic analgesia—either because they did not experience analgesia at all, or because they coped with pain by means of attention diversion and/or counter-pain fantasies—the above finding suggests that "reals" would outperform them on cognitively demanding tasks.

That suggested analgesia is at least not fully mediated by standard coping strategies is further supported by a finding that the physiological responses of high-hypnotizables to painful

stimuli differ in a hypnotic-analgesia as opposed to an attention-diversion condition (Wagstaff, 2000). We shall soon discuss (in *Sub-subsection 4.3.2.4*) some further physiological evidence that similarly suggest that "reals" undergo relevant psychological changes. Before we do, though, there is yet one more behavioral example of real–simulator differences to cover: incongruent responding.

4.2.3.3.3 Incongruence

Based both on first-person reports and third-person behavioral evidence, we concluded earlier that at least most hypnotic "reals" are honest under appropriate demands, and at least some cases of suggested analgesia involve genuine changes in pain experience. In the previous section, we then saw that driving a theoretical–explanatory wedge between suggested analgesia and more standard modes of suggested hallucination is more challenging than the skeptic would have hoped. For even disregarding proposals of cognitive penetration of perception via phenomenal mixing or phenomenal trumping by imagery experience, evidence of differential cognitive-performance interference by standard pain coping strategies and suggested analgesia imply that the latter is at least not fully attentionally and/or imaginatively mediated.

Of course, even if suggested analgesia is not a sham, and even if the skeptic has failed to explain away the phenomenon by appeal to mechanisms that could not explain more standard modes of suggested perceptual hallucination, it still doesn't follow that effects of the latter kind also have psychological reality. So a crucial question is: is there any evidence that "reals" and simulators also behave differently when given suggestions for paradigmatic perceptual hallucination? We have already mentioned some differences in first-person reports such as relating to the alleged transparency ("reals") vs. life-like character (simulators) of visual hallucinations. A behavioral difference is accordingly well exemplified by the (in)congruence of responding to auditory suggestions:

> For example, an S [subject] who responded very slowly to hand levitation was given the suggestion that he could hear only my voice. This negative auditory hallucination was tested by rubbing the thumb against the first two fingers of the hand close to the S's ear and asking him "do you hear anything but my voice?" A response congruent with the S's behavior of slow hand levitation and in a medium trance would be either no response or a slight nod or shaking of the head. Frequently, a repetition of the question would be required to elicit even a nod of the head. An immediate and emphatic verbal denial of any auditory perception other than my voice was considered a typical incongruent response—incongruent both in respect to speed and modality of response. (Orne, 1959, p. 295)

Recall that we are not concerned with the state–nonstate debate. So it is irrelevant whether "reals" in scenarios such as the above are correctly construed as being in "a medium trance." The crucial point is that despite their best efforts to remain undetected, experimentally instructed simulators manifest relevant differences in behavior that seem very hard to explain (away) without appeal to underlying differences in experience. Although various further examples could be enumerated to support this point (cf. Cox & Bryant, 2008; Lynn et al., 2008; Orne, 1959), let us move on to review some important (neuro)physiological differences.

4.2.3.4 Physiological evidence

Quick recap: Arguably, the phenomenon of suggested phenomenon is by far the best candidate of cognitive penetration. So why is the effect so widely ignored? A possible reply is that any association with hypnosis still comes with an aura of deep suspicion. Yet despite various related misconceptions both within and without the field of hypnosis research, there is good empirical reason to assume that at least some subjects in at least some cases are not merely

simulating or complying with situational demands. Of particular importance are findings of experiments that employ the real–simulator paradigm. We have surveyed such evidence of both the first-person and the third-person kind. However, so far, all third-person data discussed involved behavioral evidence. The present sub-subsection completes the discussion of the psychological reality of suggested hallucination by reviewing some third-person data from relevant (neuro)physiological studies.

4.2.3.4.1 Autonomic markers

The question of whether hypnotic subjects are lying was directly put to test by a study in which both "reals" and simulators answered questions (among others) about their purported hypnotic experiences while their electrodermal skin conductance response (SCR) was monitored with a polygraph (Kinnunen, Zamansky, & Block, 1994). SCR as a lie detection measure is generally considered the single most reliable and effective somatic indicator in the laboratory. In conformity with other studies, the first of two experiments established a lie detection accuracy of 75% for highly hypnotizable subjects both in and out of hypnosis. The results of the second, crucial experiment indicated that 89% of the (relevant) responses of "reals" met the criterion for truthfulness, whereas only 35% of the (relevant) responses of simulators met this criterion. The difference was statistically significant ($p < .001$).

A skeptic may feel inclined to argue as follows: Since 35% of simulated responses remained undetected among simulators, we should assume at least a similar proportion of undetected lying among "reals." This would suggest that the 11% of "real" responses which did not meet the criterion of truthfulness constituted at best only 11/(11+35) = 24% of all deceptive responses in the group. This corresponds to a minimum rate of 76% undetected lying. So perhaps it is not that implausible that all 89% of "real" responses that passed lie detection were in fact deceitful.

4. Suggested Hallucination

The above reasoning rests on a mistake. The 35% false positive rate among simulators in the second study corresponds to a 65% lie detection rate (true negatives)—which, recall, was 75% for high hypnotizables in the first study. Contrary to the above, this suggests that an even smaller portion of deceptive responses may have gone undetected among "reals" as opposed to simulators (25% vs. 35%).

Yet assume that there was a similar lie detection rate in the two groups. It still doesn't follow that 35% of the responses categorized as truthful among "reals" were false positives. What follows is that those deceitful responses which were categorized as untruthful (true negatives) constitute only 65% of all deceitful responses. Assuming that all 100% – 89% = 11% of "real" responses categorized as untruthful were correctly categorized (no false negatives), it follows that an additional 0.35*(11/0.65) = 5.92% of untruthful responses may have remained undetected (false positives). Subtracting this estimate from the total percentage of positives, a non-skeptic may thus maintain that 89% – 5.92% = 83.08% of "reals'" responses were indeed truthful (true positives).

So we are more or less back to where we started: unless "reals" are so determined and sophisticated at deception that they might intentionally alter their autonomic physiological responses just to fool an experimenter,[13] it would seem that an overwhelming majority of their

[13] Perhaps "reals" indeed possess an exceptional innate ability and/or learned capacity for deception that involves controlling their autonomic responses? A skeptic may note in support that, on the one hand, psychopaths also have an innate talent for lying; and, on the other hand, secret spies themselves acquire the skill to "beat" a lie detector. So perhaps "reals" are just another special group who can achieve this feat.

The proposal is both empirically and theoretically ludicrous. To start with, contrary to myth, what evidence there is indicates that psychopaths respond to polygraph testing in a similar way to other individuals (Grubin, 2010). Further, efforts to find personality differences between hypnotizable and non-hypnotizable subjects—other than those bearing on hypnotic-like behavior itself—have been notoriously unsuccessful (Hilgard, 1973). So there is no evidence to support that "reals" are natural talents at controlling autonomic functions.

As for skill acquisition, spies receive extensive dedicated training and undergo plenty of practice in bringing under voluntary control what are normally involuntary physiological responses. Is it really plausible that a relatively significant portion of the population undergoes similar training and practice only to later systematically fool experimenters of hypnosis?

responses are indeed honest. This conclusion is only further strengthened by evidence from various brain electrophysiological and brain imaging studies. Let us see some representative examples.

4.2.3.4.2 Neural markers

As we have seen, relevant evidence for real–simulator differences include both first-person and third-person data, whereby the latter include both behavioral and physiological findings. In turn, findings of the latter kind include both somatic–physiological and neurophysiological evidence. The previous sub-sub-subsection discussed physiological evidence of the somatic kind. The present sub-sub-subsection concludes the discussion of the psychological reality of suggested hallucination by surveying some neurophysiological findings relevant to the empirical dissolution of traditional skepticism about the phenomenon.

Let us start with one of the first studies of hypnotically suggested hallucination that measured event-related potentials (ERPs). ERPs are statistically averaged electroencephalographic (EEG) recordings of brain activity time locked to specific sensory, cognitive, or motor events. In the study, subjects received a negative olfactory hallucination suggestion for anosmia that they would not smell anything. Afterwards, they were presented with either a strong or a weak odor of eugenol, or no odor just an air puff. In a hypnotic but not in a non-hypnotic condition, "reals" but not simulators showed a significant increase in the amplitude of a positive waveform 300 milliseconds after stimulus presentation (P300) with both but only strong and weak odors (Barabasz & Lonsdale, 1983).

Some studies have also found ERP differences in the opposite direction. In a study of negative visual hallucination, subjects were instructed to visualize a cardboard box that blocked perception of a TV monitor on which stimuli were presented. "Reals" but not simulators showed a decrease in the amplitude of P300 throughout the brain, as well as a reduction in a

negative waveform (N200) in the (occipital) region associated with visual processing (Spiegel, Cutcomb, Ren, & Pribram, 1985).

That some studies show an increase whereas others a decrease in ERP measures is hypothesized to relate to whether subjects receive an obliterating or an obstructing hallucination suggestion, respectively (Barabasz & Barabasz, 2008; Jensen, Barabasz, Barabasz, & Warner, 2001). Suggestions of the first kind are purely negative suggestions to obliterate perception of certain stimuli or stimulus features (e.g., to not smell any odors). Especially in cases that require selective negative hallucination within a sensory modality (e.g., to not perceive slanted as opposed to horizontal lines), subjects may thus initially attend to the target stimuli or stimulus features, before trying to divert attention away from them. Further, to the extent that they expect success at a task, subjects may be surprised by any relevant unobliterated perceptual experience. Thus, presumably, the increases in ERPs.

In contrast, suggestions for obstructive hallucination involve suggesting some positive hallucination to block perception of relevant stimuli or stimulus features (e.g., a hallucinated cardboard box that visually obstructs an actual TV monitor). Rather than actively attempting to disattend to the external state of affairs, subjects in such cases are plausibly more focused on generating and sustaining appropriate positive hallucinations or imagery. Accordingly, even if they fail to completely eliminate perception of relevant (features of) stimuli, this may be of little concern or surprise to them so long as they experience their positive hallucinations or imagery as subjectively compelling. Hence, presumably, the attenuation of ERPs.

Now a skeptic may recall that “reals” are asked to try to experience suggested effects, whereas simulators are asked to refrain from undergoing relevant experiences. The differential-demand hypothesis thus allows that “reals” but not simulators are involved in suggestion-related imagery. In line with the above, this encompasses sustaining and elaborating imaginings

consistent with the aims of suggestions on the one hand, and disregarding information that is inconsistent with the aims of suggestions on the other hand (Spanos & Barber, 1974). So perhaps neurophysiological differences between "reals" and simulators of the presented kind merely reflect differences in involvement in relevant imagery?

The answer depends on what is meant exactly by involvement. If the assumption is that anyone could engage in suggestion-related imagery as "reals" do, then the answer is arguably no. It bears emphasis in this regard that there are significant neurophysiological differences between "reals" and non-simulating low-hypnotizables as well. In addition, neuroimaging studies that measure regional cerebral blood flow (rCBF) with techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) also provide evidence for differences between apparent hallucination and mere engagement in imagery. We already discussed one such study of visual hallucination (in which subjects had to add or drain colors) in the beginning of this chapter (Kosslyn et al., 2000). Let us here consider a PET study which focused on auditory hallucination (Szechtman, Woody, Bowers, & Nahmias, 1998).

As opposed to the color-adding/color-draining study of suggested visual hallucination which compared the same subjects in hypnotic-hallucination vs. nonhypnotic-imagination conditions, the auditory-hallucination study compared two different groups across four hypnotic conditions. One group consisted of high hypnotizables who seemed capable of hallucinating ("hallucinators"). The other group also included high hypnotizables, but they did not meet this criterion in prescreening ("nonhallucinators").

In the baseline condition, subjects were instructed to lay quietly, to think of nothing, and to let their minds be blank. In the hearing condition, subjects were instructed to pay attention to a voice message played with a tape recorder. In the imagining condition, subjects were asked to imagine the same voice message repeatedly as vividly as possible. Finally, in the

hallucinating condition, subjects received the same instructions as in the hearing condition, but the voice message was not played (only the sound of activating the tape recorder was made).

The results indicated a region in the right anterior cingulate cortex (Brodmann area 32) that was similarly activated for "hallucinators" (but not "nonhallucinators") by both actual and suggested (but not imagined) speech. This region is implicated in modulating autonomic activity associated with affect. The finding thus meshes with the view that a key feature of suggested hallucination is an "emotional conviction" that the suggested state of affairs is real (Sutcliffe, 1961).

The indicated neural area is also thought to take part in an "anterior attentional system." That this part of the system is only activated for "hallucinators" but not "nonhallucinators" by suggestions for auditory hallucination lends support to the view that "reals" are differentially absorbed in imaginative activity. It should be emphasized relatedly that the only higher-order personality factor ever found to at least mildly correlate with hypnotizability is absorption. The construct was originally introduced thus:

> Absorption is interpreted as a disposition for having episodes of "total" attention that fully engage one's representational (i.e., perceptual, enactive, imaginative, and ideational) resources. This kind of attentional functioning is believed to result in a heightened sense of the reality of the attentional object, imperviousness to distracting events, and an altered sense of reality in general, including an empathically altered sense of self. (Tellegen & Atkinson, 1974, p. 268)

An important upshot is that even if the phenomenon of suggested hallucination does not involve actual quasi-perceptual phenomenology, the kind or level of emotional–attentional engagement that "reals" manifest is definitely unique. This is basically all we need to dispel

the traditional skeptical worry that the phenomenon is a sham. Thus, by all reasonable empirical and theoretical standards, the verdict is that suggested hallucination is a psychologically real phenomenon.

How did we arrive at this verdict? We started from subjective retrospective reports whereby, as opposed to "reals," simulators either admit to not having had any relevant experience, or they insist that their hallucinations had a life-like perceptual character. We then surveyed behavioral evidence that show that simulators tend to overplay their role; some of their performance is differentially interfered by perceived stimuli; and the speed and modality of their responding is often incongruent with that of "reals." Finally, we saw that there is also relevant physiological evidence for real–simulator differences, involving both somatic and neural markers, with data of the latter type coming both from studies of brain electrophysiology and neuroimaging.

Now if the proposed conclusion that suggested hallucination is a psychologically genuine effect is unwarranted or at least incorrect, then so much the worse for theses of cognitive penetrability. Surely there is no current candidate for which there is better or more supporting evidence. However, so far, we have only partially raised the bar by arguing that the phenomenon cannot be dismissed as sham. In the next section, we shall further raise the bar by arguing that the effect is at least in part cognitively mediated, whereby this mediation is not fully carried by attention. The crucial remaining question to tackle will thus be whether the locus of the effect is genuinely perceptual. Defenders of cognitive penetration must establish that it is; we shall argue that it isn't.

4.3 Cognition

Here we are then: Suggested hallucination is an empirically robust phenomenon the reality of which is indubitable insofar as at least some subjects are very unlikely just faking or complying. However, the phenomenon would of course be irrelevant to the cognitive-penetrability debate if the source of the effects were not cognitive. So is it? The present section argues that in some important respects, it certainly seems so, yes.

To start with, suggestions are typically and paradigmatically verbal. Of course, as discussed in the previous chapter, it doesn't automatically follow from a lexical effect that it is cognitively mediated. In the heyday of behaviorism, Clark L. Hull—who many consider the founding father of modern experimental hypnosis research (Horton & Crawford, 2004)—proposed accordingly that the effects of suggestions are mere conditioned responses:

> a true suggestion response is one in which the subject's own symbolic processes, instead of becoming active either in facilitating or resisting the tendency to action naturally arising from the experimenter's words, remain passive so far as the particular act suggested is concerned. This withdrawal of the subject's symbolic activities would naturally leave his muscles relatively susceptible to the symbolic stimulation emanating continuously from the experimenter. (Hull, 1933, as cited in White, 1941, p. 484)

The below subsections encapsulate some major considerations against the conditioned-response view and in favor of a cognitive-mediation view. One concerns the magnitude of response (*Subsection 4.3.1*); another the orthogonality of cognitive abilities and effort (*Subsection 4.3.2*); a still further the content-specificity of cognitive facilitation and inhibition (*Subsection 4.3.3*); and a final the role of interpretation and expectations (*Subsection 4.3.4*).

4.3.1 Response magnitude

Might suggested hallucination be explained by conditioned responses? Sidestepping issues such as whether perceptual and motor responses afford similar explanation, and whether even motor effects can be satisfactorily explained by conditioned muscle responses, one worry with Hull's proposal is that it doesn't seem to account for "the disjunction between the magnitude of the [hypnotic] response and the procedure which instigates the response" (Sarbin, 1950, pp. 255, 256). For example, say exposure to the word "red" gives rise to some conditioned response involving mental imagery of red. How does this explain extended quasi-perceptual experience as of red in the face of ongoing blue stimulation?

A defender of the conditioned-response view might grant that single exposure to some word is insufficient to give rise to relevant hallucinatory experiences. Yet it might be proposed that repeated exposure within a relatively short time frame may facilitate response sufficiently to enable the desired effect. In support, the defender may point out that suggestions typically involve ample repetition of words related to the suggested state of affairs. A standard suggestion for taste hallucination serves as good illustration:

> Now I want you to think of something *sweet* in your mouth. Imagine that you have something *sweet* tasting in your mouth, like a little sugar… and as you think about this *sweet* taste you can actually begin to experience a *sweet* taste…. It may at first be faint, but it will grow and grow… and grow…. Now you begin to notice a *sweet* taste in your mouth… the *sweet* taste is increasing… *sweeter* and *sweeter*.... How much of a *sweet* taste is there now in your mouth? (SHSS:C, Item 4: Taste Hallucination; Weitzenhoffer & Hilgard, 1962; emphases added)

Still, a proponent of a cognitive view may insist, "no one can believe that mere words entering the ears of a drowsy and relaxed subject can be sufficient cause for the [radical] changes which actually take place" (White, 1941, p. 479). What further reasons may be appealed to in addition to the disjunct-magnitude problem? One is that the conditioned-response view makes the wrong prediction about which subject population is susceptible to relevant suggestions. Another is that insofar as cognitive facilitation and inhibition of conditioned responses is required, such cognitive mediation must be content-specific. Finally, at least some effects seem to be mediated by non-conditioned, idiosyncratic interpretations and expectations. The below subsections take on these issues in the presented order.

4.3.2 Cognitive abilities and effort

The objection from cognitive abilities and effort starts from the observation that, on the face of it, the conditioned-response hypothesis predicts that children as well as adults with underdeveloped cognitive capacities or low impulse control should be more susceptible to suggestions than the average population. However, there is no such evidence. Recall relatedly that efforts to find personality differences between hypnotizable and non-hypnotizable subjects—beyond those that relate to hypnotic-like behavior itself—have also been notoriously unsuccessful (Hilgard, 1973).

Perhaps it may be countered in defense of the conditioned-response view that it doesn't follow from a withdrawal of relevant symbolic processes that achieving and sustaining such passivity requires little to no cognitive effort. On the contrary, it may be proposed, to the extent that many cognitive processes (including those related to perception) are typically themselves automatic (even if not mandatory), such inhibition may be especially cognitively demanding. This meshes with the view that high responsiveness to suggestions involves at least initially a high focus of attention (Barabasz & Barabasz, 2008; Crawford, 2001; Gruzelier, 1998).

A possible issue with the high-cognitive-demand reply is that it may be overkill insofar as it connotes lower suggestibility for example among children. But let us not dwell on the empirical issue of predicted prevalence. A more general, dialectic point is that meeting the objection from weak cognitive capacities or control by appeal to increased cognitive inhibition of automatic processes in effect concedes an active role for cognition. How is this helpful for a non-cognitive account of hypnosis and suggested hallucination?

Perhaps cognitive mediation of relevant effects is compatible with a conditioned-response view to the extent that the purported processes of active inhibition are content-general. Accordingly, a proponent may argue that even if rendering passive otherwise by default active processes itself demands active cognitive intervention, the proposed intervention is akin to switching off a main engine. It requires effort, yes, but without regard for any particularities.

Recall, however, that even according to Hull, a subject's symbolic processes remain passive only "so far as the particular act suggested is concerned." So even he would seem to disagree with the content-general-inhibition view. Still, it is an essential component of his position that, given appropriate cognitive withdrawal, suggestions exert their effect via conditioned responses without the mediation of any semantic or cognitive processing. Could this be right? Here's why it doesn't seem so.

4.3.3 Facilitation and inhibition

In the preceding chapter, we discussed regarding the priming account of cognitive penetration that priming is insensitive to the compositional semantics of syntactically structured (linguistic) representations in general, and to logical operators such as the negative operator (corresponding to "not") in particular. Even if this did not hold for priming, it would still hold for conditioning. But, then, how could conditioned responses account for negative suggestions such as exemplified by this standard suggestion for anosmia?

4. Suggested Hallucination

> In a moment you are *not* going to be able to smell any odors… Even now you are becoming *less* and *less* able to smell odors… you can smell odors *less* and *less*… *less* and *less*… Very soon you will be *un*able to smell even the strongest of odors… Now you can *no* longer smell anything at all. You can *no* longer smell any odors. I am going to place a bottle of an odorous substance under your nose so that you can see for yourself that your sense is completely gone, and you *can't* smell anything… Your nose is completely *in*sensitive… See for yourself that your nose is anesthetized, *in*capable of smelling odors… (SHSS:C, Item 9: Anosmia to Ammonia; Weitzenhoffer & Hilgard, 1962; emphases added)

Notice the continuous repetition of the words "smell" (8 times) and "odors" (6 times, excluding "odorous"). Consider further that similar repeated exposure to the words "sweet" and "sugar" is supposed to automatically give rise to conditioned experience of sweet taste. Why wouldn't "smell" and "odors" give rise to corresponding olfactory experience in the context of negative suggestion then? An evident possible answer is that such responses are selectively inhibited by cognition. However, this response is not available on any view according to which a subject's symbolic (cognitive, semantic, representational, etc.) processes do not become active "either in facilitating or resisting the tendency to action naturally arising from the experimenter's words" (Hull, 1933, as cited in White, 1941, p. 484).

A made-up example further illustrates the issue of cognitive facilitation and inhibition. Say a subject is given a suggestion to perceive a blue image as if red not by reference to "red" but "the opposite color of blue." The natural tendency of the conditioned response to "blue" should thus be in the direction of blue experience. Yet perhaps the very expression "the opposite color of blue" might itself function as an unstructured unit of association that can be

independently conditioned to a response involving red experience? Even so, since "blue" appears in the expression but "red" does not, it is likely that a subject would still have to inhibit a relatively strong tendency for a blue-associated response, meanwhile facilitating an otherwise relatively weak tendency for a red-associated response.[14]

4.3.4 Interpretation and expectancies

Based on considerations such as that suggestions are typically and paradigmatically verbal; that the magnitude of such effects can be very large; that the conditioned-response hypothesis makes empirically wrong predictions about trait suggestibility; and that facilitation and inhibition of conditioned responses would have to be content-specific; it thus seems implausible that at least all suggestion effects could be sufficiently explained (away) by symbolically or semantically unmediated responses. This final subsection on cognitive mediation further strengthens the case by considering apparently context-sensitive, idiosyncratic responses that do not seem to afford an explanation in terms of conditioning even if cognitive facilitation and inhibition are allowed.

Consider for instance how well hypnotic subjects understand—and understood even when hypnosis was more closely associated with a sleep-like state—that "The command to sleep . . . means nothing more or less than an order to withdraw all interest from the world and to concentrate it upon the person of the hypnotist" (Freud, 1922, as cited in Barber, 1960/2009, p. 365). Does a suggestion to "sleep" require resisting a natural tendency for responses associated with falling asleep? Perhaps. But it also involves, among other things, figuring out which of multiple possible alternative interpretations of the word is most feasible in light of

[14] What would happen if the tendency for a blue-associated response and the tendency for a red-associated response were equally strong? A combination of the phenomenal-mixing view of cognitive penetration and the conditioned-response view of suggested hallucination would suggest that subjects might experience a mixture of blue and red (e.g., purple). However, there is no evidence of such fusion of phenomenal colors under suggested hallucination.

semantic and pragmatic considerations regarding the suggestion as a reflection of the hypnotist's intention.

Finally, that subjects are not fully cognitively withdrawn even under hypnosis is underscored by apparent evidence of expectancy-related responses. For example, following a hypnotic induction, whether a hypnotist's voice (allegedly) sounds closer or farther, or whether sounds are (allegedly) perceived as clearer or more muffled, was found to depend on a subject's preconceptions about the effects of hypnosis (Lynn et al., 2008). It is hard to imagine how such variations in (alleged) experience could be explained by differences in conditioned responses, especially for first-timers.

So is the source of suggested hallucination cognitive? It certainly seems so, yes. Accordingly, there is wide agreement in the field irrespective of theoretical viewpoint that at least some kinds of cognitive–interpretive processes are integral to paradigmatic suggestion-generated responses (Barber, 1960/2009; Dienes & Perner, 2007; Hilgard, 1986; Kihlstrom, 2008; Orne, 1959; Sarbin, 1950; Spanos, 1986; White, 1941).

Of course, as in any other case of hypnosis and suggestion, our conclusion is consistent with an assumption of relative withdrawal of critical thinking (Barber, 1960/2009; Sarbin, 1950; Spanos & Barber, 1974) and higher-order thought (Dienes & Perner, 2007). Consider the related example of "flow" experience, whereby people highly absorbed in some challenging but pleasurable activity such as rock climbing may lose a sense of their ego or self-consciousness. It hardly follows that cognition in this state is shut down or bypassed or in auto-pilot mode. Far from it, as flow is associated with an optimal challenge–skill balance, people undergoing such experience often perform at their peak level (Csikszentmihalyi, 1990/2008).

What if, despite all that's been said, no effect of suggested hallucination is in fact relevantly cognitively mediated? As has been repeatedly emphasized: then so much the worse

for finding a plausible candidate of cognitive penetration. The main thesis of this chapter is that suggested hallucination is a non-case of cognitive penetration even if all criteria except for actual perceptual hallucination are satisfied. If it turns out that the criterion of a genuine cognitive source is itself unsatisfied, then the thesis is only more supported. So we are definitely on safe ground at this point.

4.4 Attention

The phenomenon of suggested hallucination is not a sham. So the effect cannot be explained away as mere demand compliance. Nor is the effect plausibly explained by intra-perceptual or inter-modular or infra-cognitive lexical–perceptual mechanisms. For the source of the effect is at least in part plausibly cognitive. Could the effect of cognition be fully carried by attention? Which is to ask: could suggested hallucination be sufficiently explained away by attentional selection or filtering of stimuli?

If the answer to these questions is yes, then at least some skeptics of cognitive penetration might not be moved. Yet, recall, a current goal is to raise the empirical–theoretical–dialectical bar as high as possible before debunking the candidacy of suggested hallucination on account that it does not involve relevant perceptual distortion. This section argues accordingly that although attentional processes indubitably figure in the mediation of the phenomenon, they neither fully mediate the effect, nor are they in all cases of the kind that tends to be disqualified as a plausible carrier of cognitive penetration. Hence, there is both quantitative and qualitative reason to resist the hypothesis that suggested hallucination is a mere effect of attentional selection or filtering.

4.4.1 Early selection

Recall the idea that high-hypnotizables may be absorbed in suggestion-related mental activity to an extent of "total attention" (Tellegen & Atkinson, 1974). Most contemporary experts agree accordingly that hypnotic subjects sustain a high focus of attention to cues emanating from the hypnotist, meanwhile remaining relatively inattentive to other sources of stimuli (Barabasz & Barabasz, 2008; Barber, 1960/2009; Barnier & Nash, 2008; Crawford, 2001; Gruzelier, 1998; Kihlstrom, 2008; Wagstaff, 2014b). This is underscored by a recently revised definition of hypnosis, approved by the Society of Psychological Hypnosis of the American Psychological Association (APA Division 30), which emphasizes that hypnosis is "A state of consciousness involving focused attention and reduced peripheral awareness . . ." (Elkins et al., 2015, p. 6).

That suggested hallucination involves reallocation of attention has face plausibility irrespective of whether good hypnotic subjects are generally better at focusing attention (Gruzelier, 1998; Sarbin, 1950), or whether their attentional capacities may diminish during hypnosis (Crawford, 2001). For say a subject is given a suggestion to perceive a blue image as if red. Conceptually speaking, success at this task requires both a positive hallucination of quasi-perceiving red, and a negative hallucination of failing to see blue. Correspondingly, it is empirically uncontroversial that the subject may try to shift attention away from the perceived blue color of the actual image, to the imagined or hallucinated red color of the suggested image.

In principle, disattending to blue color may thus involve selecting or filtering out blue stimuli or stimulus features. However, on the one hand, as will be discussed in the next section, sensory processing of the actual stimulus state of affairs does not seem to be inhibited by suggested hallucination. But even if it were, note that our hypothetical red-hallucinating subject is not selecting between two different sources or features of external stimuli. When the subject focuses on the suggested redness as opposed to the actual blueness of the respective image,

there is no external stimulus feature that corresponds to red. So at least cases of positive hallucination cannot be sufficiently explained (away) by external attentional selection.

How about cases of negative hallucination? To start with, just as any positive hallucination may involve or require negative hallucination (e.g., hallucination of red may require obliterating perception of blue), so negative hallucination may involve or require positive hallucination. The reader may recall an example mentioned earlier (in *Sub-sub-subsection 4.2.3.4.2*) in relation to an ERP study in which subjects were delivered suggestions to hallucinate a cardboard box to obstruct perception of a TV monitor (Spiegel et al., 1985). Or consider an experiment in which it was suggested to subjects that they would be blind to an otherwise clearly visible number 8 printed in black on a sheet of white paper (Spanos, Flynn, & Gabora, 1989). To succeed in this task, subjects might need to phenomenally superimpose white color over the black numeral by means of positive hallucination.

What if some negative hallucinations do occur without corresponding positive hallucinations? Still such cases would seem to be more immune to arguments from external attentional selection than more common candidates of cognitive penetration. Recall that a common charge in relation to the latter is that external selection effects are insufficiently sensitive to cognitive contents. Yet researchers of hypnosis have for long marveled at how easy it seems for some individuals to selectively ablate from perceptual experience for example "one person in a crowded room, or the perception of face cards in a spread-out pack, or any other quite arbitrary fragment of the visual field" (White, 1941, p. 486). Such selectivity of negative hallucination requires adequate identification of the target stimuli (e.g., that a certain object is a face card). Yet such processes are paradigmatically post-(early-)perceptual or late attentional. So even pure cases of negative hallucination (if such exist) are implausibly explained by early selection of the debated kind.

4.4.2 Late selection

The upshot so far is that neither positive, nor negative suggested hallucination is likely fully mediated by early attentional selection or filtering out of external stimuli or stimulus features. Might suggested hallucination reduce to late attentional selection? This is also unlikely. Recall again the idea that a subject might fail to perceive the actual blue color of a presented image as a result of attending to the hallucinated red color of the image. This points to a possible mediating role of late attention in the creation and sustenance of negative hallucinations. However, notice that the account already presumes the co-occurrence of positive hallucination. So it cannot explain how the positive hallucination came about to start with.

It may be proposed in reply that when given a suggestion to perceive a blue image as if red, subjects initially focus their attention not on positive hallucination but positive imagery of red. Sustained focus on this imagery may then facilitate processes associated with perception of red, meanwhile inhibiting late perceptual processing of blue. Thus, if a sufficiently high focus is sustained for a sufficient time, perhaps this tradeoff between imagery processing and perceptual processing enables perception of blue to give way to hallucination of red.

On the latter proposal, then, attention mediates not between positive and negative hallucination (in virtue of focusing on positive hallucination), but between mental imagery and both positive and negative hallucination (in virtue of focusing on imagery). Still, it is unlikely that internal attention to imagery together with late selecting out of external stimuli can give rise to hallucination without any further contribution. These factors may be necessary, but they do not seem sufficient. For example, highly hypnotizable subjects can be absorbed in a story to the extent of total attention to relevant imagery without experiencing hallucinations. This is underscored by observed differences in involuntary (neuro)physiological indices of attention during mere absorption as opposed to hypnotically suggested hallucination (Gruzelier, 2005).

4.4.3 (Re)interpretation

Finally, recall an earlier discussed study (from *Sub-sub-subsection 4.2.3.3.2*) that exposed subjects to noxious stimuli meanwhile performing a cognitively demanding verbal task. Whereas pain reduction by means of attentional diversion interfered with performance on the cognitive task, this was not the case for highly hypnotizable subjects under hypnotic analgesia (Miller & Bowers, 1993). This meshes with consistent evidence that hypnotic analgesia does not block sensory processing of pain. We shall elaborate on this in the coming sections. For now, the moral is that rather than disattending to noxious stimuli, at least some subjects in at least some cases (re)interpret the stimuli in a positive way (Barber, 1972/2017; Spanos, 1991; Wagstaff, 2000). A most influential researcher of hypnosis described his own experiences thus:

> The experimenter next suggests repeatedly that my left hand is dull, numb, a piece of rubber, a lump of matter without feelings or sensations. I think with the suggestions and I picture the hand as a rubbery lump of matter that is separated from the rest of my body. The experimenter then places the hand in a pain-producing apparatus that brings a heavy weight to bear upon a finger. Although this heavy weight normally produces an aching pain in the finger, I do not think of the stimulation as pain. Instead I continue to think of the hand and finger as a rubbery lump of matter "out there" and I think of the sensations produced by the heavy weight *as sensations* that have their own unique and interesting properties. Specifically, I think of the sensations as a series of separate sensations—as a sensation of pressure, a cutting sensation, a numbness, a feeling of heat, a pulsating sensation. Although under other circumstances I would label these sensations as pain, I do not let myself think of the sensations in this way; instead, I think of them as a complex of varying sensations in a dull,

> rubbery hand and I state honestly that although I experience a variety of unique sensations I do not experience anxiety, distress, or pain. (Barber, 1972/2017, A Personal Report on Responding to Test Suggestions section, para. 5; emphasis in original)

Hence, while no one denies that attention has a crucial role in mediating the effects of hypnosis and suggestions, it is hardly the sole factor or even the most important factor. As the latter quote underscores, at least in some paradigmatic cases, cognitive (re)interpretation of sensory stimuli may figure just as heavily. Could there be stronger motivation for the candidacy of suggested hallucination as a prima facie most plausible case of cognitive penetration?

Then again, similarly to earlier points about the psychological reality and the cognitive source of the phenomenon, not much ultimately hangs on the above. On the one hand, if suggested hallucination really involves quasi-perceptual experience that is arbitrarily sensitive to source cognitive contents, then the phenomenon is plausibly a case of cognitive penetration no matter what. Whether or how or at what stage or level of processing attention may mediate the effect is then a secondary matter of detail. On the other hand, if there is good independent reason to doubt that suggested hallucination is a case of cognitive penetration, then issues concerning the attentional mediation of the effect are moot anyway.

The remainder of this chapter argues in line with the second option that, despite its prima facie plausibility, suggested hallucination is indeed a non-case of cognitive penetration. Having already established that cognition does but attention does not play a relevant role in bringing about the effect, the position to be defended is that the phenomenon does not involve relevant changes in perception. The next section (*Section 4.5*) makes this case with respect to perceptual processing. The section after that (*Section 4.6*) takes head-on the issue of perceptual experience.

Is suggested hallucination then best construed as a delusion? The final sections argue in the negative in respect of a cognitive interpretation (*Section 4.7*), but in the affirmative in respect of an affective interpretation of the delusion view (*Section 4.8*). Suggested hallucination may thus serve as a prime example of a dissociation of perception-related affect from both perceptual experience and perceptual belief (*Section 4.9*).

4.5. Perception: Processing

To recap once again: Suggested hallucination is a prime candidate of cognitive penetration of perception. The phenomenon is not a sham. It seems to be at least in part cognitively mediated, whereby this mediation does not reduce to attentional selection. Apparently and allegedly, the effect involves specific and synchronic changes in perceptual experience. In principle, such penetration of perceptual experience may occur in conjunction with cognitive penetration of (early) perceptual processing of the actual stimulus state of affairs, or despite non-penetration or even impenetrability of the latter.

The present section discusses evidence that perceptual processing is not relevantly affected by hallucination suggestions. First, we shall survey some crucial (neuro)physiological findings (*Subsection 4.5.1*); we will then turn to some key behavioral evidence (*Subsection 4.5.2*); after which we shall acquaint ourselves with some revealing first-person data (*Subsection 4.5.3*). This will set up the discussion for the next section, which argues that suggested hallucination does not involve relevant changes in perceptual experience, either. So the phenomenon is not a case of cognitive penetration of perception.

4.5.1 Physiological evidence

As discussed in the second chapter (in *Subsection 2.2.3*), although it is a logical possibility that (early) perceptual processing may be cognitively penetrated while perceptual experience is not,

this is empirically highly unlikely and no one actually holds such a view. Accordingly, if there were good (neuro)physiological evidence for cognitive penetrability of relevant processes, that would provide strong abductive support for cognitive penetrability of perceptual experience. However, there is no such evidence: if anything, the data suggest that at least early perceptual processing is not cognitively penetrated. Let us start with some data on ongoing pain processing under suggested analgesia. After that, we shall also see evidence of retained sensory processing under more standard modes of suggested hallucination such as relating to audition and vision.

4.5.1.1 Pain

Recall from the previous section that hypnotically suggested analgesia does not impede sensory processing of noxious stimuli. This is supported by consistent evidence for unhindered activity in the primary somatosensory cortex (SI), which is a foremost candidate for processing the sensory–discriminatory aspects of pain (location, intensity, duration, quality) (Miltner & Weiss, 2007; Rainville, Duncan, Price, Carrier, & Bushnell, 1997; Vaitl et al., 2005). Another apparent area is the secondary somatosensory cortex (SII). Yet for example a neuroimaging study found that hypnotic suggestion for decreased unpleasantness related to the immersion of a hand in painfully hot water decreased activation only in a frontal-lobe region called the anterior cingulate cortex (ACC), but in neither somatosensory cortex (Rainville et al., 1997).

Nor does hypnotic analgesia eliminate autonomic markers of pain such as accelerated heart rate and galvanic skin response (GSR)[15] (Crawford, 2001; Hilgard, 1977, 1986; Hilgard & Hilgard, 1994; Sutcliffe, 1961). Or, more precisely, it does not reduce these responses to a

[15] Galvanic skin response (GSR), skin conductance response (SCR), electrodermal response (EDR), electrodermal activity (EDA), psychogalvanic reflex (PGR), and sympathetic skin response (SSR) are alternate terms of the same phenomenon, whereby the electrical resistance or conductance of the skin changes momentarily as a function of psychophysiological arousal. The reader may recall that this was also the somatic marker used by Kinnunen et al. (1994) in the lie-detection study discussed in relation to real–simulator differences (cf. *Sub-sub-subsection 4.2.3.4.1*.).

greater extent than may be attributed to a lack of facilitative feedback from expressive channels of affect that can be brought under voluntary control. Let me explain.

Pain is often accompanied by automatic facial or bodily expressions of negative affect such as grimacing. However, such expressions can also be voluntarily elicited or inhibited on demand. Allegedly, eliciting such expressions increases via facilitative feedback both objective and subjective indices of pain. In turn, the inhibition of these expressions leads to a relative decrease in relevant indices. So consider hypnotic suggestions for analgesia: in response, even simulators may for example inhibit grimacing by voluntarily relaxing their facial muscles. Their physiological markers may thus correspondingly reflect a relative attenuation of pain.

This is not to imply that suggested analgesia is fully mediated by relaxation. Quite the contrary: subjects typically report significantly greater reduction of pain under hypnotic analgesia than under mere relaxation or hypnosis without relevant suggestions. It is this additional component of alleged pain relief that is not reflected by standard physiological measures (Hilgard, 1973, 1986; Hilgard & Hilgard, 1994).

What about apparent cases of positive hallucination of pain? Just as suggestions for decreased unpleasantness of pain are associated with decreased activation only in the anterior cingulate cortex (ACC) but not the primary (SI) or secondary somatosensory cortex (SII), so a suggested increase in unpleasantness of painful stimulation seems to co-occur with increased activation only in the ACC but neither somatosensory area (Rainville et al., 1997). This conforms with an earlier electrophysiological finding that whereas exposure to actual electric shock elicits a GSR response even under suggested anesthesia, no shock-appropriate GSR is produced under mere suggestion for paresthesia (Sutcliffe, 1961).

4.5.1.2 Audition

The upshot of the preceding subsection is that independently of any suggestion, actual noxious stimuli do, whereas merely suggested pain stimuli do not, lead to corresponding early sensory processing. The findings are similar for more standard modes of perceptual hallucination. For example, it was observed early on that hypnotically suggested deafness does not abolish physiological reactions to auditory stimuli (Brown & Vogel, 1938, as cited in Sutcliffe, 1960). Later studies confirmed this insofar as they also found no change in standard somatic markers such as GSR (Sutcliffe, 1960, 1961) and evoked cortical potentials (Hilgard, 1986).

Nor do suggestions for positive auditory hallucination seem to make a relevant impact. For example, a study using an electroencephalogram (EEG) investigated changes in alpha rhythm. This is the pattern or frequency of electrocortical activity observed under normal, awake, relaxed conditions. In concordance with other studies, alpha rhythm disappeared with actual noise whether or not a subject appeared hypnotically deaf, and it remained unchanged in a quiet condition whether or not a subject appeared to hallucinate noise (Lundholm & Lowenbach, 1942–1943, as cited in Sutcliffe, 1960).

Consider also an earlier mentioned neuroimaging study (in *Sub-sub-subsection 4.2.3.4.2*) which observed "hallucinator" and "nonhallucinator" high hypnotizables during baseline, hearing, auditory imagination, and hypnotically suggested auditory hallucination. The reader may recall that only the first group showed similar activation in the right anterior cingulate cortex (ACC) in the hearing and hallucination (but not the imagination) conditions. This ACC area overlaps with that affected by analgesia and pain suggestions. The authors accordingly hypothesize that their results may reflect differential affect and/or attention between apparently hallucinating and self-reportedly non-hallucinating subjects in the respective conditions (Szechtman et al., 1998).

To emphasize: contrary to common misconception, the above study found a relevant difference not in sensory–perceptual but affective–attentional processing. Indeed, ironically, it was not the "hallucinators" but the "nonhallucinators" who showed similar activation in the right auditory association cortex across the hearing and hallucination conditions. This suggests that either "hallucinators" were not literally hallucinating in the latter condition, or activity in the auditory association cortex is not only insufficient but also unnecessary for auditory experience. Even if future neuroscience were to corroborate the latter disjunct (unnecessity), it is the former disjunct (no hallucination) that is more in line with the current state of the science.

4.5.1.3 Vision

Finally, let us consider evidence on suggested visual hallucination. A study investigated the effects of hypnotic suggestions on the evoked electrocortical potentials that follow a brief flash of light. Under normal conditions, the evoked potentials change with the intensity of the stimulus, and they disappear in the absence of sensation. This pattern was not altered by hypnotic suggestions. Although subjects reported seeing the light as brighter or dimmer in accordance with suggestions, there was no corresponding increase or decrease in their cortical potentials (Beck, 1963; Beck, Dustman, & Beier, 1964; both as cited in Barber, 1965).

On the face of it, such results may seem at conflict with the findings of the color-adding/color-draining neuroimaging study of color hallucination mentioned at the very start of this chapter (*Section 4.1*). In that study, as opposed to non-hypnotic instructions to imagine a grayscale image as if colored and a colored image as if grayscale, corresponding hypnotic suggestions for positive and negative color hallucination were associated with allegedly relevant changes in left fusiform activity, which is classically associated with color processing (Kosslyn et al., 2000). Yet careful examination of the data does not support a credulous interpretation of the findings.

4. Suggested Hallucination

Consider for example the lack of difference in left fusiform activity for colored vs. grayscale stimuli in the imagination condition. Everyone agrees that subjects perceived both images relatively accurately in this condition: the colored image in color, and the grayscale image in grayscale. Why was there then no corresponding difference in left fusiform activity? That no such difference was observed casts considerable doubt on the assumption that activity in this area is a reliable neural marker of relevant color experience and hence processing.

A more recent study provided further inconsistent evidence insofar as it reported an increase in left fusiform activity for both positive and negative color hallucination (McGeown et al., 2012). But even in the original study, in which such activity increased for positive color hallucination but decreased for negative color hallucination, the absolute level of activity was in fact similar for color-hallucinated grayscale stimuli and grayscale-hallucinated color stimuli (Kosslyn et al., 2000). Unless color experience was actually similar across these conditions, it is thus untenable that similar/different left fusiform activity reflects similar/different color perception.[16]

Hence, there is little if any physiological evidence that sensory processing of the actual stimulus state of affairs is relevantly affected by suggested hallucination. Accordingly, even

[16] Why not bite the bullet and assume that subjects perceived both "color-added" and "color-drained" images as if grayish colored? This would mesh with a phenomenal-mixing view of cognitive penetration. It is also consistent with subjective ratings of perceived color in the McGeown et al. (2012) study. Still, there is good reason to resist the phenomenal-color mixing interpretation of the data.

For example, consider cases in which it is suggested that a blue image is red and vice versa. No subject ever reports seeing purple or some other hue besides red or blue. Or consider a hypnosis study in which a green number of low brightness was projected onto a screen against a diffuse, very bright, red color. Although the green number was clearly visible when projected on its own, the brightness of the red light rendered the number invisible unless looked at through a green filter. After demonstration of this effect, in some conditions, subjects were instructed to hallucinate a green filter, and then asked about the number projected and the perceived color of the background.

Though all subjects reported appropriate color hallucination, no one could tell what color was projected in green. Further, most subjects did not report anything close to how the red background appeared through the actual green filter, and there was great variance even among their off answers. Hence, even if suggested color hallucinations were perception-like, it would still be doubtful that they interact with perceived colors in accordance with psychological laws of color mixing (Miller, Lundy, & Galbraith, 1970).

the results of some most crucial neuroimaging studies are inconsistent and/or non-suggestive of cognitive penetration of relevant neural markers. Then again, none of this would matter much if there weren't also converging behavioral and subjective evidence that similarly pointed to a lack of relevant perceptual change. Let us see some such evidence.

4.5.2 Behavioral evidence

The studies in this subsection all used implicit behavioral measures. As opposed to explicit subjective measures such as verbal report, which are often used as direct indices of first-person judgments of conscious experience, implicit behavioral measures of task performance are indirect third-person indicators of experience. Importantly, many such measures also reveal unhindered sensory processing under suggested hallucination. We shall review a relevant selection of findings related to perceptual adaptation, motor control, visual priming, persisting illusions, perceptual interference, and paradoxical responding.

4.5.2.1 Perceptual adaptation

A most interesting study investigated perceptual adaptation to displacing prisms under hypnotic anesthesia, an apparent negative hallucination of the loss of tactile and kinesthetic sensitivity (Spanos & Saad, 1984). Displacing prisms are optical devices that displace the retinal image of the visual world. For example, a prism might displace retinal stimulation sideways, so that the apparent location of objects shifts, say, to the right. If subjects are then asked to point at a target object, they initially point too much to the right, in accordance with the displacement of their retinal image.

If subjects are allowed additional trials, though, their performance significantly improves within minutes. Yet when they take off the prism, their performance once again worsens: this time because they point too much to the left. Since both effects—the gradual

improvement with the prism on, and the sudden dip in performance when the prism is taken off—involve a relative leftward shift in pointing, they are considered but two manifestations of the same perceptual adaptation to the right-displacing prism (Harris, 1965).

Now an intriguing feature of such visuomotor adaptation is that the requisite visual recalibration depends on the very pointing—or, more generally, on proprioception of body movement and position. Thus, unless retinal exposure to the displaced visual stimuli is accompanied by internal generation and processing of signals related to the position and movement of the body, visuomotor adaptation does not occur.

So back to the hypnosis study. Exploiting the mentioned facts, subjects were asked to point at a visual target for two minutes with their hypnotically anesthetized dominant arm meanwhile wearing a displacing prism. The prism was then taken off, and the anesthesia suggestion and hypnosis were canceled. During post-testing, subjects in one group were asked to move their hand slowly; a second group were told that hypnotic anesthesia would enable them to overcome displacement aftereffects; and a third, control group received no special instructions. Irrespective of instruction and their earlier apparent limb-anesthesia, subjects in all groups showed a similar and significant displacement aftereffect (Spanos & Saad, 1984).

The results of the study underscore, first, that "hypnotic anesthesia is not the same thing as a nerve block . . . . There can be little question that in hypnotic anesthesia the proprioceptive information is registered in the appropriate cortical centers . . ." (Kihlstrom, Barnhardt, & Tataryn, 1992, p. 40). Second, crucially, "this registration has collateral effects on experience, thought, and action, just as the functionally blind patient sees double when his eye is deformed by external pressure" (ibid.).

4.5.2.2 Motor control

Speaking of functional blindness: it was observed in a clinical setting that a patient who claimed to have no visual awareness of his immediate environment was still reasonably able to maneuver through an examination room without bumping into the furniture (Moene & Roelofs, 2008). This resembles the behavior of sleepwalkers, who also bypass objects. As it happens, so do hypnotically blind subjects. Ironically, this further distinguishes "reals" from simulators: for it is the latter and not "reals" who often bump into objects when moving about under suggestions for blindness (Orne, 1962a).

4.5.2.3 Perceptual priming

A behavioral case study of hypnotic blindness also supports the above observations. Subjects were presented a line drawing of either a happy or a sad face. The task was to guess whether the face was definitely happy, probably happy, probably sad, or definitely sad. One highly hypnotizable subject was given instructions intended to induce high motivation for blindness; another high-hypnotizable was given instructions intended to induce weak motivation for blindness; and a low-hypnotizable subject was given instructions for simulation.

As expected, the guessing pattern of the high-hypnotizables differed from that of the simulator. Yet the proportion of their correct guesses still significantly differed from chance level. Indeed, the highly hypnotizable subject in the weak-motivation condition correctly identified the depicted emotion in all 400 trials, irrespective of reported confidence in the guesses (definitely vs. probably). The performance of the high-hypnotizable in the high-motivation condition was less extreme but more diverse: identification was above chance for low-confidence guesses, but below chance for high-confidence guesses. Thus, for both subjects, despite their insistence of not seeing anything, guessing accuracy was strongly influenced by the actual stimulus state of affairs (Sackeim, Nordlie, & Gur, 1979).

4. Suggested Hallucination

A study with a larger sample presented subjects with uncommon spellings of homophones, together with appropriate disambiguating words that biased interpretation (e.g., *window* and *pane*;[17] *wooden* and *stake*;[18] *garage* and *sale*[19]). Some of these word pairs were presented before hypnosis, while others were presented after hypnosis was induced and suggestion for blindness was given. As expected, when the suggestion was canceled, subjects reported recalling nearly none of the homophones presented during hypnotic blindness.

However, subjects were later administered a suggestion for age regression to 14 years of age, and told that they were sitting in a class at school. In addition to age-appropriate questions about arithmetic and geography, they were asked to spell some words. The spelling task included all previously presented (seen and avowedly unseen) homophones, plus some new homophones, randomly interspersed among non-homophones. Sure enough, subjects spelled significantly more of the previously presented than of the new homophones the uncommon way. Thus, even under hypnotically suggested blindness, visual exposure to the mentioned word pairs still effectively primed performance on the subsequent spelling task (Bryant & McConkey, 1989).

4.5.2.4 Persistence of illusions

Of much relevance to the issue of cognitive penetrability is that perceptual illusions also persist under suggested negative hallucination. Consider the Ponzo illusion, in which the upper of two identical horizontal lines appears longer against a background of a pair of converging lines that resemble railroad tracks in linear perspective. The illusion is eliminated if the background is in reality removed. A study inquired whether hypnotically suggested ablation of the converging

[17] The more common spelling of the homophone /pān/ is *pain* (vs. *pane*).
[18] The more common spelling of the homophone /stāk/ is *steak* (vs. *stake*).
[19] The more common spelling of the homophone /sāl/ is *sail* (vs. *sale*).

lines may have a similar effect. The results indicated that the answer is no (Miller, Hennessy, & Leibowitz, 1973).

Interestingly, the mentioned study also suggested that some highly hypnotizable subjects may actually be more susceptible to the Ponzo illusion in general. In any case, suggested negative hallucination did not affect their performance. Might suggested positive hallucination as of a relevant background when in fact there is none give rise to a similar illusion, though? An earlier study that tested for this did not find any such evidence, either (Underwood, 1960).

4.5.2.5 Perceptual interference

That suggested hallucination does not hinder consequential processing of the actual stimulus state of affairs is further illustrated by studies of negative auditory hallucination that used the method of delayed auditory feedback (DAF). This technique involves recording the speech sounds of subjects through a microphone, and then feeding back the sounds (possibly amplified) through an earphone with a short delay of a fraction of a second. In one such study, hypnotically deaf subjects were asked to say aloud each step and the solution of simple addition problems. Compared with their normal average performance, the subjects performed significantly worse with DAF (Sutcliffe, 1961).

In another study, subjects were given an oral reading task. In pretesting, DAF caused stuttering and other speech disturbances such as mispronunciations, increased vocal intensity, and slowed speech rate. One group of subjects then underwent hypnotic induction and received suggestions for deafness; a second group were simply told to try to the best of their ability not to hear and to be deaf; and a third, control group were retested without receiving either hypnotic induction or deafness suggestions. Subjects who reported partial or total deafness in the first

two groups still showed the same disturbances of speech in response to DAF as did controls (Barber, 1965).

The persistence and robustness of such perceptual interference thus further underscore that the objective effects of suggested hallucination significantly differ from those of actual perception or the lack thereof. Indeed, these effects consistently show that, however susceptible to hypnosis and suggestion a subject may be, sensory processing of the actual stimulus state of affairs is not relevantly altered even by apparently genuine hallucination.

4.5.2.6 Paradoxical responding

Finally, consider an anecdotal report of a highly hypnotizable subject's paradoxical responding to the cancelation of suggested posthypnotic deafness:

> I say to a somnambulist who has been hypnotized: "When you wake you will not see me; you will not hear me; you will be deaf and blind." I wake him; I speak to him and whisper in his ear. He does not show any sign of understanding: his face remains inert. If I then say to him decidedly, perhaps once or several times: "You hear again;" his face shows great astonishment; he hears and answers me. It is in vain that I say "You must have heard me all the time, since your pretended deafness vanished when I assured you that you again heard me." (Bernheim, 1890, as cited in Sutcliffe, 1961, p. 78)

Of course, it cannot be ruled out that the above subject or some other subjects in the cited studies were in fact pretending. But recall that, at this stage of the dialectic, we are granting that at least some hypnotic subjects are not merely faking or complying. Indeed, it bears emphasis that in many cases it is the very same effect that allows us to conclude both that a response is genuine, and that it is influenced by the actual stimulus state of affairs.

Recall for example the case of hypnotically suggested blindness, whereby simulators typically do, but hypnotic "reals" typically do not, stumble into furniture when moving about in a room. This differential behavior on the part of "reals" underscores that they are not merely complying with demands. In addition, it convincingly demonstrates that suggested blindness does not hinder perceptual processing of the actual stimulus state of affairs, at least some information regarding which is even accessible for the control of goal-directed behavior (e.g., walking from one end of the room to the other).

It was noted in the beginning of this chapter that traditional skeptics of suggested hallucination are even more suspicious of extravagant cases (complete responses) than less extreme cases (incomplete responses). Hopefully, it is by now clear why. Many defenders of a credulous view of hypnosis accordingly agree that hypnotic "reals" paradigmatically manifest only incomplete responses to suggestions for hallucination. Whether the first-person data to be presented below also reflect incomplete responding, or are a unique feature of some highly suggestible subjects, is debated. In any case, they provide crucial further evidence for the persistence of perceptual processing of the actual stimulus state of affairs under suggested hallucination.

4.5.3 First-person evidence

In the first part of this chapter, we established that suggested hallucination is a prime candidate of cognitive penetration, insofar as it is a psychologically genuine effect that is at least in part cognitively and not fully attentionally mediated. The burden of the second part of the chapter is to argue that the phenomenon is nevertheless unlikely a case of cognitive penetration, insofar as the effect does not involve relevant distortions in perception. Although our primary interest is in perceptual experience, it has abductive weight that a large body of third-person evidence including both physiological and behavioral data consistently show that suggested

hallucination has virtually no effect on early sensory processing of the actual stimulus state of affairs. The remainder of this section presents some first-person evidence that only further support this conclusion. The next section then addresses some weaknesses of credulous attempts to resolve the apparent tension in holding both that subjects experience genuine hallucinations, and that they are still aware of and guided by the actual stimulus state of affairs.

Consider for starters a fascinating true story from some decades ago. A highly hypnotizable blind person volunteered for an undergraduate class demonstration of hypnotic deafness. The subject's blindness was unrelated to hypnosis, but it served to rule out any unintended visual cues that might influence his behavior. In previous sessions, the subject had already proven to be a hypnotic virtuoso: for example, he didn't even react to the sound of pistol shots under hypnotic deafness. It was thus of no surprise in the class demonstration that the subject showed no sign of hearing at first. But, then, a student inquired whether there may be some part of the subject that still knew what was going on. Prompted by the student's question, the teacher–demonstrator–hypnotist addressed the subject as follows:

> As you know, there are parts of our nervous system that carry on activities that occur without awareness, of which the control of the circulation of the blood, or the digestive processes, are the most familiar. However, there may be intellectual processes also of which we are unaware, such as those that find expression in night dreams. Although you are hypnotically deaf, perhaps there is some part of you that is hearing my voice and processing the information. If there is, I should like the index finger of your right hand to rise as a sign that this is the case. (Hilgard, 1986, p. 186; also Hilgard, 1992, p. 75; Hilgard & Hilgard, 1994, p. 167)

The reader will have guessed: the finger rose. Immediately, the subject requested that his hearing be restored, as he wanted an explanation why he felt his finger rise in a way that did not appear a spontaneous twitch. When the deafness suggestion was canceled by a prearranged tactile cue, the subject insisted that he had heard nothing, and that he was unaware why his finger had moved. With his consent, hypnosis was then re-induced, and the following suggestion was given:

> When I place my hand on your arm like this (he demonstrated) I can be in touch with that part of you that listened to me before and made your finger rise—that part that could hear and knew what was going on when you were hypnotically deaf. When I question that part, it will be able to answer me and tell me what it knows about what happened. But this hypnotized part of you, to whom I am now talking, will not know what you are saying—or even that you are talking—until, out of hypnosis, I shall say, "Now you can remember everything." All right, now I am placing my hand on your arm. (Hilgard, 1986, p. 186; slightly different version in Hilgard, 1992, p. 76; Hilgard & Hilgard, 1994, p. 168)

In the conversation that ensued, the subject (or his "hidden observer") recalled everything that had happened earlier under hypnotic deafness, including the request to signal his auditory awareness by raising his finger. When the hypnotist subsequently lifted his hand, as suggested, the subject had no memory that the conversation even took place. However, once his amnesia was canceled and he was brought out of hypnosis, he recalled all that had occurred in both the hypnotic-deafness and the hidden-observer situation.

The hidden-observer phenomenon was later more systematically investigated in experiments of hypnotic deafness (Crawford, MacDonald, & Hilgard, 1979), hypnotic blindness (Hilgard, 1986; Zamansky & Bartis, 1985), hypnotic analgesia (e.g., Hilgard, 1973;

Hilgard, Hilgard, MacDonald, Morgan, & Johnson, 1978; Hilgard, Morgan, & MacDonald, 1975; Laurence & Perry, 1981), and hypnotic anosmia (Zamansky & Bartis, 1985). Though only a portion of even highly hypnotizable subjects seem to respond to appropriate suggestions, those who do were consistently found to be capable of reporting on the actual stimulus state of affairs.

It bears emphasis that whereas some studies required online reporting of hidden-observer experiences (e.g., in addition to verbal reports of "overt" pain, subjects would indicate "covert" pain by key pressing), other studies elicited hidden-observer reports retrospectively, as in the anecdotal classroom demonstration. Studies of the latter kind provide crucial evidence for detection and accurate identification of relevant stimuli even when this is not required.

For example, subjects in an experiment insisted under negative visual hallucination that they only saw a blank piece of paper. After receiving a hidden-observer suggestion, however, they correctly recalled that the paper had depicted the number 8 or the number 4 (Zamansky & Bartis, 1985). This underscores that suggestion-resistant perceptual processing may even enable the encoding, storage, and subsequent conscious retrieval of relevant memory traces of the avowedly unperceived stimuli.

Here's another method of corroborating the existence of such perceptual memories. In an earlier briefly mentioned experiment (cf. *Subsection 4.4.1*), it was similarly suggested to subjects that they would see only a blank sheet of paper, which in fact displayed a clearly visible number 8. Subjects who passed the suggestion were later interviewed by a second experimenter, who (mis)informed them that only simulators report having seen nothing on the paper. The subjects were further told that hypnotic reals typically initially see the figure, which gradually fades over a span of one minute. The subjects were then asked to draw what they had initially seen on the paper. Despite their earlier protestation that they had seen nothing, all but

one subject correctly drew the number 8 (Spanos et al., 1989; cf. also Spanos, Burgess, Cross, & MacLeod, 1992).

In conformity with the persistence of various physiological and behavioral indices, subjective data of the above kind thus underscore that suggested hallucination does not hinder perceptual processing of the actual stimulus state of affairs. The latter evidence also suggests that at least some phenomena may be better explained by cognitive effects resembling amnesia. Still, perhaps other effects involve genuine cognitive penetration of perceptual experience?

We have already noted (in *Sub-subsection 4.5.1.3*) why the phenomenal-mixing view is implausible. Recall that on that model, even if perceptual processing is cognitively impenetrable, perceptual experience may still be penetrated in virtue of an experiential fusion of suggested-hallucinatory content with perceived content. For example, if a subject receives a suggestion that a blue image is red, then the resulting color experience may be as of a mixture of blue and red (e.g., purple). Yet there is no trace in the literature of such experiences.[20]

The following section discusses three further models on which perceptual experience is relevantly altered despite a lack of relevant alteration in perceptual processing. As we shall see, none of these models is very plausible, and so it is correspondingly unlikely that suggested hallucination involves relevant perceptual change. Later sections will explain how and why this is compatible with non-skepticism about suggested hallucination, and also provide positive

[20] On the face of it, the phenomenal-mixing model of cognitive penetration is consistent with the hypothesis that the essence of hypnosis is "trance logic," whereby a highly hypnotizable subject can "mix freely his perceptions derived from reality with those that stem from his imagination and are perceived as hallucinations" (Orne, 1959, p. 295). However, recall that a prime example of how "These perceptions are fused in a manner that ignores everyday logic" (ibid.) is the phenomenon of transparent hallucination. So say a subject receives a suggestion to perceive a blue image as if red. The hypothesis of trance logic predicts that if the resulting hallucination of red is transparent, then the subject will "see through" the hallucination the actual blue color of the image without any distortion in blue color experience. So the experiential fusion of percepts and imagery that the hypothesis envisages is not of the kind that the phenomenal-mixing model of cognitive penetration presumes.

considerations that it is distortions in affect rather than distortions in perception or perceptual belief that lie at the heart of the phenomenon.

4.6 Perception: Experience

A hypnotized subject honestly reports perceiving a blue image as if red. On the credulous view, this reflects genuine quasi-perceptual experience as of red. Yet an overwhelming body of data suggests continued perception of blue. Bayne (2007, 2010) and the present author (Bitter, 2010) have similarly discussed—but with different conclusions—three possible credulous resolutions of this apparent inconsistency. On the implicit-perception or zombie model, hallucination of red renders perception of blue phenomenally unconscious even if not functionally inert. On the dual-stream model, hallucinated red and perceived blue are simultaneously experienced in parallel streams of consciousness. And on the switch model, hallucination of red and perception of blue are alternately conscious.

This section argues that none of these models is ultimately plausible. A crucial worry regarding the implicit-perception model is that it presumes a level of sophistication of the perceptual unconscious that is both theoretically and empirically unwarranted. A key problem with the parallel-streams model is that it does not explain the seeming efficacy of suggested analgesia. Finally, despite its prima facie merit, the switch model is still in tension with the assumed efficacy of suggested analgesia, as well as with various retrospective first-person accounts.

The section closes by meeting some final possible objections based on some most recently cited evidence by philosophers in favor of a cognitive-penetration and hence credulous interpretation of suggested hallucination.

4.6.1 Unconscious perception

The mainstream view in contemporary psychology is that suggested hallucination involves a dissociation between conscious or explicit, and unconscious or implicit perception (e.g., Bryant & McConkey, 1989; Cox & Bryant, 2008; Hilgard, 1973, 1977, 1986; Hilgard & Hilgard, 1994; Kihlstrom, 1987, 2008, 2014; Kihlstrom et al., 1992; Mallard & Bryant, 2001). Of course, that various perceptual states and processes are not conscious is in itself uncontroversial. No one assumes for example that all interlevels of early perceptual processing are conscious (cf., e.g., Fodor, 1983). The interesting thesis regarding suggested hallucination is that percepts that would normally be phenomenally conscious may themselves remain or become non-conscious.

In support of this view, it is pointed out that similar cases of unconscious or implicit perception are already well established. Implicit perception can be construed as an "effect on the subject's experience, thought, or action of an object in the current stimulus environment in the absence of, or independent of, conscious perception of that event" (Kihlstrom, 2004, p. 94). A standard example is the neuropsychological phenomenon of "blindsight." Blindsight patients seem at least partially cortically blind due to lesions in their primary visual cortex (V1). Still, if given forced-choice tasks that involve guessing certain properties of visual stimuli (e.g., whether there are any lines or motion) presented in their apparently blind visual field, they perform above chance (Weiskrantz, Warrington, Sanders, & Marshall, 1974).

Another common example of implicit perception is subliminal perception. In what was perhaps the earliest systematic study of the phenomenon, subjects were presented 10 cards one at a time in two series. Half of the cards showed a letter (each a different one), while the other half showed a numeral (also each a different one). The subjects were informed that this was the case, but they were placed at a distance that made the characters apparently undecipherable. The task was to name for each card some character that might be depicted. Although subjects

often complained that they could not see anything and that they were merely guessing, they performed above chance in identifying the characters, and even their wrong guesses were biased towards the correct character category (letters vs. numerals) (Sidis, 1898).

Finally, various attentional phenomena such as "inattentional blindness" are also considered to involve implicit perception. In the standard paradigm, subjects are briefly presented a cross on a computer display, after which a pattern mask covers the entire visible part of the screen. The task is to quickly judge which arm of the cross was longer. In critical trials, an additional shape or word is also presented together with the cross. Despite its clear visibility and closeness to the cross, about 25% of subjects still fail to recall the stimulus.[21] They are also unsuccessful at selecting it from a set of alternatives. However, say the critical stimulus is the word 'short.' Attentionally blind subjects are then more likely than chance to choose a picture of an electrical short circuit (Mack & Rock, 1998/1999).

The implicit-perception model of suggested hallucination thus relies on two crucial premises. First, it is assumed that implicit measures of perception reflect non-trivial unconscious percepts in paradigm cases such as blindsight, subliminal perception, and inattentional blindness. Second, it is held that the evidence for unhindered perception of the stimulus state of affairs under suggested hallucination relevantly resembles apparent implicit perception in paradigm cases. I doubt that either premise is sufficiently motivated. Without full elaboration of all issues, the below sub-subsections present some relevant considerations.

4.6.1.1 Paradigm candidates of implicit perception

To repeat, the argument for implicit perception under suggested hallucination incorporates two key assumptions. The first is that paradigm candidates of implicit perception show non-trivial

[21] An exception is if the stimulus depicts a subject's name, in which case no subject fails to recall the stimulus.

unconscious perception. The second is that perception under suggested hallucination relevantly resembles paradigm candidates. Let us address the premises in this order. The question of this sub-subsection is then: do standard examples of implicit perception demonstrate sophisticated unconscious perception with relevant functional–behavioral consequences? Despite according proclamations by some prominent theoreticians (e.g., Block, 1995, 2012; Greenwald, 1992; Kihlstrom et al., 1992), the issue of whether such perception exists is still a matter of much controversy (Peters, Kentridge, Phillips, & Block, 2017; Phillips, 2016; Simons, Hannula, Warren, & Day, 2007).

Consider blindsight. Some patients occasionally describe a very primitive form of visual perception, such as an experience of "a very faint flash" or "a definite pinpoint of light" (Weiskrantz, 1980, p. 378). This meshes with neuroimaging findings that indicate remaining islands of intact areas within the primary visual cortex (Fendrich, Wessinger, & Gazzaniga, 1992). Even if such relative intactness of relevant areas unlikely accounts for all cases (Fendrich, Wessinger, & Gazzaniga, 1993; Stoerig, 1993; Weiskrantz, 1993), other cases may be explained for example by scattered sensory input to neurologically spared regions directly bordering the damaged area (Campion, Latto, & Smith, 1983). In any case, that retained visual capacities are not devoid of all phenomenal experience is further supported by evidence that stimuli presented in the alleged blind field elicit seemingly conscious afterimage experiences (Weiskrantz, 2002; Weiskrantz, Cowey, & Hodinott-Hill, 2002).

This is not to suggest that blindsight patients are dishonest or confused. As an important analysis has shown, at least in the case of motion stimuli, the dissociation between subjective reports of perceptual experience and performance measures of detection in forced-choice tasks may be a function of shifts in response bias as understood within the framework of signal-detection theory (Azzopardi & Cowey, 1998). Response bias is independent of sensitivity, that

is, detectability and discriminability of target signals. It reflects the actual probability of providing a yes (no) answer to the question whether a stimulus is (was) present (absent). If the response criterion is liberal, it results in a relatively high hit rate (true positives), but also a relatively high false-alarm rate (false positives). If the criterion is conservative, the false-alarm rate decreases, but so does the hit rate, and so there is an increase in misses (false negatives).

An important proposal, then, is that blindsight patients are merely applying a more conservative response criterion if asked a yes–no question regarding their perceptual awareness, as opposed to when asked to choose from among two forced-choice alternatives regarding the properties of stimuli (Phillips, 2016). This meshes with various alleged cases of subliminal perception. For example, in the study in which subjects had to guess what letter or numeral was printed on cards placed at a relative distance, subjects in general could still see a "black, blurred, dim spot" (Sidis, 1898, p. 171). Whether they might have consciously inferred the identity of the characters then is irrelevant. The issue is not conscious vs. unconscious cognition. The crucial point is that the subjects were plausibly conscious of whatever perceptual features are paradigmatically conscious under the given optic conditions.

It bears emphasis relatedly that, despite multiple resurgences of interest in the phenomenon in the past five–six decades, very few studies of subliminal perception have applied stringent objective measures of consciousness. For example, in line with classic psychophysics, many early studies construed the conscious threshold of perception to be that level at which sensory stimuli are detected in only 50% of cases. If detectability is lowered to 0%, however, there is very little evidence for implicit perception (Simons et al., 2007).

Finally, recall that inattentional blindness does not occur for stimuli that depict subjects' names. Even if Jack cannot recollect or recognize having seen 'Jeck,' he can recollect having seen 'Jack.' This resembles the "cocktail party effect," which was also multiply

demonstrated by dichotic-listening studies. In these studies, subjects pay attention to and verbally shadow an audio stream presented to one ear, meanwhile disattending to another stream presented to the other ear. Although not much seems to be noticed in the unattended stream, many subjects do notice and correctly recall having heard their name (Conway, Cowan, & Bunting, 2001; Moray, 1959).

An influential interpretation of such findings is that all perceptual stimuli are processed even semantically, but only the most relevant or significant stimuli become conscious (Deutsch & Deutsch, 1963). However, in the dichotic-listening paradigm, for example, it has not been ruled out that subjects occasionally switch their attention between the channels (Holender, 1986). That this is possibly the case is underscored for example by a neuroimaging study that found no semantic processing for unattended words even when they were looked at directly (Rees, Russell, Frith, & Driver, 1999).

In the mentioned study, pictures were superimposed on a meaningful word or a mere string of consonants. As expected, subjects did not perform above chance on a subsequent surprise word recognition task when a picture as opposed to a meaningful word was attended beforehand. This conforms with the assumption that subjects can only encode or retrieve from memory stimuli that are or were attended. Still, perhaps there was some online difference in how unattended meaningful words as compared with unattended mere letter strings were processed? The imaging data indicated no such difference when a picture was attended. Which suggests that the implicit semantic effects thought possible under inattentional blindness do not actually occur without attention.

Of course, as the authors also emphasize, it doesn't follow that the subjects in the above study were phenomenally blind to the unattended words and letters. Indeed, considering the duration of stimulus presentation (500 ms), subjects most likely consciously experienced at

least the color and shape features of the ignored orthographic stimuli. So a lack of attention and semantic analysis should not be mistaken for a lack of perceptual phenomenology.

Possible objection: Since the shape features of words were not experimentally controlled, if the subjects had really phenomenally perceived such features, they should have recognized at least some of the unattended words based on these features. However, this does not seem to be the case, since the subjects' performance on the surprise word recognition task was at chance level. Hence, to the extent that there is little reason to assume that unattended words were semantically analyzed, there is correspondingly little reason to assume that any lower-level features of those stimuli were consciously perceived.

The objection lacks sufficient power. For say the unattended pictures were also recognized only at chance level. Should we conclude that subjects were completely blind to all parts or features of pictures while attending to the word or letter string on which the picture was superimposed? A more plausible explanation is that the respective content was not encoded into episodic memory. Or, as proponents of implicit perception may prefer, perhaps the content was encoded into memory but did not become consciously accessible. In either case, inattentional blindness may be better construed as inattentional amnesia, and hence as not a strictly perceptual but more plausibly cognitive phenomenon (Block, 2007; Moore & Egeth, 1997; Wolfe, 1999).

With that said, two points warrant emphasis. First, even if no case of interest involves relevant unconscious percepts, implicit perception may still be a theoretically and empirically valid construct. Recall that the effect is construed as occurring either in the absence of, or independent of, relevant conscious perception. The above considerations are meant to cast doubt only on the alleged absence, and not the possible independence, of conscious perception in candidate cases.

Finally, even if unconscious perception is less sophisticated and less pervasive than many suppose, the present argument neither implies nor assumes that there cannot be more modest cases of implicit perception. It is one thing to hold that relatively degraded unconscious percepts can have some minor effects on experience, thought, or action; and quite another that unconscious perception is relevantly similar to conscious perception in terms of form, content, and/or function. It is the latter position for which there is arguably insufficient evidence.

4.6.1.2 Suggested hallucination as a non-paradigm candidate of implicit perception

To recap: On the implicit-perception model of suggested hallucination, unaltered perceptual processing is compatible with suggestion-congruent changes in perceptual experience because suggestion-incongruent content may become or remain unconscious. A common dialectical first step in arguing for this claim is to provide apparent existence proof for implicit perception in connection with cases such as blindsight, wherein neural areas associated with perceptual processing are impaired; subliminal perception, wherein stimuli are presented at subthreshold intensity or duration; and inattentional blindness, wherein neurologically intact subjects seem to lack awareness even of supraliminally presented, non-masked stimuli (for further examples, cf. Kihlstrom, 2008). A second dialectical step is then to argue that some features of perception under suggested hallucination relevantly resemble such cases of implicit perception.

In the previous sub-subsection, we reflected on why the first assumption lacks sufficient warrant. The present sub-subsection accordingly turns to and questions the second assumption. The main claim is that even if, contrary to earlier argumentation, the first premise is true and hence paradigm candidates of implicit perception indeed involve relatively sophisticated unconscious processing of sensory stimuli (and so even if blindsight patients are phenomenally blind in their allegedly blind visual field, subliminally presented stimuli are non-consciously perceived, and inattentional blindness is a genuine case of functional blindness); perceptual

processing under suggested hallucination still exhibits features—such as the availability for verbal report and the rational control of behavior of allegedly unconscious percepts—that other candidates do not and which are most typically associated with phenomenal consciousness.

Why is it prima facie plausible that blindsight patients have no perceptual experience of stimuli presented in their allegedly blind visual field? For starters, a most crucial area associated with visual processing is lesioned in their brain. Accordingly, the surprise for neuroscientists and neuropsychologists was not that many patients claim not to see in their blind field. What astonished them was the extent of apparently retained visual capacity. This understandably led some to conclude that blindsight involves primarily a deficit of visual consciousness, rather than a serious deficit of visual functionality (e.g., Marcel, 1983, 1998).

Now if blindsight involves a serious deficit of visual functionality, then it is clearly dissimilar to the case of suggested hallucination. Yet so, too, is it on the conscious-deficit view. Note that the latter view is not that blindsight subjects are akin to partial visual zombies. Rather, the idea is that despite relatively spared visual functionality, content relating to the blind visual field is cognitively inaccessible for the purposes of verbal report, reasoning, and the guidance of goal-directed action. A relative lack of visual phenomenology is then inferred on the basis of such lack of cognitive access (Block, 2007) or access-consciousness (Block, 1995).

So no one assumes that blindsight is like a hypothetical case of "superblindsight," whereby subjects are access-conscious of relevant visual content without corresponding phenomenology (Block, 1995). Assume for example that a glass of water is presented in the blind visual field of a blindsight subject. Whatever implicit perceptual effects the glass of water may exert, the subject will not report (or will report not) seeing the glass. Nor will the subject reach for the glass even in case of extreme thirst (Marcel, 1986). Accordingly, the plausibility

of a relevant lack of perceptual phenomenology in such cases piggybacks on the plausibility of a relevant lack of cognitive access to perceptual content.

In contrast, recall that many subjects (or their "hidden observer") can still report on the actual stimulus state affairs both during and after suggested hallucination. Recall further that they also typically evade bumping into objects during suggested blindness. Suggested blindness would thus unlikely hinder a subject from reaching for a glass of water to quench great thirst. To assume that the subject would nevertheless still lack phenomenal consciousness of the glass of water seems to imply that suggested blindness is a case of superblindsight.

It bears emphasis that no other candidate of implicit perception has this implication. Consider as illustration a classic study of subliminal perception. The study showed that repeated exposure to emotionally neutral stimuli positively biases judged attractiveness of the stimuli even without explicit recognition. To ensure that recognition remained at chance level, stimulus presentation time had to be lowered to one thousandth of a second (1 ms), while illumination was lowered by both a neutral density and a red gelatin filter. Now it cannot be ruled out that the stimuli were still phenomenally perceived. However, given such degradation, it is uncontroversial that any perceived content was retrospectively unavailable for verbal report or explicit recognition (Kunst-Wilson & Zajonc, 1980). So whether or not the stimuli were consciously perceived, the issue of superblindsight does not arise (cf. also Zajonc, 1980).

Consider also the case of inattentional blindness. A reason this candidate of implicit perception is more controversial than neuropsychological and subliminal cases is that the alleged lack of perceptual phenomenology cannot be attributed to physical constraints such as brain damage or improper stimulation. In this respect, inattentional blindness qua functional blindness more closely resembles suggested hallucination than non-functional candidates.

Then again, similarly to non-functional cases, stimuli to which subjects appear inattentionally blind also appear relatively underprocessed and fully access-unconscious. Recall for example (from *Sub-subsection 4.6.1.1*) the neuroimaging study that found no memory or semantic processing for visually presented but unattended words (Rees et al., 1999). Inattentional blindness may accordingly explain many traffic accidents whereby drivers fail to assess the relevance of seen but unattended vehicles. This contrasts with suggested blindness whereby subjects tend not to bump into objects. Or consider the hypothetical thirsty subject who is presented a glass of water. Akin to the blindsight patient but not the subject under suggested blindness, an inattentionally blind subject would presumably not reach for the glass of water no matter how thirsty.

The upshot is that even if we grant the reality of unconscious implicit perception, two crucial differences between paradigm candidates and suggested hallucination cannot be sidestepped. First, all plausible candidates seem to involve degraded perceptual processing. Second, all such cases involve a lack of relevant cognitive accessibility. Whether or which of these factors may be a cause or consequence of the other is irrelevant. The crucial moral is that the level and kind of perceptual processing and cognitive accessibility under suggested hallucination is on a very different par than in other alleged cases of implicit perception.

4.6.1.3 Objection from partial cognitive accessibility

We have been arguing that the implicit-perception (or zombie) model of suggested hallucination is under-motivated because, one, even paradigm candidates of implicit perception may involve relevant conscious percepts; and, two, these candidates relevantly differ from suggested hallucination both in respect of their relatively degraded processing, as well as the inaccessibility of their content for explicit report or reasoning or the rational control of behavior.

4. Suggested Hallucination

Now if both claims are true, then the implicit-perception model is surely unwarranted. But even if only the first claim were granted, defending the model would seem hopeless. For assume, in line with the first claim, that paradigm candidates of implicit perception indeed involve relevant conscious percepts. Then, if the second claim were false, there would be no relevant difference between conscious cases and suggestion-incongruent perception. So it would seem conflicting to maintain that the latter is a case of unconscious implicit perception.

It is thus understandable if a defender of the implicit-perception model should wish to deny the first claim. However, considering the multitude of evidence on relatively unhindered perceptual processing and cognitive accessibility under suggested hallucination, a denial of the second claim seems untenable. Yet if it is granted that suggestion-incongruent perception relevantly differs from paradigm examples of implicit perception, then how can a defender of the implicit-perception model still hold that suggestion-incongruent perception is unconscious?

A possible argumentative strategy is this: First, note that cognitive accessibility of perceptual content is not an either–or but a multi-level category or dimensional determinable that affords various degrees. Second, argue that even if suggestion-incongruent perception under suggested hallucination differs from standard cases of implicit perception, its contents is still not globally accessible to all cognitive systems or processes. Third, argue that global accessibility of perceptual content is a necessary condition of phenomenal consciousness of such content, or at least the level or kind of accessibility characteristic of suggestion-incongruent perception is still insufficient to render relevant content phenomenally conscious.

The first premise of the objection from partial cognitive accessibility—that such accessibility is not an all-or-nothing matter—is uncontroversial. The notion was never intended otherwise (Block, 1995). In contrast, the third premise—that (more) global accessibility would be necessary for relevant conscious perceptual experience—is highly contentious. If (more)

global accessibility were a criterion of phenomenal consciousness, then for example no dream state or delirium state would qualify as phenomenally conscious, either (Bayne, 2007). Even if some would happily concur with this implication (e.g., Dennett, 1976; Malcolm, 1956), most would clearly not. The sustainability of the third premise is thus at least seriously questionable.

Yet is the second premise—that suggestion-incongruent percepts are not globally cognitively accessible—warranted to start with? Perhaps it may be reasoned that the very interest in phenomena such as the hidden-observer effect relates to that subjects seem to lack introspective access to such percepts in standard conditions of suggested hallucination. This is also much disputable. For example, many hold that hypnotic subjects do not actually surrender first-order executive control or executive monitoring, even if it is a "classic suggestion effect" (Weitzenhoffer, 1974) that they experience their responses as if they were involuntary (cf. Woody & Sadler, 2008).

Consider also that cognitive accessibility is a dispositional-property concept (Block, 1995, 2007). Accordingly, even if some perceptual content is not actually cognitively accessed, it doesn't follow that that content is cognitively inaccessible. However, it is the latter that would need to be shown for the objection from partial cognitive accessibility to go through.

It is also compatible with global accessibility if initially accessible content become temporarily hidden behind an "amnesic barrier" (cf. Hilgard, 1973, 1986, 1992; Hilgard & Hilgard, 1994). Recall for example (from *Subsection 4.5.3*) the classroom demonstration of hypnotic deafness in which the subject asked to be brought out of hypnosis after his "hidden observer" "covertly" signaled by raising his finger that he was in fact hearing everything. According to the subject's "overt" verbal account at that point, he was unaware why he had raised his finger. When hypnosis was reintroduced and the "hidden observer" was once again called on, he correctly reported all that had occurred earlier. He finished by saying: "This part

of me responded by raising my finger, so it's all clear now" (Hilgard, 1986, p. 187; also Hilgard, 1992, p. 76; Hilgard & Hilgard, 1994, p. 168).

Note the irony of the last statement: the "hidden observer" refers to him- or itself as "this part of me," and claims to have gained clarity on something which he or it just clarified. Such reports can be made much more sense of if we assume that there is a single, non-dissociated, conscious subject whose hidden-observer reports reflect memory access to content which at other times appear hidden behind an amnesic barrier. If such retrievability of prior perceptual experience is only temporary, once the "hidden observer" retreats, the subject will again seem incapable of recalling the information until appropriate suggestion or a prearranged cue is given.

That a key difference between normal hypnotized and hidden-observer states is whether or to what extent prior suggestion-incongruent experiences seem retrievable from episodic memory is underscored by this retrospective first-person account from after an experiment: "I don't think I'm totally unaware that there is a hidden observer there, but when you asked me what I had been doing, it was really hard to remember. I don't know if I could have. I was like a block" (Hilgard & Hilgard, 1994, p. 173). Accordingly, when subjective amnesia is removed, "the way is open for a reintegration of the subject's experiences" (Hilgard, 1986, p. 235). Many subjects then recount that the "covert" experiences reported by their "hidden observer" were in fact consciously "available the entire time" (p. 208).

It bears emphasis (as similarly noted in *Subsection 3.4.3*) that conscious availability of suggestion-incongruent experience does not imply that subjects reflect on or have higher-order thoughts about such content. As a particular subject noted after a hypnotic-analgesia experiment: "It's strange, because I didn't knowingly feel the pain reported, and (in automatic writing) I didn't knowingly know I was writing" (Hilgard, 1986, p. 210). This may partly

explain how or why some "reals" might misremember as if they had really experienced genuine perceptual distortions.

Yet perhaps higher-order thoughts or representations are necessary for conscious experience? Higher-order theories of consciousness seem to imply accordingly that even "hidden observers" may lack phenomenal consciousness of suggestion-incongruent perceptual content (Dienes & Perner, 2007; Spiegel, 1990/1995). This is certainly not the place to address the general plausibility of such theories. Let is suffice that they also face many of the challenges that militate against the global-accessibility criterion of phenomenal consciousness.

For example, at least some such theories seem to imply that, except perhaps for lucid dreams (LaBerge & Rheingold, 1990), even dream states are not phenomenally conscious (LeDoux & Brown, 2017). Furthermore, insofar as hidden-observer reports seem in all relevant aspects like ordinary conscious reports (Bayne, 2007; Block, 1995; Hilgard, 1986, 1992; Hilgard & Hilgard, 1994; Spanos, 1986; Spanos & Hewitt, 1980), such theories imply the psychological reality of temporary superblindsight and hence the metaphysical actuality of corresponding partial zombie states.

Of course, one philosopher's modus ponens is often another's modus tollens. What may be the motivation to accept so extreme theoretical consequences? A possible answer is that both the skeptical alternative, as well as the remaining credulous views (dual-stream model, switch model) of suggested hallucination seem even less appealing. However, recall that the delusion view is also a non-skeptical even if non-credulous account. A still further and most plausible contender, it shall be argued, is an affective account. So there are still various options to consider. Let us see how each of these alternatives fares.

4.6.2 Conscious perception

We have seen that there is little support for the view that a subject who claims to perceive blue as if red under suggested hallucination only unconsciously perceives the blue color. All things considered, it seems more plausible that perception of blue is indeed phenomenally conscious. Might the subject nevertheless also consciously experience a hallucination of red?

It seems phenomenally inconsistent that the same subject might (quasi-)perceptually experience any given spatial location at any given time as both blue and red. Recall that on the color mixing view, though, perceived blue may phenomenally fuse with hallucinated red, thereby resulting in quasi-perceptual experience as of some mixed color such as purple. However, we have already discarded this option on independent grounds. The present question is whether a subject may literally experience both perceived blue and hallucinated red, not a single mix of the two.

We shall first reflect on the dual-stream model, on which perceptual and hallucinatory content are experienced in parallel streams of consciousness. We will then consider the switch model, on which subjects are not co-experiencing perceptual and hallucinatory content, but phenomenally switching back and forth between the two. Although there is some merit to the latter view, both are ultimately unwarranted.

4.6.2.1 Parallel streams

It is at least a logical possibility that consciousness might dissociate into parallel streams. One stream might then include perceptual experience of blue, while another stream hallucination of red. This view is often—although arguably mistakenly (Weinberger, 2000)—attributed to early psychologists such as William James (1890) and Alfred Binet (1890). A related possibility is

that a subject might split into two subjects one of whom consciously perceives blue, meanwhile the other subject has hallucinatory experience as of red (cf. Block, 1995).

As a preliminary remark, it is at least debatable that either version of the parallel-streams model entails cognitive penetrability of perceptual experience. Surely it would be mind-blowing if cognition could elicit such eccentric changes in the structure of consciousness. Yet if all relevant features of perception (e.g., of blue) in the original conscious stream remained intact in one of the successor streams, then cognitive penetration would not have occurred in that stream. What about in the stream that included hallucinatory experience (e.g., of red)? If it were counterfactually true that the latter stream might itself have carried perceptual experience, then cognitive penetration would seem to have occurred. However, unless this stream were (also) numerically identical with the original stream, it is unclear that this condition would hold.

The moral is that even if some hallucinatory experience were phenomenally indistinguishable from perceptual experience; and even if it were appropriately cognitively mediated etc.; it might still not qualify as cognitively penetrated perceptual experience. Accordingly, even if dreams were both cognitively penetrable and phenomenally indistinguishable from conscious percepts, we would still not say that they constituted genuine cases of cognitive penetration. But nothing depends on this conceptual point, as we shall argue that the dual-stream model of suggested hallucination is implausible in any case.

Recall for example that illusions persist under suggested negative hallucination of stimuli the actual perceptual elimination of which would also abolish the illusions. In line with evidence from hidden-observer reports, this suggests that perceptual content under negative hallucination is functionally not dissociated from hallucinated content (cf. Hilgard, 1986). So it is inconsistent to assume that suggested hallucination might involve illusory perceptual

experience of a whole figure on the one hand (the analogue of perceiving blue); and dissociated, hallucinatory, but non-illusory experience as of only a part of that figure on the other hand (the analogue of hallucinating red).

An even more powerful consideration relates to pain experience. Standard interpretations of the dual-stream view seem to imply that rather than eliminating or significantly reducing pain, suggested analgesia merely dissociates pain experience into a separate stream of consciousness. Even if one stream experienced pain relief, if the other didn't, suggested analgesia would hardly merit clinical application. Yet rather than suing their doctors and/or hypnotists for having had to endure painful experiences in either conscious stream, highly suggestible subjects in clinical settings often ask for repeated pain relief through (hypnotically) suggested analgesia (Hilgard, 1986; Hilgard & Hilgard, 1994; Orne et al., 1986).

Some readers may be wondering whether the argument from effective analgesia is ultimately self-defeating for the currently defended view. Recall the subject who reports seeing blue as if red under suggested hallucination. The burden of this chapter is to convince that the subject still phenomenally perceives blue. Does it not follow that despite reporting a reduction or elimination of pain, the subject still experiences similar pain under suggested analgesia?

Depending on what we mean by pain—both yes and no. In short, the claim is that the sensory–discriminatory features of pain relating to its location, intensity, duration, and quality may indeed be phenomenally conscious. But not the affective–motivational component of pain, which is associated with its unpleasantness and suffering. The latter is of course significantly reduced or even eliminated in genuine cases of suggested analgesia.

Indeed, far from constituting a theoretical embarrassment, suggested analgesia serves as a prime example of the kind of dissociation that arguably also underlies more standard modes of suggested perceptual hallucination. So the reality of suggested analgesia is not denied. Nor

is it denied that the phenomenon involves some form of dissociation. What is denied is that the respective dissociation involves the splitting of a single conscious stream into two conscious streams (parallel-streams view), or into a conscious stream and a non-conscious stream (implicit-perception view). A much more plausible assumption is that the dissociation occurs between perceptual–cognitive and affective experience. But before defending this view, let us see how the switching account and the cognitive-delusion account fare.

4.6.2.2 Phenomenal switching

How to explain perception of the actual stimulus state of affairs under suggested hallucination? We have already excluded the skeptical view, on which subjects merely simulate or comply with demands for hallucination; the phenomenal-mixing view, on which perceptual and hallucinatory content fuse analogously to superimposed colors; the implicit-perception view, on which suggestion-incongruent perceptual content is phenomenally but not functionally trumped by hallucinatory content; and the parallel-streams view, on which hallucinatory and perceptual content are experienced in separate streams of consciousness.

How could perceptual and hallucinatory content both be conscious if they neither phenomenally fuse nor dissociate into separate streams? In line with Spanos and Hewitt (1980), Bayne (2007, 2010) proposes that subjects switch their attention back and forth between the actual and the suggested stimulus state of affairs. Although the proposal is discussed in relation to hypnotic analgesia, it might in principle be extended to more standard modes of perceptual hallucination.

On an extended view, then, subjects may perceive a blue image as if red, or experience noxious stimuli as if not at all painful even in a sensory sense. When their "hidden observer" is contacted, they then refocus their attention on the actual (properties of the) stimuli. So of course hidden-observer states do not involve superblindsight or partial zombiehood or parallel

conscious streams. The explanation of correct reporting in such situations is that, by shifting their attention back to the actual stimulus state of affairs, subjects' hallucinatory experiences give way to normal perceptual experiences.

The switch model has some empirical merit, but even Bayne admits that it does not do full justice to the evidence. For example, subjects in multiple experiments were not aware that their "hidden observer" would be eventually contacted. Still the results were similar (Hilgard, 1986, 1992; Hilgard & Hilgard, 1994; Zamansky & Bartis, 1985). The switch model thus predicts that subjects often spontaneously switch between hallucinatory and non-hallucinatory states. However, there is little evidence for this.

True, subjects readily and not infrequently do signal occasional intrusions of to-be-abolished content in experience. Yet for example they most typically do not report sampled or intermittent hidden-observer experiences (Hilgard, 1986; Hilgard & Hilgard, 1994). Recall also (from *Subsection 4.5.3*) that once their amnesia is lifted, many subjects recount that the "covert" content reported by their "hidden observer" was in fact consciously available all along. Or consider the case of persisting illusions under negative hallucination. The switch model wrongly predicts that the respective illusions arise and cease in accordance with attentional switches to and away from the critical stimuli (cf. *Sub-subsection 4.5.2.5*).

How well does the switch model mesh with evidence on suggested analgesia? Also less than optimally. Recall (from *Section 4.4*) that the effect is not fully mediated by attention. In fact, it can occur even for attended stimuli. Further, implicit markers of pain persist under suggested analgesia (*Sub-subsection 4.5.1.1*). This raises the suspicion that the switch model is ultimately just a special version of the implicit-perception account—which posits sensory processing without phenomenal awareness—whereby disattention to the actual stimulus state of affairs is mistakenly considered a necessary condition of suggested hallucination.

4. Suggested Hallucination

Finally, consider the relative accuracy of retrospective reports regarding the temporal dynamics of "covert" pain under suggested analgesia. If such accuracy depended on relatively frequent phenomenal switches from reduced or eliminated to fully experienced pain, and these switches were in fact most often spontaneous, then clinical applications of suggested analgesia should be still less than preferable for even highly suggestible subjects. That this is not the case—indeed, that many subjects repeatedly prefer the technique—further underscores that at least not all cases of attending to noxious stimuli are accompanied by normal pain experience.

So, sure, subjects who attend to noxious stimuli are most plausibly phenomenally conscious of the location, intensity, duration, and/or quality of pain. What is arguably both empirically and theoretically unwarranted is the inference that such attending subjects thereby also most plausibly experience the negative affect normally associated with sensory pain. If the switch account does not allow for such a dissociation between the sensory and affective components of pain, then that only further militates against the model. On the other hand, if the model is compatible with such a dissociation, then it may ultimately collapse into a hybrid of the implicit-perception view and the affective view.

In any case, as we saw, problems are aplenty for all credulous accounts. Yet perhaps there is no better theoretical alternative? We shall soon see that indeed there is. However, despite generally persisting ignorance and/or skepticism regarding suggested hallucination, two papers in defense of cognitive penetrability have recently argued to the contrary (Arstila, 2017; Newen & Vetter, 2017). Since the cited evidence is claimed to be the strongest yet for cognitive penetration, some quick remarks and debunking are in order. We shall then finally address some non-credulous, non-skeptical alternatives: first, cognitive variants of the delusion view, and then possible construals of an affect-based view.

4.6.3 Most recent credulous defenses

4.6.3.1 Grapheme–color synesthesia

A philosopher–psychologist author pair has most recently argued that the results of a study on suggested grapheme–color synesthesia—whereby perception of certain letters or digits consistently triggers quasi-perception of certain colors—provides most convincing evidence for cognitive penetrability of color experience (Newen & Vetter, 2017).[22] The study participants were neurologically normal, non-synesthetic, highly hypnotizable subjects. During hypnosis, they were presented certain digits in particular colors (e.g., 1-red, 2-yellow, 3-green), and given posthypnotic suggestions that whenever they see, think, or imagine the digits, they would always perceive them in the presented colors.

Once out of hypnosis, the subjects were given a detection task under posthypnotic amnesia. The task was to detect the presence of a black digit against a colored background that could be either congruent (e.g., 1 against red) or incongruent (e.g., 1 against blue) with the assigned color of the digit. As expected, akin to real synesthetes (but not controls), subjects in the posthypnotic-suggestion condition (but not in a no-posthypnotic-suggestion condition) detected significantly less digits correctly in the congruent than in the incongruent trials (Cohen Kadosh, Henik, Catena, Walsh, & Fuentes, 2009).

Could the poor performance in the congruent trials in the posthypnotic-suggestion condition be explained by that the black digits were perceived in the assigned colors? For example, could it be that subjects had a greater error rate for detecting 1 against red than blue because they really perceived the digit in red? Newen and Vetter (2017) buy into Cohen Kadosh et al.'s (2009) credulous interpretation without much second thought or apparent knowledge

[22] The first author, Albert Newen, was present at my earlier workshop poster presentation (Bitter, 2015a) and conference talk (Bitter, 2015b) on the topic of cognitive penetrability and hypnotic hallucination.

that a follow-up study by another group that applied more stringent methodology has meanwhile reported a null result (Anderson, Seth, Dienes, & Ward, 2014).

The main concerns of the latter group regarding the original study were that deterioration of performance is easier to simulate than facilitation of performance; that controls were inadequate insofar as non-hypnotic demands significantly differed from hypnotic demands; that it remained unclear whether the observed performance indeed resembled real grapheme–color synesthesia (e.g., the results indicated far worse deterioration in task performance for "hypnotic synesthetes" than for the single actual developmental synesthete tested on a similar version of the task); and that even if respective behavioral similarities are informative of cognitive processes, they do not provide detail of the phenomenological experience of subjects.

Anderson et al. (2014) thus applied a test (embedded figure test) in which real grapheme–color synesthetes consistently show task improvement rather than task impairment; which is less susceptible to demand characteristics; and which allows for greater comparison of results given its extensive prior use in relevant research on synesthesia. Surprise, surprise: despite a similarity in subjective reports between developmental and hypnotic grapheme–color synesthetes, there was no behavioral effect at all in the latter group. The authors accordingly conclude that "The combination of a strong behavioral effect in Cohen Kadosh et al., and none at all in our study, is most simply explained by subjects responding according to how they believe they should, without hypnotically-induced alterations in perceptual abilities" (p. 6).

Of course, as the authors themselves emphasize, "This claim is entirely consistent with subjects in both studies having subjectively compelling experiences" (ibid.). We shall soon discuss various cognitive and affective ways in which experiences may thus be subjectively compelling. The crucial moral for now is that "Theories that postulate that hypnotic

hallucination is perception-like in a way that goes beyond normal imagery . . . are challenged by the current results" (Anderson et al., 2014, p. 8).

Sure enough, the latter results have themselves been challenged most recently by a group that includes the first author of the original study (Terhune, Luke, & Cohen Kadosh, 2017). So the debate on suggested color–grapheme synesthesia may yet continue. However, given the amount of counter-considerations already touched upon in this chapter in relation to credulous models of suggested hallucination, so dubious results by a single group of credulous researchers can hardly reverse the general dialectic.

Yet perhaps there is still more weighty evidence around the corners? A recent and so far only full-paper defense of cognitive penetrability on the basis of suggested hallucination argues accordingly. Here is why this case is no more convincing.

4.6.3.2 Negative color afterimages

It is a welcome trend in current philosophy of mind that experimental psychology and empirical cognitive science are taken ever more seriously (cf. Knobe, 2015). Less understandable is a trend of digging up nearly forgotten findings of methodologically naïve studies from many decades ago, with little if any consideration of why interest in the respective findings may have originally decayed. Arstila's (2017) rediscovery of an infamous hypnosis experiment from some eight decades ago is an apparent example of such blissful ignorance.

The experiment was conducted by Milton S. Erickson and Elizabeth M. Erickson in 1938. The question was whether suggested color hallucination may give rise to negative color afterimages of the sort normal color experience does (e.g., whether hallucinated blue may give rise to a yellow afterimage). To test for this, subjects were repeatedly given a suggestion under hypnosis that they would be shown different colored pieces of paper one at a time. The

experimenter would name the color of the first piece of paper, and the subjects would nod when they saw the named color plainly and clearly. The experimenter would then show a differently colored paper, which would be named by the subjects. The following piece of paper would again be named by the experimenter, and the one after that by the subjects, and so on.

Although all pieces of paper were in fact white, four out of five subjects consistently signaled after one–two minutes that they saw the suggested colors clearly. The same subjects (and only these subjects) also appeared to name the color of the subsequent pieces of paper in accordance with negative color afterimages. For example, they would say "green" after suggested hallucination of red; and, conversely, they would say "red" after suggested hallucination of green.

Could demand compliance explain the results? Both Erickson and Erickson (1938), and Arstila (2017) think not. Why? Because, one, multiple word association tests administered both in and out of hypnosis revealed no associations between relevant color words. Two, when asked after the experiment, subjects seemed unable to define or to name complementary colors. Three, even after a definition was provided by the experimenter, subjects still appeared unable to correctly guess complementary colors (e.g., they would contrast light and dark colors, such as pink and brown).

Hence, considering that it apparently took one–two minutes to hallucinate colors clearly; that all and only those who passed suggestions for color hallucination reported seeing subsequent pieces of paper in color; that the latter reports appeared consistent with negative color afterimages; and that the subjects had no apparent knowledge or associations of complementary colors; Arstila agrees with Erickson and Erickson that the best explanation of the experimental findings is that due to relatively lengthy hallucinations when the experimenter

named a color, the subjects' visual system adapted to the hallucinated colors, thus giving rise to spontaneous experiences of negative color afterimages.

What is wrong with this reasoning? The evidence on which it relies has already been extensively reviewed more than half a century ago by two separate authors (Barber, 1964, 1965; Sutcliffe, 1960). So we shall only review here a selection of considerations.

To start with, only two other modern studies of hypnotically induced negative color afterimages reported positive results. In contrast, five studies found negative results, and a sixth study reported null results (ibid.). E. M. Erickson (1941) and Rosenthal and Mele (1952) attempted to explain away negative findings by appeal to insufficient hypnotic induction and depth of trance. However, it was in fact a positive study that showed that the effect can be obtained without a lengthy or even any hypnotic induction (Barber, 1959a). So either the objection from insufficient induction or trance does not go through, or the positive findings of the latter study are themselves unreliable. In either case, the abductive plausibility of hypnotically induced negative color afterimages would seem to lessen.

It is also noteworthy that in the mentioned positive study, some subjects reported negative color afterimages after merely imagining a disc in certain colors (Barber, 1959a). Arstila (2017) argues independently on the basis of two sources (Oswald, 1967; Phillips, 2013) that imagery and dreams may both elicit negative color afterimages. However, on the one hand, the evidence is highly dubious. For example, the case for dream afterimages relies on two anecdotal reports from more than a century (Alexander, 1904) and two ago (Gruithuisen, 1812, as cited in Oswald, 1967). The case for imagery-generated afterimages itself seems to rely largely on misinterpreted reports and evidence (cf., e.g., Alexander, 1904; Weiskrantz, 1950). Little wonder that no authoritative contemporary book on imagery (e.g., Kosslyn, Thompson, & Ganis, 2006; Pylyshyn, 2003) even makes cursory mention of such afterimages.

4. Suggested Hallucination

On the other hand, even if the evidence for imagery- and dream-related negative color afterimages were sound, in itself, that would hardly help the case for a credulous account of suggested hallucination. For assume that such afterimages do exist. Then their occurrence after suggested color hallucination is clearly non-indicative of prior perceptual hallucination. Recall that no one is denying that "real" hallucinators paradigmatically engage in vivid imagery. What is debated is whether these subjects experience perceptual hallucinations in a strict sense. If it is a real empirical possibility that they experience imagery-generated afterimages, then the credulous argument from negative color afterimages loses much of its wind.

But the real problem with the argument from negative color afterimages is not its commitment to hallucination- as opposed to imagery-generated afterimages. More serious issues are that, one, it assumes that subjects' color-naming responses were in conformity with the occurrence of actual afterimages; and, two, it interprets the null results of the administered word association tests as good indication of a lack of knowledge of complementary colors. Those who have acquaintance with the literature will know that neither premise is warranted.

Regarding the latter assumption (ignorance of complementary colors), not only critics (e.g., Hibler, 1940) but even defenders of a credulous interpretation grant that "The word association test that the Ericksons used to determine knowledge of after-images was . . . unstandardized, and, in general, a not exact method for stating confidently that the *Ss* [subjects] were naive as regards this knowledge" (Rosenthal & Mele, 1952, p. 703). It bears emphasis accordingly that, although all subjects denied relevant knowledge, they were university freshmen selected from among commerce and engineering students. Indeed, it seems that all and only studies that used university or college subjects found positive results (Barber, 1965). Still further, in both the Ericksons' and Rosenthal and Mele's study, subjects had extensive

prior practice with hypnosis. So they may have been relatively skilled at concealing their knowledge from the experimenters (and perhaps even from themselves).

Finally, consider the most crucial premise that suggestion-sensitive subjects consistently named colors in conformity with negative color afterimages. It turns out that this is false. To start with, subjects who are unacquainted with the negative afterimage phenomenon often fail to perceive the afterimages produced by actually seen colors until they have had a number of practice trials. Further, even among those who are acquainted with the phenomenon, afterimage responses tend to vary across trials even for real stimuli (Downey, 1901; cf. also Barber, 1964; Sutcliffe, 1960).

Yet the responses of subjects in positive studies in general and the Ericksons' study in particular showed neither intra- nor inter-individual variation. The latter is also surprising given both that suggestions such as that a particular piece of paper is "bright red" are ambiguous (Sutcliffe, 1960; cf. also Binet & Féré, 1888), as well as that there are individual differences even in afterimages elicited by real stimuli. For example, in a pre-experiment of the study that reported null findings in relation to hypnotically induced negative color afterimages, teenage subjects of below-average IQ from a girls training school variably described their negative afterimage of a red stimulus as light blue (8), blue (7), turquoise (3), green (3), blue-green (2), dark blue (1), and baby blue (1) (Elsea, 1961).

This brings us to a most crucial and likely surprising point. Notice that out of 25 subjects in the just-mentioned pre-experiment, only three reported an afterimage of green after seeing a red stimulus. Instead, most reported a bluish or bluish–greenish hue. Accordingly, in another pre-experiment, all subjects selected some combination of blue and green discs as most closely resembling their afterimage of red (ibid.). This is no mistake. A most recent perceptual study

similarly found that when forced to choose between green and cyan (greenish–blue), 99% of subjects reported a cyanish afterimage after a red stimulus (Manzotti, 2017).

Consider similarly some afterimage responses following a green stimulus. Of the 25 subjects in Elsea's (1961) pre-experiment, only three subjects described their afterimage as red (2) or reddish (1). An overwhelming majority instead reported pink (17), purplish–pink (1), light pink (1), light violet (1), and purplish lavender (1). In the other pre-experiment, an overwhelming majority accordingly chose some combination of red and violet discs (20), and one subject a combination of violet and blue discs (1), as most closely resembling their afterimage of green. Manzotti's (2017) recent perceptual study supports these findings as well. For when forced to choose between red and magenta (light reddish–purple or purplish–pink or mauvish–crimson), 95% of subjects reported a magentish afterimage after a green stimulus.

Hence, contrary to a longstanding mistake among philosophers and psychologists alike, including generations of mistaken textbooks, the complementary color of red on an achromatic surface is not green but cyanish, and the complementary color of green on an achromatic surface is not red but magentish. As Manzotti notes, the red–green mistake "is a remarkable error, because it has biased the way in which stimuli have been devised and outcomes have been interpreted without anyone noticing the discrepancy. Generations of college students have been trained to believe into a wrong notion" (p. 7).

So, yes, the "correct" responses of subjects in positive studies of hypnotically induced negative color afterimages "were consistent with the negative after-image phenomenon as it is described in elementary college textbooks" (Barber, 1964, p. 317). However, as was emphasized already more than half a century ago, "they were not consistent with the negative after-image phenomenon as it is actually observed in real life" (ibid.). Which is to say, whether subjects were merely complying with demands or really underwent genuine hypnotic

experiences, it is plain implausible that they actually perceptually experienced afterimages in the colors reported.

Hence, philosophers would do better to let long-buried psychological empirics rest in peace. No one contests the contribution to the study of hypnosis and suggestions of early researchers like Milton H. Erickson. But their methods did not meet the standards of current scientific practice, and many of their views have been superseded. Even if it cannot be ruled out that the superseding views are themselves mistaken—some surely are—there is generally little hope that buried dead scientific data can decompose with time into valuable philosophical fossil fuel.

So if there is really no good available evidence or argument for a credulous view of suggested hallucination, then rather than further occupy ourselves with exposing fool's gold, it is high time that we inquire in virtue of what suggested hallucination still has psychological reality. The next section addresses traditional cognitive interpretations of the delusion view accordingly. The section after that then discusses the affective interpretation of suggested hallucination. Skeptical and credulous views—rest in peace.

4.7 Delusion

Suggested hallucination is a prime candidate of cognitive penetration of perceptual experience. The phenomenon is not a sham or a mere function of compliance. It is at least in part cognitively mediated, and at least not all such mediation is carried by attention. However, suggested hallucination does not involve relevant changes in at least early perceptual processing. Contrary to the contemporary mainstream view at least in psychology, neither is it likely that it involves relevant changes in perceptual experience. Accordingly, there is little support for—but weighty considerations against—credulous views proffered by the phenomenal-mixing

model, the implicit-perception model, the parallel-streams model, and the switch model. Consideration of some recently rediscovered findings only furthers this conclusion. Hence, unless there is really no better alternative, despite its strong candidacy, the likelihood is small that suggested hallucination is a real case of cognitive penetration of perceptual experience.

So what may more plausibly explain the phenomenon? It is widely thought that the only theoretical alternative of the skeptical and credulous views is a cognitive account on which "suggested hallucinations are only forms of delusions" (Sidis, 1906, p. 257). Although there are some innocuous interpretations of this claim that are in general unproblematic, we shall grant that the delusion view as typically construed does not provide a plausible account of some crucial effects such as suggested analgesia. However, we shall argue, either the delusion view is too restrictively construed, or there is yet a further, non-skeptical, non-credulous, non-cognitive view to consider. In either case, the key explanatory component is affect. So whether the respective account is best construed as a version or an alternative of the delusion view, suggested hallucination is proposedly best explained by an affect-based model.

But first things first. Based on a novel categorization, this section summarizes three cognitive interpretations of the delusion view. On a weak reading, subjects are only deluded about the actual stimulus state of affairs. An issue with this view is that it does not explain why many subjects report perceptual distortions. On a strong reading, subjects are deluded about their current perceptual experience. A problem with this version is that it posits gross introspective error. An intermediate or hybrid view is that subjects are deluded about their past perceptual experience. Although this version plausibly explains some cases, like all cognitive accounts, it is still at odds with phenomena such as suggested analgesia.

4.7.1 Weak version

On a standard construal of the delusion view, a subject under suggested hallucination may falsely believe that a blue image is red, without corresponding perceptual experience as of red. On a weak or innocuous reading of this claim, the subject may still hold true first-person phenomenal beliefs about his or her perceptual experience of blue. Such a scenario would resemble cases of persisting illusions insofar as the subject might entertain the belief that his or her perceptual experience misrepresents the actual stimulus state of affairs.

Of course, beliefs regarding the actual stimulus state of affairs are often true in the case of persisting illusions (e.g., I can entertain the true belief that, despite my illusory perceptual experience, the lines in a Ponzo figure are of equal length), whereas similar beliefs are paradigmatically false on a weak version of the delusion view of suggested hallucination. But even disregarding this difference, the analogy does not seem to hold. For at least in many cases of suggested hallucination, subjects appear to be deluded not about the actual stimulus state of affairs but their very experience of those affairs.

Recall for example (from *Section 4.1*) the color-adding/color-draining neuroimaging study of color hallucination. It was not that subjects were led to believe that the grayscale image was colored and the color image was grayscale. Rather, "subjects were always asked to experience the stimulus in a certain way" (Kosslyn et al., 2000, p. 1281). Many subjects reported according suggestion-relevant experiences. So there is no indication that they held delusional beliefs about the actual stimulus state of affairs. Insofar as they entertained any false beliefs, those were more plausibly false first-person phenomenal beliefs concerning their own perceptual experiences.

4.7.2 Strong and intermediate versions

On a strong interpretation of the delusion view, subjects under suggested hallucination entertain distorted beliefs about the phenomenal character of their current perceptual experience, whether or not they consider their perceptual experience to accurately represent the actual stimulus state of affairs. This allows that a subject may correctly believe that a stimulus is blue; accurately perceive the stimulus as blue; but still mistakenly entertain that he or she is perceiving red.

As opposed to a weak version of the delusion view, which is theoretically uncontroversial but empirically non-fitting with the evidence, a main problem with the strong version is its prima facie theoretical implausibility. Many philosophers and psychologists alike doubt that one may entertain false first-person phenomenal beliefs about his or her current perceptual experiences.

Then again, even if subjects are not—for perhaps they cannot be—deluded about their current perceptual experiences, they may still be deluded about their past experiences, however recent those experiences may be. This suggests a third, intermediate or hybrid interpretation of the delusion view, on which suggested hallucination involves distortions not of online but offline first-person phenomenal beliefs.

Recall for example (from *Subsection 4.5.3* and *Sub-subsection 4.6.1.3*) that a typical difference between normal hypnotic and hidden-observer responding is that suggestion-incongruent content appear hidden behind an amnesic barrier in the former but not the latter condition. Some of the allegedly best evidence for suggested hallucination may accordingly reflect distorted memory of—rather than distorted online judgments about—the perceived character of stimuli.

For example, a relatively recent case study (briefly mentioned in *Section 4.1*) claimed to provide relevant electrophysiological evidence that one of two hypnotic virtuosos in fact experienced color hallucinations (e.g., a blue shape as if red and vice versa). However, the stimuli were only presented for less than one-fortieth of a second (24 ms), and the electrode recordings were pooled from three larger brain regions (occipital, parietal, posterior temporal) some of which are not specifically associated with perception. Further and most importantly, subjects also received suggestions for amnesia, so that they wouldn't remember during task performance the earlier delivered suggestions for color hallucination (Koivisto et al., 2013).

All things considered, it is not unreasonable that amnesia for the hallucination suggestions generalized to amnesia for suggestion-incongruent perceptual experiences. As hypnotic "reals" often do, the respective subject may have accordingly confabulated episodic memory of her experiences, in line with the unrecalled suggestions for color hallucination. From her point of view, these pseudomemories may have been subjectively quite compelling. So perhaps she really had little to no inkling how off her recollections were, despite relatively undistorted perceptual experiences and online perceptual and first-person phenomenal beliefs.

Hence, even if the delusion view is ultimately implausible on both its weak and strong readings, a hybrid account that emphasizes deluded episodic memory may still explain certain findings. Nonetheless, there are admittedly cases that do not seem reducible to distortions in memory. Perhaps the most evident example is suggested analgesia.

Surely it would serve medical practitioners and hypnotists in avoiding being sued by pain-enduring subjects in clinical settings if subjects had no memory of their painful experiences. But recall that we are granting the reality of suggested analgesia to the extent that subjects are at least not suffering from unpleasant feelings. Even if delusional beliefs somehow

figure into this picture, a crucial moral is that there is most plausibly some change in the affective (and not just some cognitive) component of pain as well.

Yet perhaps affective changes also lie at the core of some other cases of suggested hallucination, such as apparent occurrences of online visual or auditory hallucination? This is the proposal to be defended. So rather than try to apply some general cognitive-delusion view of suggested hallucination to specific cases such as suggested analgesia, we shall do the opposite. The next section accordingly starts by providing positive support for an affective account of suggested analgesia, from where it then extends the discussion to a more general, non-cognitive, affective account of suggested hallucination.

4.8 Affect

So here we are: suggested hallucination is insufficiently explained by skeptical, credulous, and cognitive-delusion accounts. Although there is some truth to each view (even "real" responding may be partly influenced by compliance; attention may itself elicit certain changes in perceptual experience; some subjects may entertain genuine pseudomemories; etc.), they do not provide a satisfactory (dis)solution of certain puzzles. For example: how can it be both true that a subject feels pain and yet is not in pain? Or: how can a subject both perceive blue and seemingly hallucinate red? The most plausible account, it is argued, is that subjects indeed still consciously perceive the actual stimulus state of affairs, yet they undergo affective experience that would normally only occur if the suggested perceptual state of affairs did as well.

We shall first outline the dissolution of the apparent tension between persisting pain experience and the efficacy of suggested analgesia (*Subsection 4.8.1*). The crux of the dissolution is that it is essentially the affective and not the sensory component of pain that is targeted by suggested analgesia. We will then elaborate on the idea that it is essentially an

affective and not the cognitive component of perceptual belief that is influenced by suggestions for more traditional modes of perceptual hallucination (*Subsection 4.8.2*). Good analogies are the inappropriate feeling of conviction in clinical delusion, the displacement of emotions in dreams, and the strong feeling of familiarity during déjà vu. Common allegations to the contrary notwithstanding, it is arguably similar dissociations of affect that make best theoretical and empirical sense of the neurophysiological, behavioral, and first-person findings.

4.8.1 Suggested analgesia

How can a subject under suggested analgesia both feel pain and yet not be in pain? The reader is already acquainted with the starting premise of the answer: pain experience is typically complex, involving both a sensory–discriminatory and an affective–motivational component. (A possible third component is cognitive–evaluative.) The sensory–discriminatory component relates to our capacity to detect and discriminate the location, intensity, duration, and quality of pain. This capacity is primarily associated with brain activity in the somatosensory cortices (SI, SII). In contrast, the affective–motivational component relates to the unpleasant, aversive, avoidance-related features of pain. These features are primarily associated with activity in the anterior cingulate cortex (ACC) (Grahek, 2001/2007; Hilgard, 1973, 1986; Hilgard & Hilgard, 1994; Melzack, 2001; Rainville et al., 1997).

The second premise is that the sensory and affective components of pain can and do come apart in certain conditions. For example, in the neurological syndrome of pain asymbolia, patients do not show any negative affective reaction even to severely noxious stimuli. On the contrary, they often smile, chat, and even laugh under such stimulation. All the same, they can still detect and discriminate the sensory features of pain (Grahek, 2001/2007).[23]

[23] Pain asymbolia is not the same as congenital or acquired analgesia, which involve a complete shutdown of the pain system despite relatively intact tactile sensation. Subjects inflicted with the latter conditions frequently suffer

4. Suggested Hallucination

Similar mitigation of chronic pain affect without corresponding diminution of pain sensation has also been observed with lobotomy (neurosurgical interruption of nerve tracts to and from the prefrontal cortex [PFC]), cingulotomy (neurosurgical creation of small incisions in the anterior cingulate cortex [ACC]), various opiates including morphine, and placebos. In these cases, the reduction in discomfort and suffering seems to be mediated by a more general effect whereby the alleviation of anxiety about pain and/or the bringing about of a state of contentment and bemusement minimize the readiness to respond to the sensation of pain (Barber, 1959b).

The final premise of the argument for an affective model of suggested analgesia is that the phenomenon relevantly resembles the mentioned cases. Indeed, there is relatively good evidence that suggested analgesia alters experience of the affective but not the sensory features of pain. Thus, when honest subjects under suggested analgesia report pain relief, in one sense, they are telling the truth: their pain-related stress and suffering is indeed diminished or even eliminated. In another sense, however, "hidden observers" are also correct: they still experience relevant phenomenology related to the location, intensity, duration, and quality of pain (Barber, 1959b; Hilgard, 1986; Hilgard & Hilgard, 1994; Miltner & Weiss, 2007; Rainville et al., 1997).

The above picture is both psychologically and neurologically oversimplified, but the gist of the proposal should be clear. The primary goal of analgesic techniques is to attenuate the unpleasant aspect of pain. Whether or what exact role cognition may play in achieving this goal—and so whether pain affect is itself cognitively penetrable—is ultimately immaterial to the present argument. What matters is that the strictly perceptual or sensory features of pain are still experienced and hence not cognitively penetrated.

severe physical injuries, and it is not uncommon for them to die prematurely, as they are incapable of feeling even sensory pain. This underscores that perception of sensory pain is not reducible to tactile and thermal perception, and that a reduction or elimination of sensory pain may itself occur without tactile or thermal anesthesia.

Of course, as indicated earlier (in *Sub-sub-subsection 4.2.3.3.2*), cognition may still indirectly influence the sensory processing of pain, for example by means of attentional diversion (Miller & Bowers, 1993; Miltner & Weiss, 2007). Or recall (from *Sub-subsection 4.5.1.1*) the alleged possibility of inhibiting facilitative feedback by voluntarily controlling one's bodily reactions to pain such as grimacing (Hilgard, 1973, 1986; Hilgard & Hilgard, 1994). Surely there are further examples. Yet none seems to involve a relevant influence of cognition on pain sensation. That neither does suggested analgesia is beautifully exemplified by a passage from a hypnosis researcher that was also cited (in *Subsection 4.4.3*) in relation to the (re)interpretation of attended noxious stimuli:

> The experimenter next suggests repeatedly that my left hand is dull, numb, a piece of rubber, a lump of matter without feelings or sensations. I think with the suggestions and I picture the hand as a rubbery lump of matter that is separated from the rest of my body. The experimenter then places the hand in a pain-producing apparatus that brings a heavy weight to bear upon a finger. Although this heavy weight normally produces an aching pain in the finger, I do not think of the stimulation as pain. Instead I continue to think of the hand and finger as a rubbery lump of matter "out there" and I think of the sensations produced by the heavy weight *as sensations* that have their own unique and interesting properties. Specifically, I think of the sensations as a series of separate sensations—as a sensation of pressure, a cutting sensation, a numbness, a feeling of heat, a pulsating sensation. Although under other circumstances I would label these sensations as pain, I do not let myself think of the sensations in this way; instead, I think of them as a complex of varying sensations in a dull, rubbery hand and I state honestly that although I experience a variety of unique sensations I do not experience anxiety, distress, or pain. (Barber, 1972/2017, A

> Personal Report on Responding to Test Suggestions section, para. 5; emphasis in original)
>
> This conforms with a general account of hidden-observer findings by another expert:
>
> What the subjects commonly report through automatic [hidden-observer] talking, after they have experienced no pain in the ice water following suggested anesthesia, is that the sensory experience of the cold was actually as intense as ever, but the cold did not hurt: "So cold that it ordinarily would have been very painful, but it didn't bother me at all." (Hilgard, 1973, p. 407)

Hence, there is good reason to believe that suggested analgesia does not relevantly alter sensory phenomenology of pain. If this is right, then the effect does not provide existence proof of cognitive penetrability of perceptual experience. At best, it involves cognitive penetration of affective experience. This is great dialectic progress: for what is widely considered most crucial evidence against skeptical accounts of suggested hallucination is thus at once most crucial evidence for an affective rather than a perceptual (credulous) or cognitive (delusion) account. Let us then see how the account extends to more traditional modes of apparent hallucination.

4.8.2 Perceptual hallucination

In the preceding subsection, we discussed that—similarly to pain asymbolia, frontal lobotomy, cingulotomy, opiates, and placebos—suggested analgesia likely involves a dissociation between the sensory and affective experience of pain, whereby only the latter (affect) but not the former (sensation) is relevantly responsive to suggestion-congruent cognizing. Can this affective model be generalized to more standard modes of suggested hallucination such as visual and auditory hallucination? The present subsection argues in the affirmative.

4. Suggested Hallucination

We shall start by revisiting the idea that a core feature of suggested hallucination is an emotional conviction that the suggested (perceptual) state of affairs is real. Arguably, such affect is doubly dissociable from belief. Thus, under suggested hallucination, a subject may experience a decrease or loss or detachment of the affective aspect of conviction from suggestion-incongruent perceptual belief on the one hand (negative hallucination), and an increase or generation or attachment of this affective component to suggestion-congruent imagery on the other hand (positive hallucination).

Similar displacement of affect in relation to dream or thought content has been a central theme of psychodynamic theories in psychology for over a century. More recently, in philosophy and the cognitive (neuro)sciences, there has been growing interest in dissociations of epistemic or noetic or metacognitive feelings specifically. The second sub-subsection discusses some apparent examples accordingly, with a focus on their implications for explaining suggested hallucination.

The third sub-subsection concludes the argument by replying to some final possible objections. Meeting the first challenge involves a reassessment of what is most widely cited as the strongest neurophysiological evidence for a credulous view. The second challenge is to explain why the posited feeling of conviction is an affect and hence distinct from both (meta)cognitive attitudes and (meta)cognitive phenomenology. The final challenge is to make the abductive case that an affective interpretation best accounts for phenomena such as transparent hallucinations as well.

We shall then wrap up and draw some modest implications.

4.8.2.1 Feeling of conviction

4.8.2.1.1 Non-doxastic delusion

The first defense of a delusion view of suggested hallucination was possibly put forth by Sidis (1906), who argued that "The alleged hypnotic or post-hypnotic hallucination is not at all of the nature of a hallucination, it is a *delusion.* All we do by such a suggestion is to act on the subject's belief" (p. 247; emphasis in original). Belief in what? Sidis initially hints a strong interpretation on which "the subject believes that he perceives [the suggested state of affairs]" (ibid.). However, he later seems to press only a weaker view on which

> The most we can say of hypnotic, post-hypnotic, or suggested hallucinations is that they are saturated, so to say, with the belief in the supposed presence or existence of the object suggested, somewhat in the same way as the child believes in Santa Claus, or as the school boy believes in Washington, or as we believe in the existence of Julius Cæsar. The belief, however, is not of the vital over-bubbling stimulating effect given by a direct perception of an external object, true or hallucinatory, but is one essentially representative in character. (p. 249)

It bears emphasis that even on this account, deluded belief under suggested hallucination lacks a "vital over-bubbling stimulating effect." Sidis emphasizes accordingly that when subjects are asked to report via "automatic writing" (essentially a hidden-observer method) "what you really see and not what you do not see" (p. 255), they disclaim hallucinated content and describe the actual stimulus state of affairs as they normally would.

Perhaps the thus tapped perceptual beliefs are unconscious (implicit), split off into a separate (parallel) conscious stream, or alternate with suggestion-congruent belief (switching)?

Neither seems to be Sidis's view.[24] So subjects hold inconsistent perceptual beliefs? This still doesn't seem to be his view. Rather, it appears that for him, deluded belief is ultimately "no more than a very vivid persistent idea" (p. 249). Which is to say, strictly speaking, the respective state is not doxastic.

The historical–exegetical irony is that, Sutcliffe (1960, 1961)—to whom we owe the modern-era distinction between the skeptical, credulous, and delusion views of hypnosis, and who himself argued for the latter position—probably also had a non-doxastic notion of delusion in mind. Consider his very last conclusion in the second of his two famous papers on the topic:

> The distinguishing feature of hypnosis appears to be the subjective state; and the main feature of this state is the hypnotized subject's emotional conviction that the world is as suggested by the hypnotist, rather than a pseudoperception of the suggested world. (p. 200)

What is an emotional conviction, or the emotion or feeling of conviction, and how does it relate to belief? Here's a brief (and incomplete) walkthrough.

4.8.2.1.2 Historical background

In his essay titled "On the Emotion of Conviction" (1871/1915), Walter Bagehot argued that what is commonly referred to as belief includes both an intellectual and an emotional element. The intellectual element is what is properly called assent; the emotional element is conviction.

[24] Sidis construes of automatic writing as a reflection of the subconscious. However, considering the whole of his argument and how the subconscious was understood by his teacher, William James (1890), it is unlikely that he thought of the subconscious as phenomenally or even access-unconscious. What he more likely meant by claims such as that "The automatic writing, as is usually the case, is not taken cognizance by the patient . . ." (Sidis, 1906, p. 256) is that subjects typically do not reflect on or entertain higher-order thoughts about suggestion-incongruent content (cf. *Sub-subsection 4.1.6.3* and *Subsection 4.3.4*).

Assent is based on "the laws of evidence" (p. 99); whereas conviction is probably "most closely connected with the bodily state" (p. 101).

What is the relation between these two components of belief? According to Bagehot, rigorously trained thinkers typically start with the sifting of facts and weighing of arguments. Comprehension of the results of these arguments then leads to assent, which in turn allows for an emotion of conviction to set in. However, there are cases in which despite giving clear assent to a conclusion, "our minds seem dry and unsatisfied" (p. 101). These cases involve assent without conviction. But even in cases where conviction is present, the intellectual processes leading to belief may be so weighty that a subject hardly notices the emotional component.[25]

However, most importantly, there are supposedly opposite cases as well. Everyday examples come from intellectually untrained persons, who quickly form an opinion on virtually any topic that comes before them, and who often treat their opinion as invincible knowledge even without much thinking. A more extreme example is the anecdote of the Caliph Omar, who allegedly burnt the Alexandrine Library, saying: "All books which contain what is not in the Koran are dangerous; all those which contain what is in the Koran are useless" (p. 99). Bagehot comments that

> Probably no one ever had an intenser belief in anything than Omar had in this. Yet it is impossible to imagine it preceded by an argument. His belief in Mahomet, in the Koran, and in the sufficiency of the Koran, came to him probably in spontaneous rushes of emotion; there may have been little vestiges of argument floating here and there, but they did not justify the strength of the emotion, still less did they create it, and they hardly even excused it. (p. 99)

[25] This is supposedly common among "calm and quiet minds"; "and as these quiet, careful people have written our treatises, we do not find it explained in them how important the emotional part is" (p. 99).

Accordingly, “In cases like the Caliph Omar’s,” the emotion of conviction allegedly “governs all other desires, absorbs the whole nature, and rules the whole life” (p. 101). Further, once it is acutely felt, it is indelible or at least leaves a long-lasting mark on the mind—akin to the feeling of love, which may keep on fueling feelings towards another person even if it turns into hate. In the same way, Bagehot argues,

> experience proves that no one who has had real passionate conviction of a creed, the sort of emotion that burns hot upon the brain, can ever be indifferent to that creed again. He may continue to believe it, and to love it; or he may change to the opposite, vehemently argue against it, and persecute it. But he cannot forget it. Years afterwards, perhaps, when life changes, when external interests cease to excite, when the apathy to surroundings which belongs to the old, begins all at once, to the wonder of later friends, who cannot imagine what is come to him, the grey-headed man returns to the creed of his youth. (p. 101)

Finally, of much relevance to the present discussion, Bagehot mentions two “unorthodox” states of mind in which belief is “evidently” not a purely intellectual matter: dreams and certain forms of insanity involving delusions. Had he known about the phenomenon of suggested hallucination, he may well have mentioned the effect as well.

Now most readers will probably not have heard of Bagehot before (he was a British journalist, businessman, and essayist). Or at least few if any will have recalled that he is the very first reference of William James in his chapter on belief in *The Principles of Psychology* (1890). James starts by distinguishing belief from imagining, by noting that belief but not imagining involves cognizing reality. What is the inner nature of this mental state or function? In particular, what does the sense of reality consist in? James argues that, first and foremost, it

"is a sort of feeling more allied to the emotions than to anything else. Mr. Bagehot distinctly calls it the 'emotion' of conviction" (Vol. 2, p. 283; emphasis omitted).

James argues accordingly that in a "relative" sense of reality, "in which we contrast reality with simple unreality, and in which one thing is said to have more reality than another, and to be more believed, reality means simply relation to our emotional and active life" (p. 295; emphases omitted). Thus, "The quality of arousing emotion, of shaking, moving us or inciting us to action, has as much to do with our belief in an object's reality as the quality of giving pleasure or pain" (p. 307).

On this view as well, then, a person may come to believe in something for no other reason than that he or she conceives that thing with passion. James references French philosopher Charles Bernard Renouvier,[26] who calls such beliefs "mental vertigo." In cases of mental vertigo, James explains,

> the object of passion makes us deaf to all but itself, and we affirm it unhesitatingly. Such objects are the delusions of insanity, which the insane person can at odd moments steady himself against, but which again return to sweep him off his feet. Such are the revelations of mysticism. Such, particularly, are the sudden beliefs which animate mobs of men when frenzied impulse to action is involved. (p. 309)

Would suggested hallucination fall under the category of induced mental vertigo on this view? In the case of "untutored minds," possibly, yes. For, James argues, "Every exciting thought in the natural man carries credence with it. To conceive with passion is *eo ipso* to affirm" (p. 308; emphasis omitted). However, importantly, even James allows that educated

[26] James mistakenly uses the initial "M." when citing Renouvier here. At other places, he uses the initial "Ch."

persons may resist the causal–doxastic consequences of the emotion of conviction, and hence not be deluded into holding unjustified beliefs. He emphasizes accordingly that "The greatest proof that a man is *sui compos* [legally or morally responsible in virtue of having a capacity for reasoning] is his ability to suspend belief in presence of an emotionally exciting idea. To give this power is the highest result of education" (p. 308).

Thus, despite some differences in their view, Bagehot and James ultimately agree that not only may an emotion of conviction accompany or facilitate assent or the cognitive element of belief, it may also occur completely independently. Bertrand Russell would probably second this. In his "Mysticism and Logic" (1914), he notes that "The mystic insight begins with the sense of a mystery unveiled, of a hidden wisdom now suddenly become certain beyond the possibility of a doubt. The sense of certainty and revelation comes earlier than any definite belief" (pp. 784–785). This alleged sense of certainty, which can supposedly accompany either definite or inarticulate belief or experience, is what Russell seems to refer to elsewhere as emotional conviction:

> Belief in a reality quite different from what appears to the senses arises with irresistible force in certain moods, which are the source of most mysticism and of most metaphysics. While such a mood is dominant, the need of logic is not felt, and, accordingly, the more thoroughgoing mystics do not employ logic, but appeal directly to the immediate deliverance of their insight. But such fully developed mysticism is rare in the West. When the intensity of emotional conviction subsides, a man who is in the habit of reasoning will search for logical grounds in favour of the belief which he finds in himself. But since the belief already exists, he will be very hospitable to any ground that suggests itself. (p. 793)

On Russell's view as well, then, it appears, emotional conviction and the cognitive aspects of belief are doubly dissociable. For, on the one hand, a subject may experience an emotional conviction without preceding reasoning or concurrent cognitive assent. On the other hand, the subject may retain a belief even after emotional conviction has subsided. Whether Bagehot and James would agree that one may assent to a belief after emotional conviction has subsided but before logical grounds for the belief have been established, and whether assent in such cases may itself be better construed as affective rather than cognitive, are at least unclear. But we can sidestep such issues, as our main concern is whether emotional conviction can occur without belief or assent or related cognitive factors. On the latter question, Russell and James and Bagehot clearly appear to agree.

Yet perhaps the Russellian understanding of emotional conviction as a sense of certainty differs from the Jamesean notion relating to a sense of reality? We shall touch upon some similarities and possible differences between such epistemic feelings or emotions in the next sub-subsection (*4.8.2.2*). For now, let it suffice that Richard Bevan Braithwaite,[27] who also took great interest in "The Nature of Believing" (1933), himself argued for complete independence between such feelings and belief:

> This is the point to say something about the feelings of conviction which we frequently associate with our beliefs, and which have been made the *differentiæ* of belief by many philosophers. I do not wish to deny that in a great number of cases I have a feeling of conviction when I believe: indeed I think that this feeling of conviction may reasonably be used as evidence for the existence of the belief. But I seem to have a belief frequently with no feeling of conviction: I believe quite thoroughly that the sun will rise tomorrow, but experience no

[27] He is often referenced only as R. B. Braithwaite.

> particular feeling attached to the proposition believed. And it seems possible to have the feeling of conviction without believing a proposition. So I cannot accept the feeling as part of the essence of belief. . . . [I] am convinced that belief and feeling are different, and that the latter is not part of the former although we may reasonably use it as a criterion. (pp. 141–142; emphasis in original)

Notice how the dialectic changed in the roughly half a century covered above. Bagehot's primary burden was to argue not (just) for the possible independence but the very plausibility of an affective element or notion of conviction. James then postulated such affect to lie at the core of the sense of reality that he held most crucial to belief. Breaking with Western intellectualism, Russell in turn emphasized its epistemic potential in facilitating deep insights independently of reasoning. By the time of Braithwaite, it was thus not really debated whether such feelings exist or may serve as generally reliable indicators of belief. The question was, rather, whether such affect is essential to or at least "the original source of conviction" (Jastrow, 1917, p. 523).

The majority view appears to have been that it isn't. Yet, for some reason, the baby of emotional conviction was eventually thrown out with the bathwater of affective essentialism about belief. Accordingly, even more than half a century after Sutcliffe (1960, 1961) defended his delusion view of hypnosis, it still seems to be missed that, rather than pressing a purely cognitive interpretation, he may have laid the seeds for an affective account.

Of course, it is ultimately irrelevant what Sutcliffe thought. He is merely a springboard for ideas that arguably make most theoretical and empirical sense of some key candidates of cognitive penetration such as suggested hallucination. So let us see what modern psychology and cognitive (neuro)science have to say on the affective component of conviction, before

moving on to the more general topic of epistemic feelings, after which we shall (re)assess the proposed theory in light of some most crucial evidence on suggested visual and auditory hallucination.

4.8.2.1.3 Contemporary background

A most important milestone in the modern empirical investigation of emotional conviction is the research conducted, supervised, and summarized by psychologist Robert P. Abelson in his 1987 APA (American Psychological Association) Award Address on "Conviction" (1988). Drawing on attitude research in social psychology, Abelson and his students showed via a series of questionnaire-based surveys that conviction is multidimensional; that most subjects only hold partial convictions; that there is great individual variation along each dimension; and that the affective and cognitive dimensions are separate.

They identified three dimensions in particular: emotional commitment, ego preoccupation, and cognitive elaboration. The first factor purportedly reflects items of subjective certitude, which "are primarily noncognitive and have the phenomenological flavor that seems to capture the raw stuff of emotional conviction" (p. 272). Behavioral intentions also join this factor. Accordingly, on Abelson's construal,

> emotional commitment is very often the product of a powerful and inexplicable experience, in which a sudden rush of conviction occurs when a congenial idea arises in an exciting setting. The prototype is the individual standing in a meadow looking at the stars, who, all of a sudden and in a flash, knows what to believe or what to do. (p. 274)

4. Suggested Hallucination

Never mind the factor of ego preoccupation, as it correlates relatively highly with emotional commitment anyway. However, not the factor of cognitive elaboration, which would remain separate even on a two-factor model. The latter cluster supposedly

> invokes the extent of the mental network in which the views at hand are embedded. It includes the ability to name other issues that might come up in a conversation about the issue, the ability to name consequences that might ensue if the issue was acted upon (e.g., if abortion was made illegal), and the items, "I've held my views a long time compared to most people," "I have more knowledge on the issue than the average person," and "It's easy to explain my views." (p. 273)

Recall that most subjects do not score high on all factors. Abelson emphasizes accordingly that "It is possible," among other combinations, "to have an emotional commitment without much cognitive elaboration" (ibid.). A powerful example is that the correlation between emotional commitment and cognitive elaboration in relation to questions regarding the issue of AIDS—which was a relatively new and pressing issue at the time in the United States—was only 0.18. In statistical terms, this implies that the variance in scores along the emotional factor explained only $0.18^2*100 = 3.24\%$ of the variance in scores along the cognitive factor. Even in the softest of subdisciplinary areas of psychology, this is way too low to have real explanatory value.

Following Abelson et al.'s and various converging results from both within (e.g., Zajonc, 1980) and without the field (e.g., Damasio, 1994), psychology and cognitive (neuro)science snapped out of their historical amnesia and ignorance regarding the affective component of conviction relatively quickly. Accordingly, the view soon became fashionable

again that "Conviction concerns, vitally, the role of feeling in rational thought" (Brown, 1996, p. 163). On this view,

> A logical demonstration of a mathematical proposition can give the truth of the proposition but not a conviction of its truth. The conviction is in the feeling that accompanies the demonstration. Mathematicians have commented that a proof *feels* right before it is logically demonstrated. (ibid.; emphasis in original)

This brings us to the present. Most relevantly, in contemporary cognitive neuropsychology and cognitive neuropsychiatry and their philosophy, it is increasingly thought that "strong (and misplaced) feelings of conviction are the phenomenological hallmark of delusions" (Bruno, Sachs, Demily, Franck, & Pacherie, 2012, p. 3). For example, it was proposed not long ago that due to a failure of an unconscious checking system that "tags" thoughts that require extra conscious checking, delusional and confabulating subjects lack any doubt, which manifests in an inappropriate feeling of conviction (Turner & Coltheart, 2010). This conforms with typical first-person descriptions, according to which deluded beliefs feel intuitively right, self-evident, or obviously true (Bruno et al., 2012).

Hence, it is not necessarily the case even for clinically deluded subjects that they cognitively assent to their delusions. Rather, since they experience irresistibly strong feelings of conviction, they can't rid themselves of the affective impression that things are as they entertain them to be. Does this sound familiar? Freud suggested that we often experience similar phenomena for example in our dreams. Or consider déjà vu experiences, which involve a feeling of familiarity that we have already experienced or foredreamt the present situation. The list of examples is long. Let us review some of the most illustrative among them.

4.8.2.2 Related dissociated feelings

On the proposed construal of an affective-delusion account of suggested hallucination, a subject may be cognitively and affectively so invested in suggestion-congruent imaginative activity that respective imagery become associated with a feeling of conviction that the imagery content is perceptually or externally real. Such feelings need not be accompanied by any corresponding belief. Nor do they depend on the vividness or life-like character of imagery. So even relatively dull and non-perception-like imagery may become associated with a feeling of conviction.

Correspondingly, even vivid perceptual experience may lack or lose or dissociate from a sense of reality, that is, from an affective conviction that one is perceptually related to external reality. This is plausibly why many subjects can engage in hidden-observer reporting without switching out of their pseudo-hallucinatory (or pseudo-delusional) state. This may also explain why some subjects appear to genuinely misremember as if they had experienced real perceptual hallucinations. For even if suggestion-incongruent perceptual content is not covered by amnesia, a subject may accurately remember having had a feeling of conviction only in relation to suggestion-congruent imagery content. Since a feeling of conviction that one is perceptually related to some content is most typically experienced when the content is in fact perceived, some mistaken first-person recollections may thus reflect a relatively simple attribution error.

In the previous sub-subsection, we mentioned that similar dissociation or misplacement of a feeling of conviction may constitute the phenomenological core of actual clinical disorders of delusion. In the present sub-subsection, we shall examine some non-clinical examples as well, which hopefully serve as good analogies to illustrate the gist of the idea, and which further underscore the plausibility of an affective account of suggested hallucination.

4.8.2.2.1 Displacement of affect

The idea of misplaced feelings of conviction calls to mind what Freud (e.g., 1900/1953a, 1901/1953b, 1904/1960, 1916/1963) described as displacement. This purported ego defense mechanism involves the (unconscious) transferal of affect associated with some threatening impulse or object to a more acceptable or immediate substitute. A classic example is shifting anger felt toward one's boss (the expression of which is perceived as potentially dangerous) to one's pet or spouse (who are perceived as less likely to cause harm), and therefore kicking the pet or picking a quarrel with one's spouse after work.

Of course, this is not to imply that feelings of conviction are displaced from subjectively more to subjectively less threatening targets. Nor is the assumption that suggested hallucination has anything to do with ego defense mechanisms. A relatively modest point is merely that there is a long tradition in psychology that assumes that affect can dissociate from its original source and/or bind with some apparently arbitrary target in a motivated way.

A more specific point in connection with displacement theory relates to Freud's primary area of interest in application: dreams. One need not commit to the particularities of the theory to agree that in dreams, imagery and affect often fuse and dissociate in ways that do not occur in normal waking consciousness. For example, say Mary dreams of a person who looks like Peter. She may still experience a feeling of conviction that the person in the dream is John. She might accordingly tell John the next day that she dreamt of him, although, funnily, in her dream he looked like Peter.

But say John had looked like himself—or at least how Mary tends to perceive or imagine him—in Mary's dream. Under normal waking conditions, Mary would unlikely have associated such imagery with a feeling of conviction that she is in real perceptual contact with John. Yet such feelings also seem to accompany lucid dreams, which involve awareness that

one is dreaming. Which is to say, a feeling of conviction that some dream content is perceptually or externally real may occur even if one concurrently entertains a discordant belief that the feeling is misplaced. Lucid dreams may thus provide powerful existential proof for the idiosyncratic attachment of feelings of conviction to suggestion-facilitated imagery content.

4.8.2.2.2 Feeling of familiarity

Déjà vu is the feeling of having already said or done or experienced something before, or of seeming to know what one will say or do or experience next, as if the present were a memory of some past experience, while in actuality one cannot recall any particular relevant memory, for there is nothing to recall. Although the expression literally means "already seen" in French, the experience can involve any modality. The most widely accepted definition of déjà vu accordingly states that it is "any subjectively inappropriate impression of familiarity of the present experience with an undefined past" (Neppe, 1983, as cited in Neppe, 2015, p. 2).

This phenomenon of misplaced feeling of familiarity bears relevance for multiple reasons. First, in addition to some pathological conditions such as temporal lobe epilepsy, schizophrenia, and amphetamine psychosis (Brown, 2003), the phenomenon is relatively common in healthy individuals as well. Although exact numbers are hard to provide, most researchers estimate the general prevalence of the phenomenon to be somewhere between 60% and 80% (Brown, 2003; Neppe, 2015; Probst & Jansen, 1991; Sno & Linszen, 1990), whereby possibly more than 97% of people experience déjà vu at least once in their lifetime (O'Connor, Barnier, & Cox, 2008).

Second, similarly to the case of lucid dreams, the effect can be belief-discordant. This is not to say that some people in some cases may not draw mistaken conclusions from their experience. It has been suggested accordingly that déjà vu experience may underlie various doctrines of pre-existence, reincarnation, and the transmigration of souls. The effect similarly

fascinates the parapsychological community (Brown, 2004/2005). All the same, many people in many cases are well aware that they are only experiencing déjà vu, and their construing or describing their experience as such does not make the feeling any bit less intensive or awkward.

Third, there seems to be universal agreement among researchers that the most distinctive phenomenological feature of the phenomenon is its affective quality. An important strand of theories accordingly posit that the effect occurs when there is some discrepancy between affective and cognitive functioning. For example, some argue that the experience involves an implicitly activated emotional association that cannot be explicitly connected to its source (Pagliaro, 1991; Siomopoulos, 1972, as cited by Brown, 2003, 2004/2005).

Others propose that the effect is explained by aberrant activity in the neural pathway responsible for the affective appraisal of percepts (Fleminger, 1991). Of course, there are non-affect-based rival explanations as well. But, again, even those seem to grant that the crucial experiential factor is affect—a feeling of familiarity or a feeling of conviction of having already undergone the experience—rather than distorted belief (delusion) or distorted memory (paramnesia) or some other cognitive element.

Fourth, crucially, the experience of déjà vu seems experimentally inducible under posthypnotic suggestion. In the first such study, subjects were shown some pictures before hypnosis was induced. Following induction, the subjects were then delivered a posthypnotic suggestion to forget the presented pictures after hypnosis was terminated. When hypnosis was in effect terminated, and the subjects were again shown the pictures, about 30% reported déjà vu-like experiences despite apparent amnesia for the pictures (Banister & Zangwill, 1941, as cited in O'Connor et al., 2008).

An issue with the above study, though, is that even if subjects were in fact temporarily amnesic to the pictures, their feeling of familiarity or conviction may have been driven by their

actual prior visual experience of the pictures. To correct for this methodological flaw, a more recent study compared déjà vu experiences following suggestion for posthypnotic amnesia with déjà vu experiences following suggestion for posthypnotic familiarity. Subjects in the posthypnotic amnesia group completed a puzzle game both before and after hypnosis, whereas subjects in the posthypnotic familiarity group completed the game only after termination of hypnosis. All the same, more than half of the subjects experienced a strong sense of familiarity in either group.

Yet there was a significant difference between the groups regarding déjà vu experiences. In the posthypnotic amnesia group, no subject experienced a strong sense of déjà vu, and only half experienced a weak sense of déjà vu (0 strong, 3 weak, 3 nothing). In contrast, in the posthypnotic familiarity group, all but one subject experienced déjà vu, which in all cases involved a strong sense of déjà vu (5 strong, 0 weak, 1 nothing). Post-experimental questioning revealed accordingly that whereas the experiences of subjects in the posthypnotic amnesia group did not quite resemble spontaneous déjà vu, the experiences of most subjects in the posthypnotic familiarity group did.

Particularly interesting and relevant to the issue of providing an affective account of suggested perceptual hallucination is that subjects in the posthypnotic familiarity group also talked much more often about the sensory features of the puzzle game, such as the colors of pieces or the feeling or sound of pieces sliding. Which is to say, their strong sense of déjà vu was associated with strong feelings of conviction that they had experienced particular sensory features before. And all this without having had to imagine the sensory features beforehand. Little wonder, then, that many hypnotic “reals” can pass lie detection regarding earlier suggested hallucinations.

4. Suggested Hallucination

A final important point about déjà vu is that its opposite also exists: jamais vu. The latter expression means "never seen," but this phenomenon may also involve any modality. For example, a person may walk into his or her own house and still feel as if the house were unfamiliar. Or consider the related phenomenon of semantic satiation: if a word is repeated over and over again, it can momentarily lose its connotative meaning (Brown, 2003).

By far the best known example of an inappropriate feeling of unfamiliarity—or an inappropriate lack of feeling of familiarity—is Capgras delusion. People afflicted with this syndrome appear to believe that a close relative or loved one has been replaced by an identical-looking impostor. Although their capacity for facial recognition is intact, on a prominent theory, subjects do not experience the warm and fuzzy feeling or emotional "glow" that normally—even if occasionally only very dimly—accompanies recognition of loved ones. The absence of such a feeling of familiarity may thus lead to surprise, which may in turn lead to rationalization or explaining away of the experience (Ellis & Young, 1990; Ramachandran & Blakeslee, 1998).

4.8.2.2.3 Feeling of reality

We have been discussing some analogous examples of misplaced feelings of conviction under suggested hallucination. We started with emotional displacement in dreams, and the possible persistence of a feeling of conviction about the perceptual or external reality of dream content during lucid dreaming. We then turned to the phenomenon of déjà vu, which seems to exemplify a detachment from belief of a strong feeling of familiarity. In contrast, its opposite, jamais vu, and related phenomena such as semantic satiation and Capgras syndrome, underscore that one may also lack or lose a sense of familiarity. So there is good evidence for a double dissociation between belief and the feeling of familiarity.

4. Suggested Hallucination

Now the (lack of) feeling of familiarity is closely related to a (lack of) feeling of reality. Recall for example that in Capgras delusion, subjects seemingly lack a feeling of familiarity in relation to their loved ones. What if a person loses the feeling of familiarity in relation to his or her own body or self? That results in depersonalization, which is characterized by a feeling of unreality or detachment from—as if one were an outside observer of, or unfamiliar with—one's thoughts, feelings, sensations, body, or actions (American Psychiatric Association, 2013).

Accordingly, a feeling of unreality or detachment from, or unfamiliarity with, the world—be it individuals, inanimate objects, or all surroundings—is called derealization. Some subjects describe this feeling as if being in a fog, dream, or bubble; or as if there were a veil or a glass wall between them and the world around; or as if the surroundings were artificial or lifeless (ibid.).

Dokic and Martin (2012, 2015, 2017) argue accordingly that the feeling of reality constitutes the relational phenomenology of perception whereby one experiences being related to actual things rather than mere representations of them. They emphasize that this sense of reality is two-way independent from the spatial and sensory features of perceptual experience. Thus, just as the feeling can be lacking in cases such as derealization disorder, so it can accompany unrealistic hallucinations. And just as one may believe in external reality even in the absence of this feeling, so the feeling may persist even if one is aware of hallucinating or imagining.

Hence, akin to the feelings of conviction and familiarity, the feeling of reality may itself detach from perceptual experience (apparent negative hallucination) and attach to imagery experience (apparent positive hallucination). Which suggests that such distortions of affect may themselves explain at least some cases of suggested hallucination. For example, say a subject

is delivered the suggestion that a blue image is red. On the present account, the subject will experience no relevant change in color phenomenology. No relevant change need occur in online perceptual or first-person phenomenal belief, either. For the subject's feeling of reality may independently detach from perceptual experience of blue and attach instead to conscious imagery of red.

There is in fact some pretty good evidence for this. Recall (from *Subsection 4.7.2*) the study of two hypnotic virtuosos in which subjects received (posthypnotic) suggestions that certain shapes would always appear in certain colors (e.g., "all triangles are red"). Considering for example the shortness of stimulus exposure time (24 ms), we discussed earlier that one subject plausibly confabulated congruent perceptual pseudomemories (e.g., she would misremember having seen blue triangles as if red). Although the other subject also responded to suggestions and showed alterations in brain activity,

> She reported after the behavioral and EEG sessions that sometimes she experienced a conflict between what she saw and what she "felt" the color is or that "sometimes I saw the shape as red (or blue) [i.e., in the attended-to color], but my brain said it had a different color." (Koivisto et al., 2013, p. 4)

Follow-up testing for this subject was accordingly devised, in conformity with the method of differences. In one stimulus block, it was stressed to the subject that she should only respond to what she actually saw, and to ignore completely what she felt. In this condition, the subject did not report any color change for the target shape. In a second stimulus block, the subject was then asked to respond only according to what she felt and not what she saw. In this condition, the subject reported color change for 75% of the target shapes.

In light of both her retrospective self-report and the follow-up results, the authors concede that this subject did not experience real color hallucinations. More in line with an

affective view, we may note, she seems to have experienced misplaced feelings of reality and/or conviction.[28] This conforms with a prominent view of hypnosis on which, "in the classic case, . . . experiences are associated with a degree of subjective conviction bordering on delusion, and an experienced involuntariness bordering on compulsion" (Kihlstrom, 2008, p. 21). Insofar as conviction is of course subjective, a charitable or plausible reading of the respective qualification is that subjects are not (necessarily) convinced of the reality of the suggested state of affairs in a doxastic sense. Rather, it is a misplaced feeling of conviction—which may be either identical or closely related to a misplaced feeling of reality—that lies at the core of the phenomenon.

4.8.2.2.4 Feeling of knowing

The feeling of knowing is often just another label for what is variably referred to as the emotion of conviction (Bagehot, 1871/1915; James, 1890), emotional conviction (Abelson, 1988; Russell, 1914; Sutcliffe, 1961), subjective conviction (Kihlstrom, 2008; Koriat, 2000), and the feeling of conviction (Braithwaite, 1933; Bruno et al., 2012; Koriat, 2000; Turner & Coltheart, 2010). Akin to the feeling of familiarity, it is widely presented as a paradigm example of a category of affect constituted by epistemic or noetic or metacognitive feelings. Feelings of this sort are construed as phenomenal experiences that point towards a subject's own mental capacities or processes or dispositions (Arango-Muñoz, 2014a, 2014b; Arango-Muñoz & Michaelian, 2014; de Sousa, 2008; Koriat, 2000). A most common example is the tip-of-the tongue effect, whereby one may have a feeling of knowing some sought-after name without actually being able to produce the name. Even if such "hunches" have no articulate content, the respective knowledge feels self-evident and not in need of justification (Koriat, 2000).

[28] Unfortunately (and from a methodological point of view: incomprehensibly) no follow-up was conducted with the other, plausibly confabulating subject. So we cannot tell whether she may have experienced similar feelings, either before or during her apparent confabulations regarding prior perceptual experiences.

4. Suggested Hallucination

It has been hypothesized correspondingly that a lack of feeling of knowing, or a feeling of error, may characterize the core symptoms of obsessive–compulsive disorder (OCD). On this view, it is not that OCD patients incorrectly perceive or construe or remember things related to their obsessions or compulsions. Rather, what explains their apparently irrational and repetitive mental action or behavior is that it feels to them as if something went wrong with the execution of respective actions or behavior, even when no error has in fact been made. Accordingly, even if entertaining evidence to the contrary, they may still feel as if things were "just not right" (Kropotov, 2016; Rapoport, 1989, as cited in Woody & Szechtman, 2000a, 2000b, 2007).

Extending the above hypotheses to other psychopathologies as well as to hypnotic hallucinations, the latter authors propose that a feeling of knowing may provide the biological underpinnings of the sense of external reality. On this proposal, negative suggested hallucination may resemble OCD insofar as subjects lack an appropriate feeling of knowing. Positive suggested hallucination may in turn resemble schizophrenia and the effects of marijuana insofar as even mundane, everyday stimuli may gain (illusory) epistemic relevance due to a feeling of knowing. Hence, the authors argue, "some of the subjective impressiveness and convincingness of hypnotically suggested phenomena, such as hallucinations, may be due to vivid feelings of knowing, rather than vivid sensory qualities per se" (Woody & Szechtman, 2000b, p. 10).

Is there any evidential support for the idea that suggested negative hallucination resembles OCD in that subjects act as if they didn't know what they do because it feels to them that way? It would seem so, yes. To start with, OCD involves hyperactivation of the anterior cingulate cortex (AAC) (Kropotov, 2016). This is the same area that shows increased activation under affective pain (cf. *Sub-subsection 4.5.1.1*). Accordingly, the area is traditionally

associated with affect regulation, nowadays particularly with the mediation of cognitive influences on emotion (Stevens, Hurley, & Taber, 2011). In addition, the area has been implicated in error detection and correction. An integrative proposal is accordingly that the ACC forms part of an attentional circuit that regulates both cognitive and emotional processing (Bush, Luu, & Posner, 2000).

So say a subject receives a suggestion that all triangles are red. What might happen upon the presentation of a blue triangle? If the subject is highly suggestible, then despite knowing that the triangle is perceived as blue, due to faulty affect regulation and conflict monitoring, the ACC might incorrectly signal error and tag for further conscious checking the respective blue-related belief (cf. Turner & Coltheart, 2010). Similarly to OCD patients, then, the subject may lack an appropriate feeling of knowing in relation to the blue color of the triangle because of an inappropriate feeling of error that there is something just not right about the blue-related belief.

What about cases of suggested positive hallucination? This is where Woody and Szechtman appear to depart from a non-credulous affective view. For what they essentially propose is a two-step causal model of perceptual hallucination on which

> the feeling of knowing comes first, and then the subject constructs and rationalizes a percept to go with it. . . . [I]f the conviction arises that something is out there, the situation provides very strong and clear expectations around which to construct the accompanying percept. (2000b, p. 10)

The authors thus appear to maintain the core credulous supposition that suggested hallucination involves "a dissociation between perceptual experience and reality" (2007, p. 245). Hence, unless they use the terms "percept" and "perceptual experience" liberally and idiosyncratically, their two-step model is ultimately just a causal variant of the credulous view.

However, a single-step affective model is not only more parsimonious, but also more fitting with the very data to which the authors appeal. Some of these data come from an already multiply cited neuroimaging study of auditory hallucination of which Szechtman was the first author (Szechtman et al., 1998; cf. *Sub-sub-subsection 4.2.3.4.2* and *Sub-subsection 4.5.1.2*). Recall that the study found similar activity in the ACC in the hearing and hallucination (but not the imagination and baseline) conditions among highly hypnotizable "hallucinators" (but not their "nonhallucinator" counterparts). Since activity in the ACC also showed strong positive correlations with subjective ratings of externality and clarity, the authors (mis)took these findings to support a credulous position.

Yet, as mentioned above, the ACC is traditionally associated with affect regulation on the one hand (thus the association with pain affect), and error detection and correction on the other hand (thus the association with OCD). An integrative account accordingly posits that the ACC forms part of an attentional circuit that regulates both cognitive and emotional processing. Szechtman et al. proposed accordingly that the attention of "hallucinators" is in general more affect-laden than that of "nonhallucinators." This meshes with their finding that "hallucinators" processed auditory events more extensively (involving more brain areas) than "nonhallucinators" in the hearing condition, including in the right rostral region of the ACC.

Further and most crucially, even if "hallucinators" showed similar ACC activity in the hearing and hallucination conditions, it was in fact "nonhallucinators" who showed similar activity in auditory association areas across these conditions. A credulous interpretation thus implausibly implies that differential activity in the auditory association areas is unnecessary, whereas differential ACC activity is sufficient for auditory hallucination. This is inconsistent with current neuroscientific orthodoxy. A two-step model of suggested hallucination on which

altered feelings of knowing lead to corresponding perceptual hallucination is thus unwarranted. The results conform much better with a single-step affective account of the phenomenon.

4.8.2.3 Final possible objections

We have come a long way from skeptical, credulous, and cognitive-delusion views of suggested hallucination. We started our defense of an affect-based account by arguing that not only is suggested analgesia compatible, but it is in fact a most illustrative example of the kind of affective distortion that may explain more standard modes of suggested hallucination as well. In relation to the latter phenomena, we started our inquiry by revisiting the idea that suggestion-congruent changes in the feeling of conviction may lie at the core of apparent delusions. Moving on to dream states and then the feeling of familiarity, we eventually arrived at the feelings of reality and knowing. As we saw, these feelings are closely connected. For example, lucid dreaming may involve a feeling of conviction that one is in perceptual contact with external reality; a loss of familiarity with oneself or one's surroundings may involve corresponding losses of a feeling of reality; and on some prominent theories, the feeling of knowing is more or less the same as the feeling of conviction and the feeling of reality.

A detailed account of epistemic feelings such as relating to conviction, knowing, and reality are beyond the scope of the present dissertation. However, we can still meet what appear to be some most burning possible objections against an affect-based view of suggested hallucination. The first is that the best neurophysiological evidence most plausibly indicates relevant distortions in perceptual phenomenology. The second is that the phenomenal properties of feelings of conviction, knowing, and reality are better construed as cognitive phenomenology. Finally, the third objection is that the affective view cannot make best sense of incomplete or trance-logical phenomena such as transparent hallucinations.

Since even these objections can be met, it is argued, an affect-based view of suggested hallucination must be taken more seriously than just a fringe idea. Indeed, if the account is on the right track, it has far-reaching implications for current debates of cognitive penetrability. So there is good reason for future discussions to widen and focus more on issues of suggested hallucination and related feelings.

4.8.2.3.1 Objection from neurophysiology

We ended the previous sub-subsection with a reassessment of a most widely celebrated neuroimaging study of suggested auditory hallucination. Since relevant changes in brain activity for suggested hallucination only occurred in the ACC but not the auditory association cortices; and since the lead author soon after himself proposed as second author a two-step model of hallucination on which the first step involves changes in feelings of knowing; a reinterpretation of the data in line with an affective-delusion account of suggested hallucination is not that radical of a move.

However, some might argue that the color-adding/color-draining neuroimaging study of suggested color hallucination mentioned first at the beginning of this chapter (*Section 4.1*), and then again in relation to early visual processing (*Sub-subsection 4.5.1.2*), is not amenable to such reinterpretation. Recall that in this study, Kosslyn et al. (2000) found that irrespective of whether an image was actually colored or grayscale, hypnotic suggestion to perceive the image as if colored/grayscale (but not corresponding imagery instructions) correlated with a respective increase/decrease in a classic color area in the left fusiform gyrus (LFG) of the brain. How is this not knock-out evidence for suggestion-congruent perceptual hallucination?

We already discussed some counter-considerations in the sub-subsection on early visual processing. For example, there was no difference in LFG activity for color vs. grayscale stimuli in the nonhypnotic-imagination condition. Further, in a more recent study, LFG activity

increased for both positive and negative color hallucination (McGeown et al., 2012). Still further, even in the original study, the absolute level of LFG activity was in fact similar for "color-added" grayscale and "color-drained" color stimuli. Such inconsistent results militate against the view that differential LFG activity reliably reflects differential color experience.

There are also positive considerations, which suggest that the findings may be better explained by differential attention, knowledge retrieval, and/or affect. Starting with differential attention, subjects in a study had to perform two tasks simultaneously. Increased memory load was associated with increased task interference by distractor faces. Increased task interference may reflect decreased early-selection efficiency. Since stimuli that are not selected out early on undergo further processing, it was thus expected that a brain region closely associated with face processing called the fusiform face area would show increased activation. And, indeed, it did: neuroimaging indicated that higher memory load was associated with increased fusiform activity in the respective area (de Fockert, Rees, Frith, & Lavie, 2001; cf. also Treisman, 2009).

Should we then assume that subjects perceived distractor faces as visually different under high memory load? The answer is most plausibly no. So even if certain fusiform activity is closely associated with face perception, it doesn't follow that such activity is a reliable neural marker of visual face experience. Analogously, even if LFG activity in the classic color area is closely associated with color perception, it doesn't follow that activity in this area reliably indicates color experience. Perhaps in some cases, also similarly to the case of face perception, differential activity in the LFG color area merely reflects differential attention to color.

LFG activity is also implicated in color knowledge retrieval. In the first part of a neuroimaging study that tested for this, subjects were asked on each trial to verify whether a named color or motor property was true of a named object (e.g., taxi-*yellow*, hair-*combed*). In the second part of the experiment, subjects performed a color perception task. Among other,

partly non-overlapping areas of activation, activity in the same LFG area was observed during both color perception and color knowledge retrieval, but not motor-property knowledge retrieval. Surely, this should not be interpreted as suggestive of color hallucination during color knowledge retrieval. The very point is that the LFG color area is plausibly involved in both perceptual and non-perceptual activity related to color (Simmons et al., 2007; cf. also Martin, 2009; Papanicolaou, 2017).

Finally, and of most relevance regarding an affective model of suggested hallucination, there is also suggestive neurophysiological evidence from perception research involving both healthy and brain lesioned subjects that differential LFG activity might mirror differential affective responses. For example, faces with fearful emotional expressions produce greater activation in the fusiform face area of healthy subjects than faces with emotionally neutral expressions. Similarly, both vocal and non-vocal sounds produce greater activation in relevant auditory areas when associated with emotional significance (cf. Vuilleumier & Brosch, 2009).

Importantly, the association between sensory features and affect need not be intrinsic or innate. This is underscored by increased activity in relevant sensory areas for previously neutral stimuli following aversive conditioning. Accordingly, for both conditioned and non-conditioned stimuli of emotional relevance, increased activation in relevant sensory areas correlates with increased activation in the amygdala, a foremost area associated with emotion. Indeed, it seems that the more sensitive the amygdala is to affective properties of stimuli, the higher the activation is in relevant sensory areas for respective stimuli. This meshes or is even further supported by that patients with amygdala damage do not show selectively increased sensory activity for emotionally relevant stimuli (ibid.).

Hence, converging evidence from perception research unrelated to hypnosis and suggestions provide good general support for the idea that altered activity in the LFG color

area may reflect altered color-related affective responses. Such responses may then explain (away) some prima facie puzzling findings concerning suggested hallucination. For example, LFG activity was increased for color stimuli during hypnotic but not non-hypnotic baseline in the original color-adding/color-draining study. Never mind that subjects were asked to perceive the stimuli veridically in both conditions. What is hard to make sense of on a credulous view is how saturated and bright color stimuli may be perceived as if even more colored under hypnosis. A related challenge for a cognitive-delusion model is to explain in what sense a subject may hold the belief that an image is colored to a greater or lesser extent. On the affective model, there is no analogous puzzle: a feeling of conviction or knowing may increase and decrease in intensity independently of corresponding alterations in either perception or belief.

4.8.2.3.2 Objection from cognitive phenomenology

All things considered, we may conclude that suggested hallucination is a psychologically real phenomenon that involves neither perceptual hallucination, nor cognitive delusion. Which is to say, as generally construed, the skeptical, credulous, and delusion accounts are all mistaken. It follows accordingly that subjects are not merely vividly imagining the suggested state of affairs. Crucially, they seem to experience strong feelings of conviction or knowing or reality in connection with suggestion-congruent imagery and thought, meanwhile they appear to lack such feelings in connection with suggestion-incongruent perceptual experience and belief.

Yet, despite all the arguments and evidence presented thus far, perhaps some may still wonder whether the respective feelings are better construed as cognitive. For example, it may be argued that feelings of the mentioned epistemic kind cannot be affective insofar as they lack evaluative appraisal, action tendencies, valence, and/or arousal. Accordingly, it may be noted, they do not resemble either of four paradigmatic categories of feelings: emotions, such as happiness and anger; moods, such as feelings of cheerfulness and tranquility; general bodily

conditions, such as hunger and sleepiness; or feelings of behavioral tendency, such as feeling amorous or talkative. Hence, it might be concluded, insofar as the respective feelings are neither perceptual nor affective, their respective phenomenology is most plausibly cognitive.

How might a proponent of the affective view respond? Some philosophers (e.g., Carruthers, 2017; de Sousa, 2008; Prinz, 2011) as well as psychologists and psychiatrists (e.g., Woody & Szechtman, 2002) argue that epistemic feelings are indeed emotions. For example, consider evaluative appraisal. On the face of it, epistemic feelings do not appraise the relevance of things or events or situations to our goals or values, in a way that fear of height can represent danger or sadness about death can represent loss. For even if they might represent a thing or event or situation as, say, known or familiar, such properties do not seem evaluative of those things or events or situations.

However, recall that epistemic feelings point towards a subject's own mental capacities or processes or dispositions. Accordingly, their appraisals bear on restricted domains of epistemic values such as the quality of one's knowledge (de Sousa, 2008) or cognitive competence at a given task (Dokic, 2012) or execution of some mental process (Arango-Muñoz & Michaelian, 2014). Misplaced feelings of conviction or knowing or reality under suggested hallucination may thus misappraise for example the source or quality of one's imagery or his or her success at perceptual hallucination.

Now consider action tendencies. As opposed to paradigmatic emotions such as fear and anger, epistemic feelings do not seem associable with particular facial expressions or bodily postures or approach or retreat behaviors. However, some argue that like all emotions and fringe feelings (James, 1890), epistemic feelings are also the result of feedback on bodily changes (Prinz, 2011), and as such they also carry motivational properties originating from their evolutionary origins in solving problems of survival (Woody & Szechtman, 2002).

Accordingly, epistemic feelings may play a role in the guidance of mental activity by generating action tendencies for example to reason in certain ways, or to continue or discontinue further attempts at information retrieval (de Sousa, 2008; Koriat, 2000).

Under suggested hallucination, then, feelings of conviction or knowledge or reality may perhaps play a role in generating tendencies to form acceptances. Of course, it doesn't follow that a subject in fact accepts any respective content. Recall for example (from *Sub-sub-subsection 4.8.2.2.3*) the hypnotic virtuoso who claimed that she saw the color of shown images correctly, but "my brain said it had a different color" (Koivisto et al., 2013, p. 4). As noted earlier (in *Sub-subsection 4.6.1.3*), it is not an actual loss of cognitive control but this compulsion-like feeling or experience of involuntariness that is widely considered "the classic suggestion effect" (Weitzenhoffer, 1974).

What about valence? Emotions are paradigmatically positively or negatively valenced with respect to objects towards which they are directed. Some argue that this is also typically the case with epistemic emotions. However, just as they appraise the quality of one's own cognitive dispositions or processes or competence, so their positivity or negativity of affect is directed at the epistemic value of these dispositions or processes or competence (Arango-Muñoz, 2014a, 2014b; Carruthers, 2017). Accordingly, under suggested hallucination, a strong feeling of conviction in the absence of a feeling of error may well be positively valenced.[29]

Analogously to valence, which is intimately connected with evaluative appraisal, physiological arousal is intimately connected with action tendencies. Those who hold that epistemic feelings are phenomenological manifestations of evolutionarily prepared bodily reactions (James, 1890; Prinz, 2011; Woody & Szechtman, 2002) imply that arousal may thus

[29] In turn, when receiving for example motor challenge suggestions for arm rigidity or eye catalepsy, the feeling of being incapacitated may be negatively valenced, though perhaps counterbalanced by curiosity or surprise.

characterize such feelings. Some argue so explicitly (Carruthers, 2017). Accordingly, under suggested hallucination, similarly to OCD, increased feelings of error may be associated with tension arousal; whereas, more closely resembling the effects of marijuana, increased feelings of conviction may be associated with relaxed arousal.

Hence, one possible reply to the objection from cognitive phenomenology is that feelings of conviction and knowledge and reality are indeed emotional feelings, since they typically do satisfy such conditions as evaluative appraisal, action tendencies, valence, and arousal. However, a defender of an affective interpretation of suggested hallucination need not commit to so strict criteria of emotionhood. For example, it may be noted that surprise—which is counted among the six basic emotions alongside anger, disgust, fear, happiness, and sadness (Ekman, 1992)—is itself often unvalenced; and even when it is valenced, depending on its object or content, it can take up either positive or negative valence (Carruthers, 2017).

Relatedly, when improvement in some domain is thought to be outside one's control, or when a positive feeling of admiration arises from an upward social comparison, admiration seems to be associated with passive inspiration that does not lead to relevant action tendencies (van de Ven, Zeelenberg, & Pieters, 2011). Or consider a prominent view in modern attitude and affect research in psychology according to which "preferences need no inferences" (Zajonc, 1980, pp. 151, 160). On this view, it is mistaken that emotions essentially depend on appraisal, both on account that emotion can precede cognition, and on account that feeling and thinking are underserved by partially independent systems.

The upshot of these examples is that there is plenty of theoretical room to accommodate nonstandard emotions that may lack one or more or even most standard features of emotion. This is not to imply, though, that a proponent of an affective interpretation of suggested hallucination must defend by all means a view on which feelings of an epistemic or noetic or

metacognitive kind are emotions. A further alternative is to grant that the respective feelings are not emotions, but to argue that they are still affective. This is consistent with a distinction in the literature between theories of epistemic emotions and theories of epistemic feelings (Arango-Muñoz & Michaelian, 2014).

Does it follow that feelings of conviction or knowing or reality are moods like cheerfulness, general bodily conditions like hunger, or feelings of behavioral tendency like feeling talkative? No, it doesn't. Even if there are differences between epistemic feelings of different kinds, a key theoretical motivation of their study is that they may constitute a proprietary affective category. Although assuming so may be less parsimonious than subsuming respective feelings under some more standard affective category, it is hardly more parsimonious to construe respective experiential features as reducible to proprietary cognitive phenomenology. Here are some relevant considerations.

Firstly, recall that at least many subjects under suggested hallucination unlikely (doxastically) believe in the (online) reality of the suggested states of affairs. Still they may experience genuine effects. Second, nor are subjects merely imagining or hoping or wishing or desiring the suggested effects. As genuine as one's intention and effort to undergo relevant experiences may be, if a subject is not suggestible, he or she will unlikely experience genuine suggested hallucinations.

Third, the very idea of dissociated or misplaced affect assumes that subjects may experience the same feeling not only in relation to different content, but also across different attitudes and states and modes. This suggests that the phenomenology in question cannot be particular to any first-order attitude or state or mode. However, fourth, it is unclear whether or how some general background cognitive phenomenology might explain specific alterations in experience unless that phenomenology were itself sensitive to the influence of suggestion. This

suggests that the prima facie most plausible account of suggested hallucination in terms of distorted cognitive phenomenology is one that posits the distortion of relevant metacognitive representations.

Yet, fifth, it would seem that at least some epistemic feelings are nonconceptual and/or opaque with respect to first-order representational content (Arango-Muñoz, 2014b; Arango-Muñoz & Michaelian, 2014; Dokic, 2012). Sixth, there seems to be a double dissociation between such feelings and corresponding metacognitive representations. For, on the one hand, the respective feelings often precede or occur in the absence of any apparent metacognitive representation. Indeed, it would seem that it is often the interpretation of the feeling or affective response that produces corresponding metacognitive judgment, rather than the other way around. On the other hand, in at least some instances, one may well be aware of metacognitive content without any corresponding feeling (Arango-Muñoz, 2014b).

Seventh, even if epistemic feelings are in normal conditions sensitive to certain properties of cognitive processes such as their fluency; and even if they may appear to misrepresent such properties in certain conditions such as suggested hallucination; strictly speaking, they are unlikely metacognitive. This is not merely to say that they are not explicitly metarepresentational; arguably, they are not even implicitly metarepresentational. As Carruthers (2017) concludes,

> one can only continue to regard them as implicitly metacognitive by defining "metacognition" in such weak terms that we would have to allow that human and animal cognition is almost ubiquitously metacognitive. This drains such a claim of any of the interest that might otherwise attach to it. (p. 74)

Finally, and perhaps most crucially, it is unclear whether or how metacognitive phenomenology might explain (away) first-person accounts of suggested hallucinations such

as when a subject experiences a conflict between what is seen and what is felt, or when it appears as if one's brain were saying a different color than actually perceived (cf. *Sub-sub-subsection 4.8.2.2.3*). Such cases clearly do not involve any (meta)cognitive delusion. So how could the respective phenomenology be (meta)cognitive? All things considered, it is much more plausible that the feeling is affective.

Still, perhaps those subjects who confabulate pseudomemories of their perceptual experiences indeed undergo metacognitive distortions and hence experience corresponding metacognitive phenomenology? All things considered, the answer is probably still no. But assume that the answer is yes. Then some cases of suggested hallucination may involve distortions in both affective and cognitive phenomenology. This is an interesting possibility, but it has no bearing on the two central theses of this chapter: one, that suggested hallucination does not involve perceptual hallucination and hence is not a case of cognitive penetration of perceptual experience; and, two, that a core feature of some most relevant effects of suggested hallucination (including but not limited to suggested analgesia) is a distortion in affect. From this point, defenders of cognitive phenomenology are welcome to fill in the details.

4.8.2.3.3 Objection from transparent hallucination

Suggested hallucination is a psychologically genuine phenomenon with a most exciting twist: at its core are changes in affective experience that typically manifest in suggestion-congruent (loss of) feelings of conviction, knowing, and reality. After meticulous argumentation covering a wide range of theoretical and empirical issues, including in the last two sub-sub-subsections possible objections from neurophysiology and cognitive phenomenology, we may now address a final possible challenge: the objection from transparent hallucination.

Recall that as opposed to simulators, many hypnotic "reals" report "seeing through" their hallucinations. On a credulous interpretation, the phenomenon demonstrates a subject's

ability "to mix freely his perceptions derived from reality with those that stem from his imagination and are perceived as hallucinations" (Orne, 1959, p. 295). We already discussed why a fusion of perceptual and hallucinatory experiences is implausible for example in relation to phenomenal-color mixing. Insofar as the respective experiences "are fused in a manner that ignores everyday logic" (ibid.), such cases of "trance logic" also seem hard to explain on any purely cognitive-delusion account.

However, perhaps it may be objected that the phenomenon is also inconsistent with the hypothesis of misplaced affect. For, the challenge may go, on the proposed model, it would seem that the respective feeling of conviction or knowing or reality should attach either to perception or imagery but not both. So perhaps a perceptual-fusion model is more plausible after all. How can the affective model dissolve the apparent puzzle?

A first consideration is that of course subjects can "see through" their "hallucinations": the latter are only mental images. A second consideration is that the detachment from perception and the attachment to imagery of relevant feelings is not a zero-sum affair. So even if one is affectively absorbed in suggestion-congruent imagery, it doesn't follow that s/he has succeeded in affective disengagement from suggestion-incongruent perceptual experience.

Relatedly, that multiple electrophysiological studies have found incomplete responding to be the norm rather than the exception—especially when obliterating (negative) rather than obstructive (positive) suggestions are delivered (cf. Barabasz & Barabasz, 2008; Jensen et al., 2001)—underscores that affective disengagement from perceptual experience is likely much harder than affective engagement with imagery experience.

So in cases involving duality experience such as transparent hallucination, perceptual and imagery experience may both associate with a feeling of conviction or knowing or reality. In a sense, then, socio-cognitive theorists were right to claim that the phenomenon constitutes

incomplete responding (Spanos, 1986; Spanos et al., 1983). For no feature of the effect involves relevant changes in perceptual experience.

In another sense of incomplete responding, though, it was credulous proponents of a trance-logical fusion of perception and imagery who were right. For even if subjects succeed in eliciting or increasing or attaching relevant affective experience to imagery, they may still fail to eliminate or reduce or detach such feelings associated with perceptual experience. In which case both types of content may involve feeling perceptually related to external reality.

Further, there is a relevant sense in which the respective experiences may indeed be phenomenally unified. Of course, even in normal waking consciousness, perceptual and imagery experiences are typically tied together into a single, all-encompassing, unified state of consciousness. A plausible particular feature of transparent hallucination is that the respective experiences are also affectively unified.

What distinguishes between transparent (incomplete) and non-transparent (complete) suggested hallucination, then, is not whether or which features or to what extent perceptual content is penetrated by hallucinatory content. Nor is it whether a subject doxastically believes in the suggested state of affairs. A more plausible difference is whether or to what extent or in what ratio feelings of conviction or knowing or reality can be dissociated from suggestion-incongruent perceptual experience meanwhile associating them with suggestion-congruent imagery. Insofar as achieving the former is harder than achieving the latter, it is only expected that a significant portion of suggestion responses will be accordingly incomplete.

Hence, far from posing theoretical embarrassment or casting empirical doubt on the reality of misplaced feelings of conviction or knowing or reality, the phenomenon of transparent hallucination only underscores the relative abductive power of an affective model of suggested hallucination.

4.9 Conclusion

Is suggested hallucination a case of cognitive penetration? Decades of credulous research and argumentation to the contrary notwithstanding, there is little reason to believe that it is. Yet there is also little reason to assume a skeptical or purely cognitive position. So it is high time that philosophers and psychologists alike seriously (re)consider the proposal that suggested hallucination is at its core an affective affair.

If the present analysis of the phenomenon is on the right track, then the likelihood of finding convincing cases of cognitive penetration of perceptual experience cannot be very far from zero. But irrespective of epistemic likelihoods and the empirical fact of the matter, an affective account of suggested hallucination also has important implications for the general structure of the cognitive-penetrability debate.

For example, it is widely taken for granted that cognitive penetration may involve penetration of either perception or perceptual belief. Many defenses of cognitive penetrability of perception accordingly place emphasis on negative arguments against the plausibility of explaining some or another phenomenon by appeal to distorted perceptual belief. The conclusion of this chapter implies that such a dialectic is unwarranted.

Consider relatedly an apparent assumption of many debates that if some experience in addition to or as opposed to low-level perceptual experience is cognitively penetrable, then either high-level perceptual phenomenology or cognitive phenomenology or both exist. The affective interpretation of suggested hallucination underscores that even if both the antecedent and the consequent of this conditional are true, the conditional as such is still false. For at least in some relevant cases, it is plausibly affective phenomenology rather than perceptual or cognitive phenomenology that is cognitively or otherwise penetrated.

4. Suggested Hallucination

Dokic and Martin (2015) argue accordingly that some apparent cases of cognitive penetration are most plausibly explained by alterations in affective phenomenology. They focus on much-discussed candidates of diachronic cognitive penetration in particular, such as that, allegedly, yellow banana shapes are perceived as more yellow than similarly colored but non-yellow-associated shapes, as a function of cognitively mediated perceptual learning about the typical color of bananas. Similarly to the presently defended view, the authors contend that subjects do not really perceive bananas as more yellow: it only feels to them as if this were so.

The contribution of this chapter relative to Dokic and Martin, then, is a defense of an application of an affective interpretation to candidates of synchronic cognitive penetration. In this respect, their thesis and the present theses are not only compatible but also complementary. They are also complementary insofar as suggested hallucination is a less known but ultimately more powerful candidate of cognitive penetration than the standard diachronic candidates.

Yet perhaps an affective interpretation of diachronic candidates of cognitive penetration is in tension with the intra-perceptual account of the ambiguous-race face lightness illusion proposed in *Chapter 3*? After all, we granted that the illusion may involve genuine perceptual distortion. By extrapolation, one might expect that we would also grant for example that bananas are perceived as more yellow than similarly colored non-yellow-associated objects.

However, there may be a way to have our cake and eat it too. Recall that the stimuli used in the face–race lightness illusion experiments may have been artefacts. For example, as a result of controlling for average luminance, the white of the eyes of the black face were perceptibly lighter than the white of the eyes of the white face. So perhaps an intra-perceptual account of the illusion still holds due to such mishaps in the stimuli, whereas an affective account better explains other candidates of cognitive penetration.

In any case, as opposed to Dokic and Martin (2015), who tackle the issue of cognitive penetrability head on, but who only focus on diachronic candidates, Woody and Szechtman (2000a, 2000b, 2007) do not discuss cognitive penetrability, but they explicitly propose an affective view of suggested hallucination. The reader may recall from the discussion of the feeling of knowing, though, that the authors appear to defend a two-step causal account of suggested hallucination whereby the second step involves the generation of hallucinatory percepts around feelings of knowing (*Sub-sub-subsection 4.8.2.2.4*).

If the mentioned reading is correct, then the defended view of this chapter significantly departs from that of Woody and Szechtman with respect to suggested positive hallucination. But even if the reading is incorrect, this chapter will still have provided a more in-depth and wide-ranging and up-to-date defense of a non-credulous affective view, meanwhile also explicitly connecting issues of suggested hallucination with issues of cognitive penetrability. So whether the defended theses are rival alternatives or an extension of previously proposed related theses is ultimately immaterial. In either case, they hopefully fruitfully contribute to an ever more interesting debate about cognitive penetrability.

5. General conclusion

Is the way we perceive the world influenced by how we think about or desire the world to be? *Chapter 2* argued that whether this question is posed in a computational–functional framework or in relation to conscious experience, the answer has implications for both mental architecture and issues relating for example to the justificatory role of perception and the theory-ladenness of scientific observation. The chapter also argued that as of the present state of the debate, the most plausible candidates of cognitive penetration are imaginatively mediated.

Chapter 3 accordingly addressed a most widely discussed candidate of synchronic cognitive penetration that would seem to involve distortions in lightness perception of faces as a function of beliefs about or concepts relating to the typical skin tone of different races. The chapter challenged a prima facie powerful argument to this effect, and proposed a low-level, intra-perceptual account of the phenomena instead.

Chapter 4 then raised the empirical, theoretical, and dialectic bar as high as possible, by introducing and defending suggested hallucination as a psychologically real phenomenon. However, it was argued, no credulous or cognitive-delusion view is as plausible as an affective construal of the effect. On the latter view, subjects under suggested hallucination do not experience real perceptual hallucinations. Nor need they undergo genuine cognitive delusions. What more plausibly explains their responding is a distortion in the affective dimension of pain in the case of suggested analgesia, and feelings of conviction or knowing or reality in more standard modes of suggested hallucination such as suggested visual and auditory hallucination.

The verdict, then, is that perception is unlikely cognitively penetrable. Or at least it is unlikely cognitively penetrable by imagery, which is at least a prima facie strongest candidate intermediary of alleged penetration effects. This is the main conclusion of the dissertation.

5. General conclusion

Two auxiliary conclusions have important implications for philosophy and psychology irrespective of cognitive impenetrability. The first is that intra-perceptual states and processes may be richer and hence theoretically more consequential than is generally assumed. The second is that affective phenomenology may be more integral to both perceptual experience and perceptual belief or judgment than is typically reckoned. Since the latter implies that the perception–cognition divide is a false dichotomy, it warrants two final points.

An interesting question is whether perceptual belief or judgment may be intentionally withheld in certain cases. Some retrospective first-person accounts of suggested hallucination point in this direction. For example, a subject reflected after an experiment of suggested analgesia that "When you're hypnotized, there are certain questions that just aren't answered, and you just don't probe them in your mind. I think you're aware that the pain exists, but it's not appropriate to deal with it just then" (Hilgard, 1986, p. 275). Another noted in relation to suggested blindness that she "did not register" visual percepts because she believed "it was not important to notice them" (Cox & Bryant, 2008, p. 321). Even if "hidden observers" do register relevant perceptual information, their experiences still seem similar insofar as "The hidden part doesn't deal [e.g.] with pain. It looks at what is, and doesn't judge it" (Hilgard, 1986, p. 209).

Finally, skeptics of cognitive penetrability who argue that many alleged penetration effects are carried by perceptual misjudgments, response bias, or demand compliance are often taken to task to explain why psychologically healthy and honest experimental subjects would commit so gross errors in judgment or response. A possible reply is that of course most subjects are neither misjudging the perceptual state of affairs, nor merely complying with demands. However, the responses of a relatively significant portion of subjects may still reflect distorted feelings of conviction or knowing or reality. As long as this possibility is not taken into account, no final word can be given on cognitive penetrability of perception, either.

References

Abelson, R. P. (1988). Conviction. *American Psychologist*, *43*(4), 267–275. https://doi.org/10.1037/0003-066x.43.4.267

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science*, *11*(2), 118–124. https://doi.org/10.1111/1467-9280.00226

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press. Retrieved from https://www.semanticscholar.org

Alexander, H. B. (1904). Some observations on visual imagery. *Psychological Review*, *11*(4–5), 319–337. https://doi.org/10.1037/h0069793

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

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

Arango-Muñoz, S. (2014a). Metacognitive feelings, self-ascriptions and mental actions. *Philosophical Inquiries*, *2*(1), 145–162. Retrieved from https://philinq.it

Arango-Muñoz, S. (2014b). The nature of epistemic feelings. *Philosophical Psychology*, *27*(2), 193–211. https://doi.org/10.1080/09515089.2012.732002

Arango-Muñoz, S., & Michaelian, K. (2014). Epistemic feelings, epistemic emotions: Review and introduction to the focus section. *Philosophical Inquiries*, *2*(1), 97–122. Retrieved from https://philinq.it

References

Arstila, V. (2016). Perceptual learning explains two candidates for cognitive penetration. *Erkenntnis*, *81*(6), 1151–1172. https://doi.org/10.1007/s10670-015-9785-3

Arstila, V. (2017). Cognitive penetration, hypnosis and imagination. *Analysis*, *77*(1), 3–10. https://doi.org/10.1093/analys/anx048

Azzopardi, P., & Cowey, A. (1998). Blindsight and visual awareness. *Consciousness and Cognition*, *7*(3), 292–311. https://doi.org/10.1006/ccog.1998.0358

Bagehot, W. (1915). On the emotion of conviction. In R. Barrington (Ed.), *The works and life of Walter Bagehot: The works in nine volumes. The life in one volume* (Vol. 5, pp. 99–111). Retrieved from https://archive.org (Original work published 1871)

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. https://doi.org/10.3758/s13423-016-1048-z

Banyai, E. I., & Hilgard, E. R. (1976). A comparison of active-alert hypnotic induction with traditional relaxation induction. *Journal of Abnormal Psychology*, *85*(2), 218–224. https://doi.org/10.1037/0021-843x.85.2.218

Barabasz, A. F., & Barabasz, M. (2008). Hypnosis and the brain. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 337–363). https://doi.org/10.1093/oxfordhb/9780198570097.013.0013

Barabasz, A. F., & Barabasz, M. (2015). The new APA definition of hypnosis: Spontaneous hypnosis MIA. *American Journal of Clinical Hypnosis*, *57*(4), 459–463. https://doi.org/10.1080/00029157.2015.1011507

Barabasz, A. F., & Lonsdale, C. (1983). Effects of hypnosis on P300 olfactory-evoked potential amplitudes. *Journal of Abnormal Psychology*, *92*(4), 520–523. https://doi.org/10.1037/0021-843x.92.4.520

References

Barber, T. X. (1959a). The afterimages of "hallucinated" and "imagined" colors. *Journal of Abnormal and Social Psychology*, *59*(1), 136–139. https://doi.org/10.1037/h0047334

Barber, T. X. (1959b). Toward a theory of pain: Relief of chronic pain by prefrontal leucotomy, opiates, placebos, and hypnosis. *Psychological Bulletin*, *56*(6), 430–460. https://doi.org/10.1037/h0046676

Barber, T. X. (1964). Hypnotically hallucinated colors and their negative after-images. *The American Journal of Psychology*, *77*(2), 313–318. https://doi.org/10.2307/1420146

Barber, T. X. (1965). Physiological effects of "hypnotic suggestions": A critical review of recent research (1960-64). *Psychological Bulletin*, *63*(4), 201–222. https://doi.org/10.1037/h0021931

Barber, T. X. (1970). Hypnosis, suggestions, and auditory-visual "hallucinations": A critical analysis. In W. Keup (Ed.), *Origin and mechanisms of hallucinations: Proceedings of the 14th Annual Meeting of the Eastern Psychiatric Research Association held in New York City, November 14–15, 1969* (pp. 167–180). https://doi.org/10.1007/978-1-4615-8645-6_15

Barber, T. X. (2009). The necessary and sufficient conditions for hypnotic behavior. *American Journal of Clinical Hypnosis*, *51*(4), 363–375. Retrieved from http://www.tandfonline.com/toc/ujhy20/current (Reprinted from *American Journal of Clinical Hypnosis*, 1960, *3*(1), 31–42)

Barber, T. X. (2017). Suggested ("hypnotic") behavior: The trance paradigm versus an alternative paradigm. In E. Fromm & R. E. Shor (Eds.), *Hypnosis: Research developments and perspectives* (Ch. 5). Retrieved from https://books.google.com (Original work published 1972)

Barnier, A. J., Dienes, Z., & Mitchell, C. J. (2008). How hypnosis happens: New cognitive theories of hypnotic responding. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford*

handbook of hypnosis: Theory, research, and practice (pp. 141–177). https://doi.org/10.1093/oxfordhb/9780198570097.013.0006

Barnier, A. J., & Nash, M. R. (2008). Introduction: A roadmap for explanation, a working definition. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 311–336). https://doi.org/10.1093/oxfordhb/9780198570097.001.0001

Bayne, T. (2007). Hypnosis and the unity of consciousness. In G. A. Jamieson (Ed.), *Hypnosis and conscious states: The neuroscience perspective* (pp. 93–109). New York: Oxford University Press.

Bayne, T. (2010). Hypnosis. In *The unity of consciousness* (pp. 173–188). https://doi.org/10.1093/acprof:oso/9780199215386.003.0008

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 640–645. https://doi.org/10.1037/0278-7393.13.4.640

Binet, A. (1890). *On double consciousness: Experimental psychological studies.* Retrieved from https://archive.org

Binet, A., & Féré, C. (1888). *Animal magnetism*. Retrieved from https://archive.org

Bitter, D. (2010). *The dissociation of consciousness* (Master's thesis). Retrieved from https://library.ceu.edu/ceu-library/electronic-theses-and-dissertations-etds/

Bitter, D. (2013a, March). *Distortions in the perceived lightness of faces: A reply to Macpherson.* Paper presented at the 9th International Symposium of Cognition, Logic and Communication: Perception and Concepts, Riga, Latvia.

References

Bitter, D. (2013b, June). *Distortions in the perceived lightness of faces: A defense of cognitive impenetrability*. Paper swap at the 39th Annual Meeting of the Society for Philosophy and Psychology, Providence, Rhode Island.

Bitter, D. (2013c, July). *Distortions in the perceived lightness of faces: A defense of cognitive impenetrability*. Paper presented at the 21st Annual Meeting of the European Society for Philosophy and Psychology, Granada, Spain.

Bitter, D. (2014). Is low-level visual experience cognitively impenetrable? A critical analysis of some of the purported best evidence. *Baltic International Yearbook of Cognition, Logic and Communication*, *9*. https://doi.org/10.4148/1944-3676.1082

Bitter, D. (2015a, March). *Hypnotic hallucination: A best case for cognitive penetration?* Poster session presented at the Interdisciplinary Conference on Predictive Coding and Cognitive Penetration, Bochum, Germany.

Bitter, D. (2015b, July). *Hypnotic hallucination: A best case for cognitive penetration?* Paper presented at the 23rd Annual Meeting of the European Society for Philosophy and Psychology, Tartu, Estonia.

Bitter, D. (2017, January). *Two dogmas of the cognitive-penetrability debate*. Paper presented at the CEU-Rutgers MIND Workshop, Budapest, Hungary.

Bitter, D., & Farkas, K. (2012, March). *Perceptual learning and cognitive penetration of perceptual experience*. Paper presented at the Perceptual Learning and Perceptual Recognition Workshop, York, UK.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/s0140525x00038188

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5–6), 481–499. https://doi.org/10.1017/s0140525x07002786

References

Block, N. (2012). The grain of vision and the grain of attention. *Thought: A Journal of Philosophy*, *1*(3), 170–184. https://doi.org/10.1002/tht3.28

Block, N. (2016). Tweaking the concepts of perception and cognition [Peer commentary on the paper "Cognition does not affect perception: Evaluating the evidence for 'top-down' effects," by C. Firestone & B. J. Scholl]. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/s0140525x15002733

Braithwaite, R. B. (1933). VI.—The nature of believing. *Proceedings of the Aristotelian Society*, *33*(1), 129–146. https://doi.org/10.1093/aristotelian/33.1.129

Briscoe, R. (2015). Cognitive penetration and the reach of phenomenal content. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 174–199). https://doi.org/10.1093/acprof:oso/9780198738916.003.0008

Brogaard, B. (2014). Seeing as a non-experiential mental state: The case from synesthesia and visual imagery. In R. Brown (Ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience* (pp. 377–394). https://doi.org/10.1007/978-94-007-6001-1_26

Brogaard, B., & Gatzia, D. E. (2017a). Cortical color and the cognitive sciences. *Topics in Cognitive Science*, *9*(1), 135–150. https://doi.org/10.1111/tops.12241

Brogaard, B., & Gatzia, D. E. (2017b). Is color experience cognitively penetrable? *Topics in Cognitive Science*, *9*(1), 193–214. https://doi.org/10.1111/tops.12221

Brössel, P. (2017). Rational relations between perception and belief: The case of color. *Review of Philosophy and Psychology*, *8*(4), 721–741. https://doi.org/10.1007/s13164-017-0359-y

Brown, A. S. (2003). A review of the déjà vu experience. *Psychological Bulletin*, *129*(3), 394–413. https://doi.org/10.1037/0033-2909.129.3.394

References

Brown, A. S. (2005). *The déjà vu experience*. Retrieved from https://books.google.com (Original work published 2004)

Brown, J. W. (1996). *Time, will, and mental process*. Retrieved from https://books.google.com

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, *42*(1), 33–44. https://doi.org/10.1037/h0058484

Bruno, N., Sachs, N., Demily, C., Franck, N., & Pacherie, E. (2012). Delusions and metacognition in patients with schizophrenia. *Cognitive Neuropsychiatry*, *17*(1), 1–18. https://doi.org/10.1080/13546805.2011.562071

Bryant, R.A., & McConkey, K.M. (1989). Hypnotic blindness, awareness, and attribution. *Journal of Abnormal Psychology*, *98*(4), 443–447. https://doi.org/10.1037//0021-843x.98.4.443

Burnston, D. C. (2017). Cognitive penetration and the cognition–perception interface. *Synthese*, *194*(9), 3645–3668. https://doi.org/10.1007/s11229-016-1116-y

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222. https://doi.org/10.1016/s1364-6613(00)01483-2

Campion, J., Latto, R., & Smith, Y. M. (1983). Structure, function, and consciousness in residual vision and blindsight [Response to peer commentary on the paper "Is blindsight an effect of scattered light, spared cortex, and near-threshold vision?"]. *Behavioral and Brain Sciences*, *6*(3), 469–486. https://doi.org/10.1017/s0140525x0001709x

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

Carruthers, P. (2017). Are epistemic emotions metacognitive? *Philosophical Psychology*, *30*(1–2), 58–78. https://doi.org/10.1080/09515089.2016.1262536

References

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science*, *55*(2), 167–187. https://doi.org/10.1086/289425

Coe, W. C., & Sarbin, T. R. (1977). Hypnosis from the standpoint of a contextualist. *Annals of the New York Academy of Sciences*, *296*(1), 2–13. https://doi.org/10.1111/j.1749-6632.1977.tb38157.x

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

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295x.82.6.407

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, *21*(4), 843–860. https://doi.org/10.3758/s13423-013-0564-3

Coltheart, M. (2006). Perhaps functional neuroimaging has not told us anything about the mind (so far). *Cortex*, *42*(3), 422–427. https://doi.org/10.1016/S0010-9452(08)70374-5

Coltheart, M. (2010). Lessons from cognitive neuropsychology for cognitive science: A reply to Patterson and Plaut (2009). *Topics in Cognitive Science*, *2*(1), 3–11. https://doi.org//10.1111/j.1756-8765.2009.01067.x

Coltheart, M. (2013). How can functional neuroimaging inform cognitive theories? *Perspectives on Psychological Science*, *8*(1), 98–103. https://doi.org/10.1177/1745691612469208

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331–335. https://doi.org/10.3758/bf03196169

References

Cox, R. E., & Bryant, R. A. (2008). Advances in hypnosis research: Methods, designs and contributions of intrinsic and instrumental hypnosis. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 311–336). https://doi.org/10.1093/oxfordhb/9780198570097.013.0012

Crawford, H. J. (2001). Neuropsychology of hypnosis: Toward an understanding of how hypnotic interventions work. In G. D. Burrows, R. O. Stanley, & P. B. Bloom (Eds.), *International handbook of clinical hypnosis* (pp. 61–82). https://doi.org/10.1002/0470846402.ch5

Crawford, H. J., Macdonald, H., & Hilgard, E. R. (1979). Hypnotic deafness: A psychophysical study of responses to tone intensity as modified by hypnosis. *The American Journal of Psychology*, *92*(2), 193–214. https://doi.org/10.2307/1421920

Csikszentmihalyi, M. (2008). *Flow: The psychology of optimal experience* (1st Harper Perennial Modern Classics ed.). New York: Harper Perennial Modern Classics. (Original work published 1990)

Cyna, A. M., & Lang, E. V. (2011). How words hurt. In A. M. Cyna, M. I. Andrew, S. G. M. Tan, & A. F. Smith (Eds.), *Handbook of communication in anaesthesia & critical care: A practical guide to exploring the art* (pp. 30–37). Retrieved from https://books.google.com

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Avon Books.

de Fockert, J. W. (2001). The role of working memory in visual selective attention. *Science*, *291*(5509), 1803–1806. https://doi.org/10.1126/science.1056496

Delk, J. L., & Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *The American Journal of Psychology*, *78*(2), 290–293. https://doi.org/10.2307/1420503

References

Dennett, D. (1976). Are dreams experiences? *The Philosophical Review*, *85*(2), 151–171. https://doi.org/10.2307/2183728

Deroy, O. (2013). Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies*, *162*(1), 87–107. https://doi.org/10.1007/s11098-012-9989-1

de Sousa, R. (2008). Epistemic feelings. In G. Brun, U. Doguoglu, & D. Kuenzle (Eds.), *Epistemology and emotions* (pp. 185–204). Aldershot, UK: Ashgate.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 51–60. https://doi.org/10.1037/h0042712

Dienes, Z., & Perner, J. (2007). Executive control without conscious awareness: The cold control theory of hypnosis. In G. A. Jamieson (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective* (pp. 293–314). New York: Oxford University Press.

Dokic, J. (2012). Seeds of self-knowledge: Noetic feelings and metacognition. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 302–321). https://doi.org/10.1093/acprof:oso/9780199646739.003.0020

Dokic, J., & Martin, J.-R. (2012). Disjunctivism, hallucinations, and metacognition. *WIREs Cognitive Science*, *3*, 533–543. https://doi.org/10.1002/wcs.1190

Dokic, J., & Martin, J.-R. (2015). Looks the same but feels different: A metacognitive approach to cognitive penetrability. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 241–267). https://doi.org/10.1093/acprof:oso/9780198738916.003.001

Dokic, J., & Martin, J.-R. (2017). Felt reality and the opacity of perception. *Topoi*, *36*(2), 299–309. https://doi.org/10.1007/s11245-015-9327-2

Downey, J. E. (1901). An experiment on getting an after-image from a mental image. *Psychological Review*, *8*(1), 42–55. https://doi.org/10.1037/h0074691

References

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Elkins, G. R., Barabasz, A. F., Council, J. R., & Spiegel, D. (2015). Advancing research and practice: The revised APA Division 30 definition of hypnosis. *International Journal of Clinical and Experimental Hypnosis*, *63*(1), 1–9. https://doi.org/10.1080/00207144.2014.961870

Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *The British Journal of Psychiatry*, *157*(2), 239–248. https://doi/10.1192/bjp.157.2.239

Elsea, O. C. (1961). *A study of the effect of hypnotic suggestion on color perception* (Doctoral dissertation). Retrieved from https://shareok.org

Erickson, E. M. (1941). Critical comments on Hibler's presentation of his work on negative after-images of hypnotically induced hallucinated colors. *Journal of Experimental Psychology*, *29*(2), 164–170. https://doi.org/10.1037/h0060698

Erickson, M. H., & Erickson, E. M. (1938). The hypnotic induction of hallucinatory color vision followed by pseudo-negative after-images. *Journal of Experimental Psychology*, *22*(6), 581–588. https://doi.org/10.1037/h0060512

Fazekas, P., & Nanay, B. (2017). Pre-cueing effects: Attention or mental imagery? *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00222

Fendrich, R., Wessinger, C. M., & Gazzaniga, M. S. (1992). Residual vision in a scotoma: Implications for blindsight. *Science*, *258*(5087), 1489–1491. https://doi.org/10.1126/science.1439839

Fendrich, R., Wessinger, C. M., & Gazzaniga, M. S. (1993). [Response to peer commentary on the paper "Residual vision in a scotoma: Implications for blindsight"]. *Science*, *261*(5120), 494–495. https://doi.org/10.1126/science.261.5120.494-a

References

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. https://doi.org/10.1177/0956797613485092

Firestone, C., & Scholl, B. J. (2015a). Can you *experience* "top-down" effects on perception? The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*, *22*(3), 694–700. https://doi.org/10.3758/s13423-014-0711-5

Firestone, C., & Scholl, B. J. (2015b). Enhanced visual awareness for morality and pajamas? Perception vs. memory in "top-down" effects. *Cognition*, *136*, 409–416. https://doi.org/10.1016/j.cognition.2014.10.014

Firestone, C., & Scholl, B. J. (2015c). When do ratings implicate perception versus judgment? The "overgeneralization test" for top-down effects. *Visual Cognition*, *23*(9–10), 1217–1226. https://doi.org/10.1080/13506285.2016.1160171

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

Firestone, C., & Scholl, B. J. (2016b). Seeing and thinking: Foundational issues and empirical horizons [Response to peer commentaries on the paper "Cognition does not affect perception: Evaluating the evidence for 'top-down' effects"]. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/s0140525x16000029

Firestone, C., & Scholl, B. J. (2017). Seeing and thinking in studies of embodied "perception": How (not) to integrate vision science and social psychology. *Perspectives on Psychological Science*, *12*(2), 341–343. https://doi.org/10.1177/1745691616679944

Fleminger, S. (1991). Déjà vu phenomena [Peer commentary on the paper "The déjà vu experience: Remembrance of things past?" by H. N. Sno & D. H. Linszen]. *American Journal of Psychiatry*, *148*(10), 1418–1419. https://doi.org/10.1176/ajp.148.10.1418-b

References

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, *28*(2), 97–115. https://doi.org/10.1007/bf00485230

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT press.

Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, *51*(1), 23–43. https://doi.org/10.1086/289162

Fodor, J. A. (1985). Précis of *The modularity of mind*. *Behavioral and Brain Sciences*, *8*(1), 1–5. https://doi.org/10.1017/s0140525x0001921x

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1988). A reply to Churchland's "Perceptual plasticity and theoretical neutrality". *Philosophy of Science*, *55*(2), 188–198. https://doi.org/10.1086/289426

Fodor, J. A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford: Calendron Press.

Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's "ecological approach". *Cognition*, *9*(2), 139–196. https://doi.org/10.1016/0010-0277(81)90009-3

Franklin, B., Majault, Le Roy, A., Sallin, Bailly, D'Arcet, . . . Lavoisier (1785). *Report of Dr. Benjamin Franklin, and other commissioners, charged by the King of France, with the examination of the animal magnetism as now practised at Paris: Translated from the French: With an historical introduction* (Unknown, Trans.). Retrieved from https://archive.org (Original work published 1784)

References

Freud, S. (1953a). The interpretation of dreams. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vols. 4–5, pp. 1–622). London: Hogarth Press. (Original work published 1900)

Freud, S. (1953b). On dreams. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 5, pp. 633–686). London: Hogarth Press. (Original work published 1901)

Freud, S. (1960). The psychopathology of everyday life. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 6). London: Hogarth Press. (Original work published 1904)

Freud, S. (1963). Introductory lectures on psycho-analysis: Part II: Dreams. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 15, pp. 83–239). London: Institute of Psycho-analysis. (Original work published 1916)

Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research*, *46*(23), 4032–4047. https://doi.org/10.1016/j.visres.2006.07.014

Gatzia, D. E. (2017). Cognitive penetration and memory colour effects. *Erkenntnis*. Advance online publication. https://doi.org/10.1007/s10670-017-9951-x

Gendler, T. S. (2008a). Alief and belief. *The Journal of Philosophy*, *105*(10), 634–663. https://doi.org/10.5840/jphil20081051025

Gendler, T. S. (2008b). Alief in action (and reaction). *Mind & Language*, *23*(5), 552–585. https://doi.org/10.1111/j.1468-0017.2008.00352.x

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., . . . Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*(4), 795–834. https://doi.org/10.1037/0033-295X.106.4.795

References

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

Goleman, D. (1987, March 31). Hypnosis still provokes some skeptics. *The New York Times*. Retrieved from https://www.nytimes.com

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grahek, N. (2007). *Feeling pain and being in pain* (2nd ed.). Cambridge, MA: MIT Press. (Original work published 2001)

Green, J. P., Barabasz, A. F., Barrett, D., & Montgomery, G. H. (2005). Forging ahead: The 2003 APA Division 30 definition of hypnosis. *International Journal of Clinical and Experimental Hypnosis*, *53*(3), 259–264. https://doi.org/10.1080/00207140590961321

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. *American Psychologist*, *47*(6), 766–779. https://doi.org/10.1037/0003-066x.47.6.766

Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00221

Gross, S., Chaisilprungraung, T., Kaplan, E., Menendez, J. A., & Flombaum, J. (2014). Problems for the purported cognitive penetration of perceptual color experience and Macpherson's proposed mechanism. *Baltic International Yearbook of Cognition, Logic and Communication*, *9*. https://doi.org/10.4148/1944-3676.1085

Grubin, D. (2010). The polygraph and forensic psychiatry [Editorial]. *The Journal of the American Academy of Psychiatry and the Law*, *38*(4), 446–451. Retrieved from https://jaapl.org

Gruzelier, J. (1998). A working model of the neurophysiology of hypnosis: A review of evidence. *Contemporary Hypnosis*, *15*(1), 3–21. https://doi.org/10.1002/ch.112

References

Gruzelier, J. (2005). Altered states of consciousness and hypnosis in the twenty-first century. *Contemporary Hypnosis*, *22*(1), 1–7. https://doi.org/10.1002/ch.14

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

Harnad, S. (1986). [Editorial commentary on the paper "Hypnotic behavior: A social-psychological interpretation of amnesia, analgesia, and 'trance logic,'" by N. P. Spanos]. *Behavioral and Brain Sciences*, *9*(3), 488–489.

Harris, C. S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review*, *72*(6), 419–444. https://doi.org/10.1037/h0022616

Helton, G. (2016). Recent issues in high-level perception. *Philosophy Compass*, *11*(12), 851–862. https://doi.org/10.1111/phc3.12383

Hilgard, E. R. (1965). *Hypnotic susceptibility*. New York: Harcourt, Brace & World.

Hilgard, E. R. (1973). The domain of hypnosis: With some comments on alternative paradigms. *American Psychologist*, *28*(11), 972–982. https://doi.org/10.1037/h0035452

Hilgard, E. R. (1977). The problem of divided consciousness: A neodissociation interpretation. *Annals of the New York Academy of Sciences*, *296*(1), 48–59. https://doi.org/10.1111/j.1749-6632.1977.tb38160.x

Hilgard, E. R. (1986) *Divided consciousness: Multiple controls in human thought and action* (Exp. ed.). New York: John Wiley & Sons.

Hilgard, E. R. (1991). A neodissociation interpretation of hypnosis. In S. J. Lynn & J. W. Rhue (Eds.), *Theories of hypnosis: Current models and perspectives* (pp. 83–104). New York: Guilford Press.

Hilgard, E. R. (1992). Dissociation and theories of hypnosis. In E. Fromm & M. R. Nash (Eds.), *Contemporary hypnosis research* (pp. 69–101). New York: Guilford Press.

References

Hilgard, E. R., & Hilgard, J. R. (1994). *Hypnosis in the relief of pain* (Rev. ed.). New York: Brunner-Routledge.

Hilgard, E. R., Hilgard, J. R., Macdonald, H., Morgan, A. H., & Johnson, L. S. (1978). Covert pain in hypnotic analgesia: Its reality as tested by the real-simulator design. *Journal of Abnormal Psychology*, *87*(6), 655–663. https://doi.org/10.1037/0021-843x.87.6.655

Hilgard, E. R., Morgan, A. H., & Macdonald, H. (1975). Pain and dissociation in the cold pressor test: A study of hypnotic analgesia with "hidden reports" through automatic key pressing and automatic talking. *Journal of Abnormal Psychology*, *84*(3), 280–289. https://doi.org/10.1037/h0076654

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Science*, *9*(1), 1–23. https://doi.org/10.1017/s0140525x00021269

Horton, J. E., & Crawford, H. J. (2004). *Hypnosis and suggestibility: An experimental approach*. Clark H. Hull. Carmarthen, Wales: Crown House Publishing. 2002. (Originally published 1933) [Book review]. *American Journal of Clinical Hypnosis*, *46*(3), 258–261. https://doi.org/10.1080/00029157.2004.10403608

Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social and Personality Psychology Compass*, *2*(2), 1052–1072. https://doi.org/10.1111/j.1751-9004.2008.00090.x

Hugenberg, K., Young, S. G., Sacco, D. F., & Bernstein, M. J. (2011). Social categorization influences face perception and face memory. In A. J. Calder, G. Rhodes, M. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 245–261). Retrieved from https://books.google.com

James, W. (1890). *The principles of psychology* (Vols. 1–2). Retrieved from https://archive.org

References

Jastrow, J. (1917). The psychology of conviction. *The Scientific Monthly*, *5*(6), 523–544. Retrieved from https://www.jstor.org/journal/sciemont

Jensen, S. M., Barabasz, A., Barabasz, M., & Warner, D. (2001). EEG P300 event-related markers of hypnosis. *American Journal of Clinical Hypnosis*, *44*(2), 127–139. https://doi.org/10.1080/00029157.2001.10403468

Johnson, R. F., Maher, B. A., & Barber, T. X. (1972). Artifact in the "essence of hypnosis": An evaluation of trance logic. *Journal of Abnormal Psychology*, *79*(2), 212–220. https://doi.org/10.1037/h0032681

Kallio, S., & Koivisto, M. (2013). Posthypnotic suggestion alters conscious color perception in an automatic manner. *International Journal of Clinical and Experimental Hypnosis*, *61*(4), 371–387. https://doi.org/10.1080/00207144.2013.810446

Kallio, S., & Koivisto, M. (2016). Seeing blue as red: A hypnotic suggestion can alter visual awareness of colors. *International Journal of Clinical and Experimental Hypnosis*, *64*(3), 261–284. https://doi.org/10.1080/00207144.2016.1171088

Kallio, S., & Revonsuo, A. (2003). Hypnotic phenomena and altered states of consciousness: A multilevel framework of description and explanation. *Contemporary Hypnosis*, *20*(3), 111–164. https://doi.org/10.1002/ch.273

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science*, *237*(4821), 1445–1452. https://doi.org/10.1126/science.3629249

Kihlstrom, J. F. (2004). Availability, accessibility, and subliminal perception. *Consciousness and Cognition*, *13*(1), 92–100. https://doi.org/10.1016/j.concog.2003.09.004

Kihlstrom, J. F. (2008). The domain of hypnosis, revisited. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 21–52). https://doi.org/10.1093/oxfordhb/9780198570097.013.0002

References

Kihlstrom, J. F. (2014). Hypnosis and cognition. *Psychology of Consciousness: Theory, Research, and Practice*, *1*(2), 139–152. https://doi.org/10.1037/cns0000014

Kihlstrom, J. F., Barnhardt, T. M., & Tataryn, D. J. (1992). Implicit perception. In R. F. Bornstein & T. S. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 17–54). Retrieved from https://www.ocf.berkeley.edu/~jfkihlstrom/PDFs/1990s/1992/ImpPercep1992.pdf

Kinnunen, T., Zamansky, H. S., & Block, M. L. (1994). Is the hypnotized subject lying? *Journal of Abnormal Psychology*, *103*(2), 184–191. https://doi.org/10.1037/0021-843x.103.2.184

Kirsch, I., Cardeña, E., Derbyshire, S., Dienes, Z., Heap, M., Kallio, S., . . . Walters, V. (2011). Definitions of hypnosis and hypnotizability and their relation to suggestion and suggestibility: A consensus statement. *Contemporary Hypnosis & Integrative Therapy*, *28*(2), 107–115. Retrieved from https://www.researchgate.net/directory/publications

Knobe, J. (2015). Philosophers are doing something different now: Quantitative data. *Cognition*, *135*, 36–38. https://doi.org/10.1016/j.cognition.2014.11.011

Koivisto, M., Kirjanen, S., Revonsuo, A., & Kallio, S. (2013). A preconscious neural mechanism of hypnotically altered colors: A double case study. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0070900

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9*(2), 149–171. https://doi.org/10.1006/ccog.2000.0433

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. https://doi.org/10.1093/acprof:oso/9780195179088.001.0001

References

Kosslyn, S. M., Thompson, W. L., Costantini-Ferrando, M. F., Alpert, N. M., & Spiegel, D. (2000). Hypnotic visual illusion alters color processing in the brain. *The American Journal of Psychiatry*, *157*(8), 1279–1284. https://doi.org/10.1176/appi.ajp.157.8.1279

Kropotov, J. D. (2016). *Functional neuromarkers for psychiatry: Applications for diagnosis and treatment*. Amsterdam: Academic Press.

Kunst-Wilson, W., & Zajonc, R. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*(4430), 557–558. https://doi.org/10.1126/science.7352271

LaBerge, S., & Rheingold, H. (1990). *Exploring the world of lucid dreaming*. New York: Ballantine.

Laurence, J.-R., & Perry, C. (1981). The "hidden observer" phenomenon in hypnosis: Some additional findings. *Journal of Abnormal Psychology*, *90*(4), 334–344. https://doi.org/10.1037/0021-843x.90.4.334

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114

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. https://doi.org/10.1037/0096-3445.135.4.501

Lupyan, G. (2015a). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, *6*(4), 547–569. https://doi.org/10.1007/s13164-015-0253-4

Lupyan, G. (2015b). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, *161*, 117–130. https://doi.org/10.1016/j.actpsy.2015.08.006

References

Lynn, S. J., Kirsch, I., & Hallquist, M. N. (2008). Social cognitive theories of hypnosis. In S. J. Lynn & J. W. Rhue (Eds.), *Theories of hypnosis: Current models and perspectives* (pp. 111–139). New York: Guilford Press.

Lyons, J. (2005). Perceptual belief and nonexperiential looks. *Philosophical Perspectives*, *19*(1), 237–256. https://doi.org/10.1111/j.1520-8583.2005.00061.x

Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues*, *21*(1), 289–311. https://doi.org/10.1111/j.1533-6077.2011.00205.x

Lyons, J. C. (2015). Unencapsulated modules and perceptual judgment. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 103–122). https://doi.org/10.1093/acprof:oso/9780198738916.001.0004

Machery, E. (2015). Cognitive penetrability: A no-progress report. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 59–74). https://doi.org/10.1093/acprof:oso/9780198738916.003.0002

Mack, A., & Rock, I. (1999). *Inattentional blindness: An overview* by Arien Mack and Irvin Rock. *Psyche*, *5*(3). Retrieved from http://journalpsyche.org (Reprinted from Chapter 1 of *Inattentional blindness*, pp. 1–26, by A. Mack & I. Rock, 1998, Cambridge, MA: MIT Press)

Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, *84*(1), 24–62. https://doi.org/10.1111/j.1933-1592.2010.00481.x

Macpherson, F. (2015). Cognitive penetration and predictive coding: A commentary on Lupyan [Peer commentary on the paper "Cognitive penetrability of perception in the

age of prediction: Predictive systems are penetrable systems"]. *Review of Philosophy and Psychology*, *6*(4), 571–584. https://doi.org/10.1007/s13164-015-0254-3

Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition*, *47*, 6–16. https://doi.org/10.1016/j.concog.2016.04.001

Malcolm, N. (1959). *Dreaming*. London: Routledge & Kegan Paul.

Mallard, D., & Bryant, R. A. (2001). Hypnotic color blindness and performance on the Stroop test. *International Journal of Clinical and Experimental Hypnosis*, *49*(4), 330–338. https://doi.org/10.1080/00207140108410082

Mandelbaum, E. (2017). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*. Advance online publication. https://doi.org/10.1111/phpr.12368

Manzotti, R. (2017). A perception-based model of complementary afterimages. *SAGE Open*, *7*(1). https://doi.org/10.1177/2158244016682478

Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, (*15*)2, 238–300. https://doi.org/10.1016/0010-0285(83)90010-5

Marcel, A. J. (1986). Consciousness and processing: Choosing and testing a null hypothesis. *Behavioral and Brain Sciences*, *9*(1), 40–41. https://doi.org/10.1017/s0140525x00021439

Marcel, A. J. (1998). Blindsight and shape perception: Deficit of visual consciousness or of visual function? *Brain*, *121*(8), 1565–1588. https://doi.org/10.1093/brain/121.8.1565

Marchi, F., & Newen, A. (2015). Cognitive penetrability and emotion recognition in human facial expressions. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00828

References

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press. (Original work published 1982)

Martin, A. (2009). Circuits in mind: The neural foundations for object concepts. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 1031–1045). Cambridge, MA: MIT Press.

Mazzoni, G., Rotriquenz, E., Carvalho, C., Vannucci, M., Roberts, K., & Kirsch, I. (2009). Suggested visual hallucinations in and out of hypnosis. *Consciousness and Cognition*, *18*(2), 494–499. https://doi.org/10.1016/j.concog.2009.02.002

McConkey, K. M. (2008). Generations and landscapes of hypnosis: Questions we've asked, questions we should ask. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 53–77). https://doi.org/10.1093/oxfordhb/9780198570097.013.0003

McDowell, J. (1996). *Mind and world: With a new introduction* (1st Harvard University Press pbk. ed.). Cambridge, MA: Harvard University Press. (Original work published 1994)

McDowell, J. (2009). Conceptual capacities in perception. *Having the world in view: Essays on Kant, Hegel, and Sellars* (pp. 127–144). Cambridge, MA: Harvard University Press. (Reprinted from *Kreativität: XX. Deutscher Kongreß für Philosophie 26.–30. September 2005 an der Technischen Universität Berlin: Kolloquienbeiträge*, pp. 1065–1079, by G. Abel, Ed., 2006, Hamburg: Felix Meiner Verlag.)

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

References

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meeker, W. B., & Barber, T. X. (1971). Toward an explanation of stage hypnosis. *Journal of Abnormal Psychology*, *77*(1), 61–70. https://doi.org/10.1037/h0030419

Melzack, R. (2001). Pain and the neuromatrix in the brain. *Journal of Dental Education*, *65*(12), 1378–1382. Retrieved from http://www.jdentaled.org

Miller, M. E., & Bowers, K. S. (1993). Hypnotic analgesia: Dissociated experience or dissociated control? *Journal of Abnormal Psychology*, *102*(1), 29–38. https://doi.org/10.1037//0021-843X.102.1.29

Miller, R. J., Hennessy, R. T., & Leibowitz, H. W. (1973). The effect of hypnotic ablation of the background on the magnitude of the ponzo perspective illusion. *International Journal of Clinical and Experimental Hypnosis*, *21*(3), 180–191. https://doi.org/10.1080/00207147308409122

Miller, R. J., Lundy, R. M., & Galbraith, G. G. (1970). Effects of hypnotically induced hallucination of a color filter. *Journal of Abnormal Psychology*, *76*(2), 316–319. https://doi.org/10.1037/h0029785

Miltner, W. H. R., & Weiss, T. (2007). Cortical mechanisms of hypnotic pain control. In G. A. Jamieson (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective* (pp. 51–66). New York: Oxford University Press.

Moene, F. C., & Roelofs, K. (2008). Hypnosis in the treatment of conversion and somatization disorders. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 625–645). https://doi.org/10.1093/oxfordhb/9780198570097.013.0026

References

Mole, C. (2015). Attention and cognitive penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 218–237). https://doi.org/10.1093/acprof:oso/9780198738916.003.0010

Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 339–352. https://doi.org/10.1037/0096-1523.23.2.339

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Nanay, B. (2015). Cognitive penetration and the gallery of indiscernibles. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01527

Neppe, V. M. (2015). An overview perspective on what déjà vu is (part 1). *Journal of Psychology & Clinical Psychiatry*, *2*(6). https://doi.org/10.15406/jpcpy.2015.02.00111

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

Noë, A., & Thompson, E. (1999). Seeing beyond the modules toward the subject of perception. *Behavioral and Brain Sciences*, *22*(3), 386–387. https://doi.org/10.1017/s0140525x99452021

O'Connor, A. R., Barnier, A. J., & Cox, R. E. (2008). Déjà vu in the laboratory: A behavioral and experiential comparison of posthypnotic amnesia and posthypnotic familiarity. *International Journal of Clinical and Experimental Hypnosis*, *56*(4), 425–450. https://doi.org/10.1080/00207140802255450

References

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

Orne, M. T. (1959). The nature of hypnosis: Artifact and essence. *Journal of Abnormal Psychology*, *58*(3), 277–299. https://doi.org/10.1037/h0046128

Orne, M. T. (1962a). Hypnotically induced hallucinations. In L. J. West (Ed.), *Hallucinations* (pp. 211–219). Retrieved from http://www.psych.upenn.edu/history/orne/orne1962inwestbc.html

Orne, M. T. (1962b). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424

Orne, M. T., Dinges, D. F., & Orne, E. C. (1986). Hypnotic experience: A cognitive social psychological reality. *Behavioral and Brain Sciences*, *9*(3), 477–478. https://doi.org/10.1017/s0140525x00046641

Oswald, I. (1957). After-images from retina and brain. *Quarterly Journal of Experimental Psychology*, *9*(2), 88–100. https://doi.org/10.1080/17470215708416225

Pagliaro, L. (1991). Déjà vu phenomena [Peer commentary on the paper "The déjà vu experience: Remembrance of things past?" by H. N. Sno & D. H. Linszen]. *American Journal of Psychiatry*, *148*(10), 1418. https://doi.org/10.1176/ajp.148.10.1418-a

Papanicolaou, A. C. (2017). In search of the mnemonic traces of concepts. In A. C. Papanicolau (Ed.), *The Oxford handbook of functional brain imaging in neuropsychology and cognitive neurosciences* (pp. 231–251). https://doi.org/10.1093/oxfordhb/9780199764228.013.20

Peters, M. A. K., Kentridge, R. W., Phillips, I., & Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix015

Phillips, I. (2013). Afterimages and sensation. *Philosophy and Phenomenological Research*, *87*(2), 417–453. https://doi.org/10.1111/j.1933-1592.2012.00580.x

Phillips, I. (2016). Consciousness and criterion: On Block's case for unconscious seeing. *Philosophy and Phenomenological Research*, *96*(2), 419–451. https://doi.org/10.1111/phpr.12224

Prinz, J. J. (2006). Is the mind really modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22–36). Malden, MA: Blackwell Publishing.

Prinz, J. J. (2011). The sensory basis of cognitive phenomenology. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 174–196). https://doi.org/10.1093/acprof:oso/9780199579938.003.0008

Probst, P., & Jansen, J. (1991). Depersonalization and déjà vu experiences: Prevalences in nonclinical samples. *Zeitschrift Fur Klinische Psychologie, Psychopathologie Und Psychotherapie*, *39*(4), 357–368. Abstract retrieved from https://www.ncbi.nlm.nih.gov/pubmed

Pylyshyn, Z. W. (1979a). The rate of "mental rotation" of images: A test of a holistic analogue hypothesis. *Memory & Cognition*, *7*(1), 19–28. https://doi.org/10.3758/bf03196930

Pylyshyn, Z. W. (1979b). Validating computational models: A critique of Anderson's indeterminacy of representation claim. *Psychological Review*, *86*(4), 383–394. https://doi.org/10.1037/0033-295x.86.4.383

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*(1), 111–132. https://doi.org/10.1017/s0140525x00002053

References

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*(1), 16–45. https://doi.org/10.1037/0033-295x.88.1.16

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a theory for cognitive science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. (1999a). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(3), 341–365. https://doi.org/10.1017/s0140525x99002022

Pylyshyn, Z. (1999b). Vision and cognition: How do they connect? [Response to peer commentaries on the paper "Is vision continuous with cognition? The case for cognitive impenetrability of visual perception"]. *Behavioral and Brain Sciences*, *22*(3), 401–414. https://doi.org/10.1017/s0140525x99622026

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

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

Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cognitive Science*, *25*(3), 423–451. https://doi.org/10.1207/s15516709cog2503_4

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

Raftopoulos, A. (2014a). 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. https://doi.org/10.1080/09515089.2012.729486

Raftopoulos, A. (2014b). Nonconceptual content: A reply to Toribio's "Nonconceptualism and the cognitive impenetrability of early vision". *Philosophical Psychology*, *27*(5), 643–651. https://doi.org/10.1080/09515089.2014.926441

Raftopoulos, A. (2017a). Cognitive penetration lite and nonconceptual content. *Erkenntnis*, *82*(5), 1097–1122. https://doi.org/10.1007/s10670-016-9861-3

Raftopoulos, A. (2017b). Pre-cueing, the epistemic role of early vision, and the cognitive impenetrability of early vision. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01156

Raftopoulos, A., & Zeimbekis, J. (2015). The cognitive penetrability of perception: An overview. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 1–56). https://doi.org/10.1093/acprof:oso/9780198738916.003.0001

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, *277*(5328), 968–971. https://doi.org/10.1126/science.277.5328.968

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*. New York: William Morrow.

Ramachandran, V. S., Rogers-Ramachandran, D., & Cobb, S. (1995). Touching the phantom limb. *Nature*, *377*(6549), 489–490. https://doi.org/10.1038/377489a0

Raz, A., Kirsch, I., Pollard, J., & Nitkin-Kaner, Y. (2006). Suggestion reduces the Stroop effect. *Psychological Science*, *17*(2), 91–95. https://doi.org/10.1111/j.1467-9280.2006.01669.x

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science*, *286*(5449), 2504–2507. https://doi.org/10.1126/science.286.5449.2504

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-x

Rosenthal, B. G., & Mele, H. (1952). The validity of hypnotically induced color hallucinations. *Journal of Abnormal and Social Psychology*, *47*(3), 700–704. https://doi.org/10.1037/h0057046

Rozin, P., Markwith, M., & Ross, B. (1990). The sympathetic magical law of similarity, nominal realism and neglect of negatives in response to negative labels. *Psychological Science*, *1*(6), 383–384. https://doi.org/10.1111/j.1467-9280.1990.tb00246.x

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, *50*(4), 703–712. https://doi.org/10.1037/0022-3514.50.4.703

Russell, B. (1914). Mysticism and logic. *The Hibbert Journal*, *12*, 780–803. Retrieved from https://archive.org

Sackeim, H. A., Nordlie, J. W., & Gur, R. C. (1979). A model of hysterical and hypnotic blindness: Cognition, motivation, and awareness. *Journal of Abnormal Psychology*, *88*(5), 474–489. https://doi.org/10.1037/0021-843x.88.5.474

Sarbin, T. R. (1950). Contributions to role-taking theory: I. Hypnotic behavior. *Psychological Review*, *57*(5), 255–270. https://doi.org/10.1037/h0062218

Sarbin, T. R. (2005). Reflections on some unresolved issues in hypnosis. *International Journal of Clinical and Experimental Hypnosis*, *53*(2), 119–134. https://doi.org/10.1080/00207140590927581

Sheehan, P. W., & Perry, C. W. (2015). *Methodologies of hypnosis: A critical appraisal of contemporary paradigms of hypnosis*. Retrieved from https://books.google.com (Original work published 1976)

Sidis, B. (1898). *The psychology of suggestion: A research into the subconscious nature of man and society*. Retrieved from https://archive.org

Sidis, B. (1906). Are there hypnotic hallucinations? *Psychological Review*, *13*(4), 239–257. https://doi.org/10.1037/h0073552

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

Siegel, S. (2015). Afterword: Epistemic evaluability and perceptual farce. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 405–424). https://doi.org/10.1093/acprof:oso/9780198738916.003.0017

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810. https://doi.org/10.1016/j.neuropsychologia.2007.05.002

Simons, D. J., Hannula, D. E., Warren, D. E., & Day, S. W. (2007). Behavioral, neuroimaging, and neuropsychological approaches to implicit perception. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 207–250). https://doi.org/10.1017/cbo9780511816789.010

Sno, H. N., & Linszen, D. H. (1990). The déjà vu experience: Remembrance of things past? *American Journal of Psychiatry*, *147*(12), 1587–1595. https://doi.org/10.1176/ajp.147.12.1587

Spanos, N. P. (1986). Hypnotic behavior: A social-psychological interpretation of amnesia, analgesia, and "trance logic." *Behavioral and Brain Sciences*, *9*(3), 449–467. https://doi.org/10.1017/s0140525x00046537

Spanos, N. P. (1991). A sociocognitive approach to hypnosis. In S. J. Lynn & J. W. Rhue (Eds.), *Theories of hypnosis: Current models and perspectives* (pp. 324–361). New York: Guilford Press.

Spanos, N. P., & Barber, T. X. (1974). Toward a convergence in hypnosis research. *American Psychologist*, *29*(7), 500–511. https://doi.org/10.1037/h0036795

Spanos, N. P., Bridgeman, M., Stam, H. J., Gwynn, M., & Saad, C. L. (1983). When seeing is not believing: The effects of contextual variables on the reports of hypnotic hallucinators. *Imagination, Cognition and Personality*, *2*(3), 195–209. https://doi.org/10.2190/7rjd-8mfu-8b69-uye8

Spanos, N. P., Burgess, C. A., Cross, P. A., & MacLeod, G. (1992). Hypnosis, reporting bias, and suggested negative hallucinations. *Journal of Abnormal Psychology*, *101*(1), 192–199. https://doi.org/10.1037/0021-843x.101.1.192

Spanos, N. P., Flynn, D. M., & Gabora, N. J. (1989). Suggested negative visual hallucinations in hypnotic subjects: When no means yes. *British Journal of Experimental and Clinical Hypnosis, 6*(2), 63–67.

Spanos, N. P., & Hewitt, E. C. (1980). The hidden observer in hypnotic analgesia: Discovery or experimental creation? *Journal of Personality and Social Psychology*, *39*(6), 1201–1214. https://doi.org/10.1037/h0077730

Spanos, N. P., & Saad, C.-L. (1984). Prism adaptation in hypnotically limb-anesthetized subjects: More disconfirming data. *Perceptual and Motor Skills*, *59*(2), 379–386. https://doi.org/10.2466/pms.1984.59.2.379

Spiegel, D. (1995). Hypnosis, dissociation, and trauma: Hidden and overt observers. In J. L. Singer (Ed.), *Repression and dissociation: Implications for personality theory, psychopathology, and health* (Pbk. ed., pp. 121–142). Retrieved from https://books.google.com (Original work published 1990)

Spiegel, D., Cutcomb, S., Ren, C., & Pribram, K. (1985). Hypnotic hallucination alters evoked potentials. *Journal of Abnormal Psychology*, *94*(3), 249–255. https://doi.org/10.1037/0021-843x.94.3.249

Stevens, F. L., Hurley, R. A., & Taber, K. H. (2011). Anterior cingulate cortex: Unique role in cognition and emotion. *Journal of Neuropsychiatry*, *23*(2), 121–125. https://doi.org/10.1176/appi.neuropsych.23.2.121

Stoelb, B. L., Molton, I. R., Jensen, M. P., & Patterson, D. R. (2009). The efficacy of hypnotic analgesia in adults: A review of the literature. *Contemporary Hypnosis*, *26*(1), 24–39. https://doi.org/10.1002/ch.370

Stoerig, P. (1993). Sources of blindsight. *Science*, *261*(5120), 493–494. https://doi.org/10.1126/science.8380016

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

Stokes, D. (2015). Towards a consequentialist understanding of cognitive penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives* (pp. 75–100). https://doi.org/10.1093/acprof:oso/9780198738916.003.0003

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

References

Sutcliffe, J. P. (1960). “Credulous” and “sceptical” views of hypnotic phenomena: A review of certain evidence and methodology. *International Journal of Clinical and Experimental Hypnosis*, *8*(2), 73–101. https://doi.org/10.1080/00207146008415837

Sutcliffe, J. P. (1961). “Credulous” and “skeptical” views of hypnotic phenomena: Experiments on esthesia, hallucination, and delusion. *Journal of Abnormal and Social Psychology*, *62*(2), 189–200. https://doi.org/10.1037/h0046745

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/10.1016/s0022-5371(79)90355-4

Szechtman, H., Woody, E., Bowers, K. S., Nahmias, C. (1998). Where the imaginal appears real: A positron emission tomography study of auditory hallucinations. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(4), 1956–1960. https://doi.org/10.1073/pnas.95.4.1956

Tellegen, A. (1978). On measures and conceptions of hypnosis. *American Journal of Clinical Hypnosis*, *21*(2–3), 219–237. https://doi.org/10.1080/00029157.1978.10403973

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences (“absorption”), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*(3), 268–277. https://doi.org/10.1037/h0036681

Terhune, D. B., Luke, D. P., & Cohen Kadosh, R. (2017). The induction of synaesthesia in non-synaesthetes. In O. Deroy (Ed.), *Sensory blending: On synaesthesia and related phenomena* (pp. 215–247). https://doi.org/10.1093/oso/9780199688289.003.0012

Teufel, C., & Nanay, B. (2017). How to (and how not to) think about top-down influences on visual perception. *Consciousness and Cognition*, *47*, 17–25. https://doi.org/10.1016/j.concog.2016.05.008

Toribio, J. (2014). Nonconceptualism and the cognitive impenetrability of early vision. *Philosophical Psychology*, *27*(5), 621–642. https://doi.org/10.1080/09515089.2014.893386

Treisman, A. (2009). Attention: Theoretical and psychological perspectives. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 189–204). Cambridge, MA: MIT Press.

Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research*, *45*(9), 1095–1098. https://doi.org/10.1016/j.visres.2004.11.001

Turner, M., & Coltheart, M. (2010). Confabulation and delusion: A common monitoring framework. *Cognitive Neuropsychiatry*, *15*(1-3), 346–376. https://doi.org/10.1080/13546800903441902

Underwood, H. W. (1960). The validity of hypnotically induced visual hallucinations. *Journal of Abnormal and Social Psychology*, *61*(1), 39–46. https://doi.org/10.1037/h0041667

Vaitl, D., Birbaumer, N., Gruzelier, J., Jamieson, G. A., Kotchoubey, B., Kübler, A., . . . Weiss, T. (2005). Psychobiology of altered states of consciousness. *Psychological Bulletin*, *131*(1), 98–127. https://doi.org/10.1037/0033-2909.131.1.98

van de Ven, N., Zeelenberg, M., & Pieters, R. (2011). Why envy outperforms admiration. *Personality and Social Psychology Bulletin*, *37*(6), 784–795. https://doi.org/10.1177/0146167211400421

Varga, S. (2017). Perceptual experience and cognitive penetrability. *European Journal of Philosophy*, *25*(2), 376–397. https://doi.org/10.1111/ejop.12239

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

Vuilleumier, P., & Brosch, T. (2009). Interactions of emotion and attention in perception. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 925–934). Cambridge, MA: MIT Press.

Wagstaff, G. F. (1977). Goal-directed fantasy, the experience of nonvolition, and compliance. *Social Behavior and Personality: An International Journal*, *5*(2), 389–393. https://doi.org/10.2224/sbp.1977.5.2.389

Wagstaff, G. F. (1978). [Review of the book *Methodologies of hypnosis: A critical appraisal of contemporary paradigms of hypnosis*, by D. W. Sheehan & C. W. Perry]. *British Journal of Psychology*, *69*(1), 150. https://doi.org/10.1111/j.2044-8295.1978.tb01641.x

Wagstaff, G. F. (1981). *Hypnosis, compliance, and belief*. New York: St. Martin's Press.

Wagstaff, G. F. (1991). Compliance, belief, and semantics in hypnosis: A nonstate, sociocognitive perspective. In S. J. Lynn & J. W. Rhue (Eds.), *Theories of hypnosis: Current models and perspectives* (pp. 362–396). New York: Guilford Press.

Wagstaff, G. F. (2000). On the physiological redefinition of hypnosis: A reply to Gruzelier. *Contemporary Hypnosis*, *17*(4), 154–162. https://doi.org/10.1002/ch.206

Wagstaff, G. F. (2004). High hypnotizability in a sociocognitive framework. In Heap, M., Brown, R. J., & Oakley, D. A. (Eds.), *The highly hypnotizable person: Theoretical, experimental and clinical issues* (pp. 85–114). Retrieved from https://books.google.com

Wagstaff, G. F. (2014a). On the centrality of the concept of an altered state to definitions of hypnosis. *The Journal of Mind-Body Regulation*, *2*(2), 90–108. Retrieved from https://journalhosting.ucalgary.ca/index.php/mbr

Wagstaff, G. F. (2014b). Making the distinction between the "hypnotic" and "non-hypnotic" [Response to peer commentaries on the paper "On the centrality of the concept of an

altered state to definitions of hypnosis"]. *The Journal of Mind-Body Regulation*, *2*(2), 132–150, Retrieved from https://journalhosting.ucalgary.ca/index.php/mbr

Wagstaff, G., & Cole, J. (2005). Levels of explanation and the concept of a hypnotic state. *Contemporary Hypnosis*, *22*(1), 14–17. https://doi.org/10.1002/ch.16

Wagstaff, G. F., Toner, S., & Cole, J. (2002). Is response expectancy sufficient to account for hypnotic negative hallucinations? *Contemporary Hypnosis*, *19*(3), 133–138. https://doi.org/10.1002/ch.250

Weinberger, J. (2000). William James and the unconscious: Redressing a century-old misunderstanding. *Psychological Science*, *11*(6), 439–445. http://doi.org/10.1111/1467-9280.00286

Weiskrantz, L. (1950). An unusual case of after-imagery following fixation of an "imaginary" visual pattern. *Quarterly Journal of Experimental Psychology*, *2*(4), 170–175. https://doi.org/10.1080/17470215008416594

Weiskrantz, L. (1980). Varieties of residual experience. *Quarterly Journal of Experimental Psychology*, *32*(3), 365–386. https://doi.org/10.1080/14640748008401832

Weiskrantz, L. (1993). Sources of blindsight. *Science*, *261*(5120), 494–495. http://doi.org/10.1126/science.8332916

Weiskrantz, L. (2002). Prime-sight and blindsight. *Consciousness and Cognition*, *11*(4), 568–581. https://doi.org/10.1016/s1053-8100(02)00015-6

Weiskrantz, L., Cowey, A., & Hodinott-Hill, I. (2002). Prime-sight in a blindsight subject. *Nature Neuroscience*, *5*(2), 101–102. https://doi.org/10.1038/nn793

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, *97*(1), 709–728. https://doi.org/10.1093/brain/97.1.709

Weitzenhoffer, A. M. (1974). When is an "instruction" an "instruction"? *International Journal of Clinical and Experimental Hypnosis*, *22*(3), 258–269. https://doi.org/10.1080/00207147408413005

Weitzenhoffer, A. M., & Higard, E. R. (1962). *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto, CA: Consulting Psychologists Press. Retrieved from https://hypnosisandsuggestion.org/suggestibility-scales.html

White, W. R. (1941). A preface to the theory of hypnotism. *Journal of Abnormal and Social Psychology*, *36*(4), 477–505. https://doi.org/10.1037/h0053844

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

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 71–94). Cambridge, MA: MIT Press.

Woody, E. Z., & Barnier, A. J. (2008). Hypnosis scales for the twenty-first century: What do we need and how should we use them. *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 255–281). https://doi.org/10.1093/oxfordhb/9780198570097.013.0010

Woody, E. Z., & Sadler, P. (2008). Dissociation theories of hypnosis. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 81–110), https://doi.org/10.1093/oxfordhb/9780198570097.013.0004

Woody, E., & Szechtman, H. (2000a). Hypnotic hallucinations: Towards a biology of epistemology. *Contemporary Hypnosis*, *17*(1), 4–14. https://doi.org/10.1002/ch.186

Woody, E., & Szechtman, H. (2000b). Hypnotic hallucinations and yedasentience [Response to peer commentaries on the paper "Hypnotic hallucinations: Towards a biology of epistemology"]. *Contemporary Hypnosis*, *17*(1), 26–31. https://doi.org/10.1002/ch.189

References

Woody, E., & Szechtman, H. (2002). The sensation of making sense: Motivational properties of the "fringe" [Peer commentary on the paper "Sensation's ghost: The non-sensory 'fringe' of consciousness," by B. Mangan]. *Psyche*, *8*(20). Retrieved from http://journalpsyche.org

Woody, E., & Szechtman, H. (2007). To see feelingly: Emotion, motivation and hypnosis. In G. A. Jamieson (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective* (pp. 241–256). New York: Oxford University Press.

Wu, W. (2017). Shaking up the mind's ground floor: The cognitive penetration of visual attention. *Journal of Philosophy*, *114*(1), 5–32. https://doi.org/10.5840/jphil201711411

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175. https://doi.org/10.1037/0003-066x.35.2.151

Zamansky, H. S., & Bartis, S. P. (1985). The dissociation of an experience: The hidden observer observed. *Journal of Abnormal Psychology*, *94*(3), 243–248. https://doi.org/10.1037//0021-843x.94.3.243

Zeimbekis, J. (2012). Color and cognitive penetrability. *Philosophical Studies*, *165*(1), 167–175. https://doi.org/10.1007/s11098-012-9928-1