

7

Cognitive Penetration and the Reach of Phenomenal Content

Robert Briscoe

1 Introduction

The phenomenal character of a perceptual experience depends, in part, on its representational content. What it is like to enjoy the experience of seeing a bowl of fruit, for example, is determined by (among other things) the various shapes, sizes, colors, and textures that objects are represented as having in the experience. In keeping with recent usage (see e.g. Kriegel 2002 and Bayne 2009), I shall employ the expression ‘phenomenal content’ to refer to that component of a perceptual state’s representational content that supervenes on its phenomenal character. Features that are represented in a perceptual experience, but that can vary without any change in the experience’s phenomenal character—supposing there be such—are outside the scope of phenomenal content as understood here.¹

Are the phenomenal contents of visual experience informationally encapsulated? Can information originating from outside of the visual system influence the way an object appears to an observer, where ‘appears’ is interpreted in a *phenomenal* as opposed to epistemic sense and where the relevant influence on experience is *direct*, e.g. not mediated by shifts in selective attention? If so, then it is psychologically possible for two observers (or for one observer at different times) to have visual experiences with different phenomenal contents while seeing and attending to the same distal stimuli under the same external conditions as a result of differences in other representational mental states.²

This chapter critically examines two different ways of defending the claim that visual phenomenal content is *not* informationally encapsulated, that it can be ‘penetrated’ by

¹ See Prinz (2006) for an argument that conscious perceptual experiences can have contents that are not reflected in their phenomenal character. A visual experience, e.g., on Prinz’s account can represent an object as having the high-level property of being an *apple* even if that property makes no difference to the way the object (non-epistemically) visually appears.

² This formulation is adapted from Siegel (2012).

nonvisual sources of information about the world or the perceiver's body. I shall refer to this claim, in what follows, as the 'penetrability thesis' (or 'Thesis P' for short).

Some philosophers have recently argued that acquiring the capacity to categorize an object as belonging to a certain high-level kind can alter the way the object subsequently appears. For example, your visual experience of seeing a tiger before acquiring the ability to distinguish tigers from non-tigers, it is claimed, can have a different phenomenal character than your visual experience after acquiring the ability. It is then argued that the best explanation of the putative phenomenal contrast between the two experiences is that the latter, unlike the former, represents the high-level kind property *tiger*. The conclusion that visual phenomenal content can come to include high-level properties—and thus is not restricted to low-level attributes such as shape, size, and color—has been referred to as 'expansionism' about phenomenal content, while its denial has been referred to as 'restrictivism' (Prinz 2012). I shall use this terminology here.

Theorists who pursue one or another variation on this line of argument take what I shall call the 'high road' to Thesis P: they seek to show that visual experience can come to represent high-level properties via some sort of top-down influence on lower-level processing from cognitive systems involved in object recognition. Outputs emanating from systems dedicated to categorizing perceived objects on the basis of the high-level kinds to which they belong can, under certain conditions, expand the stock of attributes that objects may be represented as having in conscious vision and hence, it is claimed, the way they thereafter look (see Siegel 2006: 501).

A rather different approach to defending Thesis P eschews appeal to intuitions about how objects may phenomenally appear to an observer after acquiring certain recognitional dispositions. Further, it does not undertake to show that outputs from object-recognition systems can extend the reach of visual phenomenal content. Rather, it relies on psychophysical and neuroscientific evidence to motivate the claim that information originating outside the visual system can modulate the way an object's low-level attributes visually appear. Accordingly, I shall refer to this second approach as the 'low road' to Thesis P. Unlike theorists who take the high road, theorists who take the low road attempt only to show that information present in areas outside the visual system can directly affect the value that some low-level, representational variable assumes, for example, by changing the weighting assigned to one or another source of optical stimulus information.

This chapter is divided into three parts: Section 2 is dedicated to preliminaries and lays out the distinction between informational encapsulation and cognitive impenetrability. In Section 3, I critically examine three recent high-road arguments for Thesis P deployed respectively by Susanna Siegel (2006; 2010), Tim Bayne (2009), and William Fish (2013). I argue that none of them ultimately proves to be successful. In Section 4, I switch from offense to defense. In particular, I argue—in keeping with the low-road approach—that there is a substantial body of psychophysical and neuroscientific evidence that information originating outside the visual system can modulate

the way an object's low-level attributes visually appear. Visual phenomenal content, I show, is significantly influenced not only by crossmodal interactions between vision and other exteroceptive senses such as touch and audition but also by interactions between vision and nonperceptual systems involved in motor planning and construction of the proprioceptive body image.

2 Some Preliminary Definitions and Remarks

Informational encapsulation is the most important of the nine definitive properties of a mental module enumerated by Jerry Fodor (1983; 2000). According to Fodor, if a perceptual input system such as vision or audition is informationally encapsulated, then the way it processes a set of inputs is computationally insensitive to whatever information might be present elsewhere in other modular or non-modular systems. Cognitive impenetrability, by contrast, only excludes high-level penetration, e.g. penetration by the subject's beliefs, expectations, desire, and other centrally accessible mental states. As Pylyshyn puts it, 'if a system is cognitively penetrable then the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to *what the person knows* (1999: 343; emphasis added). That a perceptual input system is cognitively impenetrable, in this sense, is consistent with informational unencapsulation. Even if processing internal to the system is impervious to high-level influences of various kinds, it may be computationally sensitive to low-level information originating elsewhere in the brain.

Fodor and Pylyshyn have focused primarily on the question of whether the non-conscious, 'early' visual system is informationally encapsulated and/or cognitively impenetrable. The notions of informational encapsulation and cognitive penetrability, however, can be usefully extended to discussions of conscious visual experience (e.g. Macpherson 2012; Siegel 2012; Wu 2013). Visual experience, to a first approximation, is informationally encapsulated if its phenomenal contents are not directly influenced by sources of information emanating from outside of the visual system. And visual experience can be characterized as cognitively impenetrable if its phenomenal contents are not directly influenced by high-level, centrally generated mental states, such as beliefs, desires, and intentions.³

Orthogonal to the distinction between informational encapsulation and cognitive impenetrability—a distinction with respect to putative sources or causes of penetration—is a distinction between different kinds of *effects* on phenomenal content. High-road arguments for Thesis P, as characterized above, maintain that information emanating from object recognitional systems can cause high-level properties

³ A direct causal link of some kind, however, isn't enough. As Wayne Wu (2013) has argued, to establish that visual phenomenal content is penetrated by information originating in some non-visual system *Y*, it is further necessary to provide a mechanism whereby the computations that support visual experience are able to exploit *Y* as a representational resource.

to be represented in visual phenomenal content. For example, acquiring the ability to distinguish tigers from non-tigers might cause the high-level property *tiger* to be represented in one's visual experience of a tiger. On the assumption that the source of penetrating influence here is a properly cognitive one, this would be a case in which a high-level penetrator has a high-level effect on phenomenal content. Alternatively, a high-level, cognitive penetrator might have a low-level effect on phenomenal content, e.g. by modifying the way an object's shape, or orientation, or lightness appears. Correspondingly, a low-level penetrator could, in principle, have either low-level or high-level effects on phenomenal content.

3 High-Road Approaches to Thesis P

In this section, I critically assess three different arguments for the view that acquiring object-recognitional dispositions can cause high-level properties to be represented in the phenomenally conscious contents of visual experience. The first two arguments, developed respectively by Susanna Siegel (2010) and Tim Bayne (2009), employ an essentially phenomenological methodology. The third argument, developed by William Fish (2013), by contrast, draws support from an array of recent psychophysical and neuroscientific findings.

3.1 *The method of phenomenal contrast*

In the method of phenomenal contrast, a putative difference in phenomenology between two overall experiences, OE1 and OE2, is treated as an *explanandum* that different, rival hypotheses about the contents of visual experience compete to explain (Siegel 2010). Siegel brings the method to bear on two cases in which acquired recognitional dispositions seem to affect phenomenology. The first case contrasts one's overall experience when looking at a page of Cyrillic text before learning Russian with one's overall experience of looking at the same page after learning Russian. The second case contrasts one's overall experience when looking for pine trees in a forest before and after learning to discriminate pines from other kinds of trees. After acquiring the relevant recognitional skills, 'you can spot the pine trees immediately: they become visually salient to you' (Siegel 2010: 100).

The argument from these cases to what Siegel calls the 'Rich Content View' has the following structure (let 'VE1' and 'VE2' designate the visual experiences one has when enjoying overall experiences OE1 and OE2, respectively):

- (0) The overall experience OE1 of which VE1 is a part differs phenomenologically from the overall experience OE2 of which VE2 is a part.
- (1) If OE1 differs phenomenologically from OE2, then there is a phenomenological difference between visual experiences VE1 and VE2.
- (2) If there is a phenomenological difference between VE1 and VE2, then VE1 and VE2 differ in phenomenal content.

- (3) If there is a difference in phenomenal content between VE1 and VE2, then it is a difference with respect to the high-level properties represented in VE1 and VE2. For example, the change in how salient or attention-grabbing the pine trees are to you in the second case, Siegel suggests, is to be explained by the presence of the high-level property *pine* in the phenomenal content of VE1 and its absence in the phenomenal content of VE2.

Siegel defends premise (1) against re-descriptions according to which OE1 and OE2 differ in respect of either occurrent non-sensory, cognitive phenomenology or ‘mood-like’ background phenomenology. There are other relevant alternatives in the Cyrillic text case, however. In particular, reading words in a language that one understands often elicits auditory (and sometimes visual) imagery that is absent when one scans a page of text written in an unfamiliar alphabet. Grasping the meaning of what we read can also, in many cases, affect our emotional responses. A sign on a forest trail that reads ВНИМАНИЕ: МЕДВЕДИ ГРИЗЛИ ПОБЛИЗОСТИ! (BEWARE: GRIZZLY BEARS NEARBY!) will elicit quite different feelings from Russian speakers and non-Russian speakers, respectively.

When it comes to the pine-tree example, the most promising skeptical tack, it seems to me, is rather to challenge premise (3), i.e. to accept that there is some sort of difference in phenomenology between VE1 and VE2, but to deny that it is specifically a difference with respect to the high-level kind properties that the two experiences respectively represent.

It is plausible, as Siegel suggests, that an important respect in which the experience of seeing pine trees changes after one learns to recognize them has to do with their *visual salience*. Once you know what a pine tree looks like, the presence of a pine tree in your field of view (at least when you are looking for one) is apparent in a way that it wasn't before. (And plausibly such a change in your experience is *itself* phenomenologically salient: you notice that it is now easier to find pine trees than it was before.) There are two different ways, however, in which visual salience can be conferred on a perceived object. On the one hand, objects sometimes have low-level features in virtue of which they ‘pop out’ from their background and capture the perceiver’s attention in an exogenous, bottom-up manner. A bright yellow triangle, for instance, will be especially noticeable when presented against a background filled with dark, purple discs. Evidently, acquiring the ability to recognize pines does not cause them to be visually salient in this sense.

On the other hand, an object can be visually salient as the result of top-down factors that influence the allocation of overt or covert selective attention. The paradigm here is having a stored, inner representation of some kind—for example, a mental image or concept—that guides or otherwise influences you when searching for the object in a structured, visual scene. Presumably, it is this latter, recognition-based kind of visual salience that is at issue in Siegel’s second example.

In what does recognition-based, visual salience consist? Recent models of top-down or ‘guided’ visual search suggest that high-level knowledge about a target object can

prime or configure the visual system to find it in a scene, in particular, by amplifying the conspicuity of the object's location in what vision researchers refer to as the perceiver's 'saliency map' (Koch and Ullman 1985; Wolfe 1994; Blaser et al. 1999; Wolfe et al. 2004; Underwood et al. 2006). A saliency map is an abstract, topographical representation that records the overall attentional strength of each location in visual space. The greater the attentional strength of a location on the map, the more likely it is to be selected. Evidence gathered in research on top-down search suggests that a location's attentional strength is a function of both pre-attentive, bottom-up feature processing and top-down, cognitive influences of various kinds. For example, stored knowledge about an object's visual appearance might increase the selection-relevant weightings assigned to visual representations of certain low-level features, e.g. certain shapes or colors, if they are distinctive of the object (Wolfe et al. 2004). Alternatively, knowledge about the object's most probable locations in the distal scene might increase the attentional strength of certain regions on the saliency map while decreasing the attentional strength of others. Subjects, for example, are much more likely to fixate locations on a sidewalk or road when looking for bicycles than when looking for chimneys or clouds (Henderson and Hollingworth 1999; Henderson et al. 2007).

For present purposes, the main point is that the increased visual salience or attention-grabbingness of the pine trees in Siegel's second case can be parsimoniously explained without recourse to the hypothesis that acquiring the ability to recognize pines causes a certain high-level kind property to be represented in visual phenomenal content. Rather, in the context of visual search, the visual salience of pines may be amplified in a top-down manner by knowledge about their distinctive, low-level features and/or their statistically most probable locations in the scene. Plausibly, both kinds of knowledge are often acquired in the course of learning to discriminate pine trees.

It should be emphasized in this connection that implicitly categorized stimuli can capture visual attention even when subjects are not explicitly looking for them. Studies of inattention blindness (IB) reported in Mack and Rock (1998) provide evidence that nonconsciously perceived stimuli are often implicitly categorized at the highest levels of perceptual processing. According to Mack and Rock's model of late attentional selection, objects that are implicitly categorized as highly meaningful or task-relevant have the ability to defeat IB and capture the perceiver's attention (1998: chs 8 and 11; also see Mack 2002). Another way, then, in which acquiring the ability to recognize an object can contribute to its visual salience is by making the object capable of defeating IB even when it isn't the target of deliberate visual search. The object will seem especially salient, in other words, because implicit categorization has made it attention-grabbing.

These proposals need to be fleshed out in more detail than is possible here. They hopefully serve to show, however, that the boost in a kind's visual salience after subjects have learned to recognize its members can be parsimoniously explained without

recourse to the expansionist conclusion that the relevant high-level kind property has come to be represented in visual phenomenal content.

3.2 *Associative visual agnosia and the method of phenomenal contrast*

A contrast argument recently developed by Tim Bayne (2009) appeals to changes in phenomenology that are consequent not on the acquisition of recognitional dispositions but rather on their loss in 'pure' associative agnosia. In relevant cases, visually acuity, form discrimination, 3D perceptual organization, and other low-level perceptual abilities remain intact, but patients are incapable of categorizing objects as belonging to certain high-level kinds, for example, as *tigers*, or *teapots*, or *tomatoes*.

Bayne contends that associative agnosia is characterized by a loss of a 'layer' of high-level phenomenal content. The three brief arguments he presents in support of this claim are motivated by a description of deficits in a patient with visual agnosia provided by Rubens and Benson (1971):

For the first three weeks in the hospital the patient could not identify common objects presented visually and did not know what was on his plate until he tasted it. He identified objects immediately on touching them. When shown a stethoscope, he described it as 'a long cord with a round thing at the end', and asked if it could be a watch. He identified a can opener as 'could it be a key?' Asked to name a cigarette lighter, he said, 'I don't know', but named it after the examiner lit it. He said he was 'not sure' when shown a toothbrush. Asked to identify a comb, he said, 'I don't know'. When shown a large matchbook, he said, 'It could be a container for keys'. He correctly identified glasses. For a pipe, he said, 'Some type of utensil, I'm not sure'. Shown a key, he said, 'I don't know what that is; perhaps a file or a tool of some sort'. He was never able to describe or demonstrate the use of an object if he could not name it. If he misnamed an object his demonstration of its use would correspond to the mistaken identification. . . . Remarkably, he could make excellent copies of line drawings and still fail to name the subject. . . . He easily matched drawings of objects that he could not identify, and had no difficulty in discriminating between complex non-representational patterns differing from each other only subtly. He occasionally failed in discriminating because he included imperfections in the paper or in the printer's ink. He could never group drawings by class unless he could first name the subject. (Rubens and Benson 1971: 308–9)

Bayne uses associative visual agnosia to develop a quick contrast argument for expansionism. Although the patient's experience of low-level perceptual attributes is intact, it is nonetheless 'extremely plausible', he suggests, 'to suppose that the phenomenal character of his visual experience has changed' (Bayne 2009: 391). But since the patient, by hypothesis, has not lost low-level phenomenal content, it follows that he must have lost high-level or categorical phenomenal content.

Restrictivists will regard this contrast argument as simply question-begging. There is nothing in the quoted passage from Rubens and Benson (1971) that goes beyond the standard, textbook characterization of associative visual agnosia: viz. the presence of normal abilities to discriminate low-level perceptual attributes in the absence of abilities to recognize (properly categorize) the objects to which those attributes belong.

There is nothing in this characterization, by itself, that underwrites the claim that associative agnosia leads to a ‘disruption to the phenomenal looks of objects’ (p. 394).

Bayne deploys two brief additional arguments in support of the proposal that “*x* looks *F* to *S*” can capture a genuine phenomenal-looks claim even when “*F*” expresses a property that is not sensory in any natural sense of that term’ (2009: 394). Both proceed by defending view that the kind of object recognition that is impaired in visual agnosia falls on the perceptual side of the perception/cognition divide.

The first argument is based on the claim that ‘object recognition of this kind resists doxastic penetration. It does not matter what one believes about an object; it *still looks* like a pipe, a stethoscope, or a cigarette lighter’ (Bayne 2009: 395; emphasis added). This argument, however, begs the question against restrictivism once again. The restrictivist about phenomenal content, as Bayne acknowledges, is quite happy to allow that a thing can look in a non-phenomenal, epistemic, or comparative sense like a pipe or some other high-level kind of object. What the restrictivist will not allow, without a convincing argument, is that a thing can also sometimes look in a phenomenal sense like a pipe.

The second argument relies on the claim that object recognition cannot be restored in associative agnosia by the ‘insertion’ of a belief about the perceived object’s high-level kind properties. The argument goes like this:

- (1) According to nonperceptual, doxastic models of object recognition, there are two components in visual recognition, a *belief* component and a *looking* component: seeing that such and such a type of object is present is a matter of forming an appropriate belief or judgment on the causal basis provided by one’s visual experience of the object.
- (2) But suppose that a patient with pure associative agnosia ‘suffers from a freak neurophysiological condition that causes him to believe that every object he is looking at is a pipe. This case satisfies . . . the causal condition on visual recognition, but it seems doubtful whether it suffices to reinstate the missing experiential content’ (Bayne 2009: 396).
- (3) Hence, doxastic models of object recognition are inadequate. ‘Perceptual recognition is not simply a matter of believing that such and such a type of object is present whilst enjoying low-level visual experience’ (Bayne 2009: 396).

This line of argument is doubly problematic. First, premise (2) again simply assumes that a layer of experiential content *is* missing in visual agnosia. Second, it is doubtful that proponents of nonperceptual models of object recognition would accept the freak neurophysiological condition described in premise (2) as an adequate causal basis for recognizing something as a pipe. In order to recognize an object *O* as a pipe, it might be argued, it isn’t sufficient that seeing *O* just *somehow* causes you to form the belief that *O* is a pipe. Rather, most psychological theories of object recognition suggest that the adequate causal basis minimally would involve perceiving *O* to have a certain cluster of visual attributes, e.g. a certain shape and size, and, in addition, computing

the similarity between *O* and a stored, inner representation of the category *pipe*. From this empirically informed perspective, a ‘freak’ neurophysiological condition that somehow causes you to believe that every visually perceived object is a pipe is a condition in which *no* object is visually recognized as a pipe.

Bayne, I conclude, fails to show that cases of pure visual associative agnosia provide any non-question-begging reason to think that the reach of perceptual phenomenality extends to high-level properties.

3.3 Empirical arguments for high-level properties in visual experience

Unlike Siegel and Bayne, William Fish (2013) has evinced skepticism about the prospects of using an essentially phenomenological methodology to determine the scope of phenomenal content. Instead, he appeals to three different sources of empirical evidence to motivate the conclusion that high-level properties appear in the ‘sensory, presentational component’ of visual experience—in its phenomenal character—and not merely in our cognitive or ‘interpretive’ responses to that component.⁴

Although Fish is primarily concerned with the question of *which* properties appear in the sensory component of visual experience rather than with the question of whether or not visual experience is cognitively penetrable, the empirical evidence he reviews is certainly germane to the latter. As Fiona Macpherson (2012) writes:

People who subscribe to the existence of high-level content in visual experience are likely to reject cognitive impenetrability. This is because many of the arguments for high-level content proceed by arguing that *learning* can affect which visual experience one has. . . . For example, plausibly, the visual system does not come ready-made to represent pine trees or the specific individual that is my brother. This is a type of tree and a particular human that not all humans will encounter. If all such specific representational abilities had to be built-in to the visual system it would be enormous and unwieldy. But, perhaps on repeated exposure to pine trees or to my brother, a subject can come to notice features that all and only such trees have and that all and only that person has. And perhaps the subject’s knowledge of these features can feed into their visual system so that they come to have visual experiences that are sensitive to those features, at least sensitive in a way that they were not before. In this way, the subject would come to have visual experiences that they did not have before—visual experiences that represent pine trees or my brother. In other words, a very plausible mechanism for visual experiences coming to have high-level content is that the visual system is penetrated by the cognitive system. (2012: 32–3)

Hence, if psychological findings suggest that high-level properties *are* represented in the phenomenal contents of visual experience—or, as Fish would rather put it, in its presentational component—then this would be seemingly good reason to suppose that visual phenomenal content has been penetrated by information originating outside of the visual system.

⁴ Although Fish’s naïve realist commitments lead him to speak of properties that appear in the presentational component of visual experience rather than properties that are *represented* in visual phenomenal content, this doesn’t affect my assessment of the evidence he adduces or the arguments he develops.

Fish's first source of evidence involves the finding that high-level object and scene properties can be perceived in the absence of focal attention. In particular, subjects engaged in a task that required them to attend to a 5-letter array presented in central vision were found to be able to determine with about 75% accuracy whether a scene presented in peripheral vision contained either animals or vehicles (Li et al. 2002). Here is the argument that Fish develops for the conclusion that high-level kind properties figure in the presentational component of visual experience:

- (1) 'It would seem independently plausible . . . to suppose that, for processes of interpretation to take place, we would need to allocate additional cognitive resources to the task. . . . [I]f a property requires attention to be perceived, perhaps this is evidence that it should be located in the interpretative component of a visual experience; *if it can be perceived preattentively, this is reason to think that it appears in the presentational component*' (Fish 2013: 49; emphasis added).
- (2) The best interpretation of findings reported by Li et al. (2002) is that kind properties such as *animal* or *vehicle* can be perceived without attention.
- (3) So there is reason to think that kind properties such as *animal* or *vehicle* can appear in the non-interpretive, presentational component of visual experience.

There is good evidence that an object's properties can be perceived without attention. Indeed, psychophysical findings garnered in the previously mentioned studies of inattentive blindness suggest that a pre-attentively perceived object may be (implicitly) categorized as belonging to a certain high-level kind (Mack and Rock 1998; Mack 2002). The problem is that the same studies also provide compelling evidence that attention is necessary for *conscious* perception. In other words, a pre-attentively perceived object may be categorized by the visual system as an *animal*, or a *vehicle*, or some other high-level kind even though the object is absent from the sensory, presentational component of visual experience, contrary to premise (1).⁵ In addition, the interpretation of the findings reported by Li et al. (2002) in premise (2) is contestable. In particular, there is no reason to think that the scenes presented in subjects' peripheral vision were completely unattended. Rather, as Jesse Prinz has pointed out, diffuse, nonfocal attention may have been allocated to them (Prinz 2012: 118–22). It is worth mentioning that the experimenters themselves conclude only that 'rapid visual categorization of novel natural scenes requires very little or no *focal attention*' (Li et al. 2002: 9599; emphasis added).

Fish's second source of evidence comes from studies that suggest that high-level properties can be rapidly detected. For example, Fabre-Thorpe et al. report that human subjects can detect the presence of an animal with 94% accuracy and a median reaction time of 445 ms in natural images that are flashed for only 20 ms (2001: 172).

⁵ See Cohen et al. (2012: 414) for further evidence in support of this assessment.

Fish assumes that, if properties can be detected quickly, then this is evidence that they appear in phenomenal character. Why? Quoting Prinz, he appeals to a dominant view of visual processing in cognitive science:

[P]erception may work in the following way. After transduction, a signal is propagated through a hierarchically organized sequence of subsystems, which begin by producing representations of local features and move on to representations that are more global and invariant.

(Prinz 2006: 456)

The idea, I take it, is that if a property is detected quickly, then, given the above view of how visual perception works, its detection must occur at a relatively early stage in the visual processing hierarchy and, hence, prior to the involvement of cognitive, interpretative mechanisms.

The problem is that the above account of perceptual processing actually implies that detection of viewpoint-invariant, high-level kind properties occurs at the uppermost stages of the processing hierarchy—for example, in the lateral occipital complex—not at earlier, feature-representing stages such as V3 or V4 (see e.g. Grill-Spector and Malach 2004). Categorization may be astonishingly rapid, but it is driven to a significant extent by even faster local feature processing. So, contrary to Fish, the rapidity with which objects and scenes are categorized as having certain high-level properties does not provide evidence that those properties are represented at a ‘pre-interpretive’ level of perceptual processing.

Fish’s final source of evidence comes from studies of visual adaptation and its aftereffects. Adaptation to an oblique grating, for example, can cause vertical lines subsequently to look tilted in the opposite direction. Similarly, adaptation to the downward flow of water in a waterfall can make stationary objects on the riverbank appear to flow upwards for a few seconds. While the underlying mechanism of adaptation isn’t fully understood, it is generally thought to involve reduction of activity and/or sensitivity in neural populations involved in processing the adapting stimulus attribute (Grill-Spector et al. 2006).

Fish appeals to the existence of seemingly high-level adaptation aftereffects to motivate the claim that high-level properties appear in the presentational component of visual experience. In particular, he appeals to evidence provided by Burr and Ross (2008) that *numerosity* is susceptible to adaptation. Adapting to a large number of dots on a circular patch, Burr and Ross report, decreases the apparent numerosity of a neutral (averagely populated) patch, while adapting to a small number of dots has the opposite effect.⁶ The conclusion that high-level properties can sometimes appear in the presentational component of visual experience is then inferred by means of the following argument:

⁶ A demonstration is available here: <<http://www.current-biology.com/cgi/content/full/18/6/425/DC1/>>

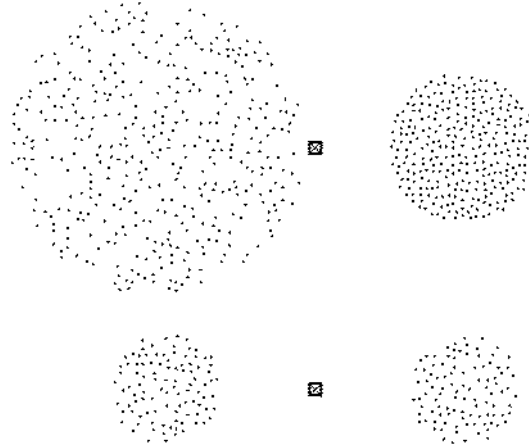


Figure 7.1 Adapting texture density versus adapting numerosity. Reproduced, with permission, from Durgin (2008)

- (1) '[A]ll agreed primary visual properties—the properties (such as size, orientation, shape, colour and motion) that everyone agrees appear in phenomenal character—are susceptible to adaptation' (Fish 2013: 52).
- (2) 'So if we can show that another property is also susceptible to adaptation, we have an argument that this property appears in phenomenal character too' (Fish 2013: 52).⁷
- (3) Burr and Ross (2008) provide evidence that numerosity is susceptible to adaptation.
- (4) So there is evidence to think that numerosity, a high-level property, appears in phenomenal character. As Burr and Ross conclude: 'just as we have a direct visual sense of the reddishness of half a dozen ripe cherries, so we do of their sixishness. In other words there are distinct qualia for numerosity, as there are for color, brightness, and contrast' (2008: 426).

It is open to question, however, whether Burr and Ross (2008) do in fact provide evidence for adaptation to numerosity. An empirically well-motivated, alternative explanation is that the numerosity judgments made in Burr and Ross's experiment are based on perceived, relative *texture density* rather than perceived, relative numerosity (Durgin 2008). Consider Figure 7.1. Following adaptation to the two circular texture patches in the top row, a neutral patch subsequently presented within the region adapted to the less numerous but more dense texture (bottom row, right) appears less dense, and hence less numerous, than a neutral patch presented within the region

⁷ The inference from (1) to (2) commits the converse error, but I shall put this objection aside to focus on premise (3).

adapted to the more numerous but less dense texture (bottom row, left). Unlike numerosity, however, texture density is a low-level visual property. So evidence that texture density is susceptible to adaptation does not support the claim that high-level properties can appear in phenomenal character.

Fish also appeals to evidence that adaptation to a human face can affect the visual appearance of subsequently seen faces. Adapting to a picture of a face that has been horizontally or vertically extended, for example, makes a picture of the original, non-distorted face appear distorted in the opposite direction (Webster and MacLin 1999). Adapting to an 'anti-face' (see Figure 7.2)—a synthetic face created by locating the position of a target face relative to the average face in high-dimensional face space and then inverting the differences between them—makes the average face appear more like the target face and also makes it easier for subjects to identify the target face (Leopold et al. 2001). Adapting to a female face makes a subsequently presented androgynous face—formed by morphing between a female and male face—appear more masculine (Webster et al. 2004). Adapting to certain facial structures and textures and can lead to significant aftereffects in apparent age (O'Neil and Webster 2011).⁸

Corresponding to specific differences in configural and surface properties between faces, then, there are specific differences in adaptation and resultant visual aftereffects. Webster and MacLeod (2011) argue that these aftereffects support a norm-based model of face coding and a normalization model of face adaptation. According to the norm-based coding model, individual faces are represented in the visual brain in terms of their direction and distance from a prototypical or average face in a high-dimensional, perceptual face space (see Loffler et al. 2005 and Leopold et al. 2006 for consistent fMRI evidence). And according to the normalization model of face adaptation, adapting to an individual face 'renormalizes' perceptual face space so that the face that appears perceptually neutral or average is shifted in the direction of the adapting face.⁹

Studies and current theoretical models of face adaptation, while suggestive, don't seem to provide compelling evidence that high-level face properties figure in what Fish calls the 'presentational component' of conscious visual experience. One source of difficulty is that aftereffects induced by complex, real-world objects are likely to reflect

⁸ For a detailed review of findings and theoretical assessment, see Webster and MacLeod (2011).

⁹ Webster and MacLeod review several sources of evidence for a norm-based coding system. Here are two. First, adapting to a configurally distorted face significantly alters the appearance of a neutral, undistorted face, but adapting to the neutral face does not affect the appearance of the distorted face (Webster and MacLin 1999). The renormalization model predicts this asymmetry: 'the neutral face simply reinforces the current norm or neutral point in face space and hence changes nothing, whereas the distorted adapting face induces a shift in the neutral point so that the previously neutral faces are no longer seen as such' (Webster and MacLeod 2011: 1708). Second, the appearance of the adapting face is itself altered by the adaptation process so that it looks less distinctive, i.e. less divergent from the norm, and, further, produces a global shift in the same direction for all other faces in face space, e.g. adapting to a horizontally extended face makes every subsequently seen face look narrower. This again suggests the immediate result of adaptation is a 'recentering' of face space nearer to adapting stimulus so that the adapting face appears perceptually less distinctive and closer to the average.

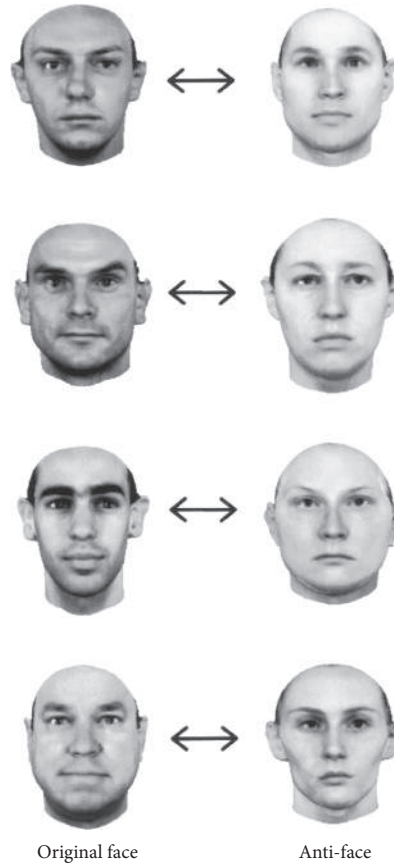


Figure 7.2 Faces and their corresponding anti-faces. Reproduced, with permission, from Leopold et al. (2001)

changes in sensitivity at all levels of the visual processing hierarchy, starting as early as the retina (this is the problem of what Xu et al. 2008 call ‘aftereffect inheritance’). In consequence, aftereffects that are suggestive of adaptation to some high-level attribute or category may in many cases be caused by adaptation of simple feature detectors or mechanisms that represent ‘holistic’ but low-level stimulus attributes. Thus Webster and MacLeod write:

[T]he selectivity of the adaptation for different natural facial categories [e.g. *race* or *gender*] does not require that the mechanisms of face coding are directly tuned to these specific categories, for as long as the mechanisms can be differentially adapted by different categories they will result in selective aftereffects. (2011: 1715)

[I]t remains unclear whether there are special axes in the representation of faces—perhaps corresponding to prominent natural categories—or if the selectivity of the aftereffects *simply reflects the distances between the stimuli in face space* . . . The point is that we cannot be confident that facial attributes that appear salient or ecologically relevant are the stimulus attributes that face processing is directly encoding. (2011: 1715; emphasis added)

In short, the available scientific evidence underdetermines the question of whether the aftereffects observed in face adaptation studies reflect changes in the sensitivity of neurons that explicitly represent high-level facial attributes or at the level of neurons that code for the lower-level ‘building blocks’ of possible faces. Hence, it is unclear at present how best to characterize the nature of face adaptation aftereffects.

Preliminary grounds for skepticism about Fish’s assessment, however, comes from a recent study of cross-category adaptation. Javadi and Wee (2012) adapted subjects to images of objects strongly associated with one gender, e.g. lipstick for females or motorcycles for males, and then asked them to judge the gender of an androgynous face. They found that subjects were more likely to characterize the appearance of the androgynous face as masculine when the images were of objects associated with women, and vice versa. It is highly implausible, however, that there is a high-level, visible property or ‘look’ that is common to feminine faces, tubes of lipstick, and tiaras (or to masculine faces, motorcycles, and electric razors). But, if not, then at least some gender-related visual aftereffects do not seem to arise in consequence of adaptation to the presence of high-level gender properties in visual phenomenal content.¹⁰

4 Low-Level Effects on Visual Phenomenal Content

According to Thesis P, visual phenomenal content is *not* informationally encapsulated, i.e. it can be penetrated by information from outside the visual system. In the last section, I criticized recent ‘high-road’ approaches according to which acquiring the ability to categorize an object as belonging to a certain high-level kind can cause the relevant kind property to be represented in visual experience (or to be present in what Fish calls its ‘presentational component’) and thereby alter the way the object subsequently appears.

The approach to defending Thesis P, examined in this section, doesn’t undertake to show that inputs from cognitive systems can extend the reach of visual phenomenal content: it is neutral with respect to the question of whether such inputs can augment the range of attributes that objects may be represented as having in conscious seeing.

¹⁰ It should also be emphasized that there are numerous respects in which perceptual processing for faces is special, i.e. involves face-specific cognitive and neural processes (for discussion, see McKone et al. 2006). Hence, even if expansionist conclusions with respect to face perception were empirically warranted by face adaptation research, there would be no guarantee that they could be extended unproblematically to perception of other kinds of objects.

Rather, it relies on psychophysical and neuroscientific evidence to motivate the claim that information originating outside the visual system can modulate the way an object's low-level properties visually appear.

There is a large body of evidence that when visual information about the environment conflicts with information received in another modality, vision typically dominates. In the ventriloquism effect, for example, the visually perceived location of a stimulus captures auditory localization of a sound source. Vision has also been found to exert a dominating influence on both touch and proprioceptive estimates of limb position and body posture (Stratton 1899; Harris 1980; Rock and Harris 1967). The effect of vision on other modalities is not limited to spatial localization. It can also have a striking influence on speech perception. In the McGurk effect, for example, which syllable a speaker is heard to pronounce strongly depends on how her lips visually appear to move.

It would be a mistake, however, to think that when the senses conflict, vision is always the recalcitrant modality. While evidence that vision can affect phenomenal contents in other modalities has been available for over a century (Stratton 1899), recent findings from studies of multisensory integration suggest that the influence can also go in the other direction. In what follows, I present four cases in which sources of information originating outside of the visual system affect the way an object's low-level properties appear in visual experience. The first three are cases of what I shall refer to as 'synchronic, low-level informational penetration' (or SLIP). A fourth case, by contrast, suggests that visual phenomenal content can sometimes be cognitively penetrated by the subject's high-level intentions to engage in motor actions.

4.1 *Tactile-visual SLIP*

There are numerous, independently variable sources of information in the light sampled by the eye. The more important of these include binocular disparity, vergence, occlusion, motion parallax, texture gradients, shading, reflections, and relative size (for a comprehensive review, see Bruce et al. 2003). In many cases, however, visual estimation of an environmental property is optimized by recruiting sources of auxiliary information from outside the visual system. These may be directly integrated with available sources of optical stimulus information and/or used to change the weighting assigned to one or another source of such information.

Receiving touch-based information about a surface's 3D orientation, for example, can influence the way the surface subsequently visually appears. Ernst et al. (2000) performed an experiment in which texture-based and disparity-based sources of optical information about a surface's slant were set in conflict. They found that haptic feedback consistent with one or the other source of visual information changed the subsequent visual appearance of the surface's slant for up to twenty-four hours. For example, when haptic signals were consistent with the texture-specified slant, the visual appearance of the surface's 3D orientation was closer to the texture-specified slant than before training.

This finding indicates that the weighting of a source of visual information in perceptual processing is not always determined by its overall ecological reliability, but can also be modulated contextually by feedback from other senses. In a recent study, van Beers et al. (2011) undertook to estimate how rapidly the weightings assigned to visual cues change during training with haptic feedback. Subjects were instructed to place a cylinder flush onto a slanted, checkerboard surface. The apparent 3D orientation of the surface was perceptually determined on the basis of both monocular and binocular cues. In some trials, these cues could conflict; when this was the case, the haptic feedback that subjects received as they placed the cylinder on the surface was consistent with one but not the other cue. Van Beers et al. estimated the weights respectively assigned to monocular and binocular cues in each trial by monitoring the orientation of the cylinder just before it contacted the surface. They discovered that significant cue reweighting occurred after only about 32 conflict trials with small but non-negligible changes in weights after *each* trial (2011: 10). This result constitutes empirical evidence that visual orientation perception is susceptible to synchronic, low-level informational penetration (SLIP) by touch.

4.2 Audiovisual SLIP

As mentioned above, in the most familiar examples of multisensory integration, vision exerts a dominating influence on other senses. The previous subsection, however, shows that the influence can flow in the other direction, with vision exhibiting susceptibility to change in the face of intersensory conflict with touch. In this subsection, I review evidence that what we see can also be influenced by what we *hear*.

In the 'stream-bounce illusion' (Sekuler et al. 1997), audition influences an object's perceived trajectory in space. Subjects are shown an ambiguous display in which two objects can be perceived either as streaming through each other (without a change in their directions) or as bouncing off of each other (with a reversal in their directions), depending on whether or not a brief sound is heard when they coincide. The nature of the illusory effect, however, isn't fully understood. One possibility is that the sound modulates motion perception simply because it causes attention to be momentarily withdrawn from the moving targets (Watanabe and Shimojo 1998). Consistent with this hypothesis it has been found that other transient attention-capturing stimuli, e.g. a visual flash or a brief tap on the subject's finger, bias the visual system toward the bouncing interpretation of object motion. This speaks against the presence of audio-visual SLIP in the etiology of the illusion.

Persuasive evidence for audio-visual SLIP, however, comes from the 'sound-induced flash illusion' (Shams et al. 2000; 2002). When a black disk is flashed once, but accompanied by two brief beeps, it is visually experienced as having been flashed twice. 'This phenomenon', Shams and co-authors suggest, 'clearly demonstrates that sound can alter the visual percept qualitatively even when there is no ambiguity in the visual stimulus' (2002: 23). Information originating in the auditory system, it appears, can sometimes influence the phenomenal contents of visual experience.

Several sources of evidence point to a perceptual, non-cognitive interpretation of this effect. First, the illusion is insensitive to the observer's beliefs: knowing that the disc has only been flashed once does not eliminate the illusion. Such belief-independence is often taken to be a hallmark of perception. Second, the effect does not appear to be due to either task difficulty or to a cognitive strategy that involves counting the number of beeps. Third, neuroimaging studies have shown that the illusion is correlated with changes in activity in low-level visual cortical areas, including V1 (Watkins et al. 2007; Mishra et al. 2007). This speaks strongly against a cognitive interpretation of the effect. Last, Rosenthal et al. (2009) found that the illusion is significantly resistant to feedback training in which subjects are apprised about their errors. Together, these findings suggests that audition is genuinely affecting visual phenomenal content by means of SLIP.

4.3 *Proprioceptive-visual SLIP*

Size constancy is a basic feature of visual experience. We see a fire hydrant, for instance, as having roughly the same intrinsic or 'absolute' size despite variations in our perspective and other viewing conditions. What is it, however, to perceive an object's intrinsic size and not merely its size relative to other objects in our field of view? Clearly, it is not enough that there be some proximal stimulus variable whose values correspond to variations in the extent of the object. Perceivers must also make use of a spatial scaling scheme of some kind that assigns intrinsic size meaning to those values.¹¹ In the case of human perception, how is such scaling achieved? If the units in terms of which perceived spatial extents are scaled are not conventional units such as feet or inches (Peacocke 1986), then *how* are they scaled?

One philosophically influential answer to this question, dating back to Malebranche in the 17th century, is that an object's perceived size is scaled to one's sense of the size of one's own body. 'Our sight', Malebranche writes, 'does not represent extension to us as it is in itself, but only as it is in relation to our body' (1997[1674–5]: I.6, §1; quoted by Simmons 2003). The purpose of spatial vision isn't to provide us with information about objects suitable for precise, mathematical calculation, but rather to adapt our actions to the structure of the 3D environment. In keeping with this purpose, visual experience represents properties such as size, shape, and distance using a body-relative scaling scheme that action-planning systems understand. As Simmons puts it, 'Sensory perception . . . reflects one's own body in the way it represents objects. This in turn helps to explain why sensory perception is especially suited to helping the mind to direct its body safely through the world' (2003: 396).

The idea that visual perception makes use of an egocentric, body-relative scaling system finds supports in several decades of psychophysical work on size and distance perception. One reliable source of information about spatial layout, for example, is the 'horizon ratio' (Sedgwick 1973; 1986). Because the horizon line is always at eye level

¹¹ An analogous problem arises with respect to distance perception. See Schwartz (2006: 37).

in the optic array, the height of an object from its base to the point where it visually intersects the horizon is equal in height to the elevation of the perceiver's eyes above the ground. Hence, the ratio of the object's height (h) to the height of the perceiver's eyes (e) is approximately equal to the ratio of the vertical visual angle subtended by the object (H) to the vertical visual angle subtended by the portion of the object below the horizon (E):

$$h/e = H/E$$

In other words, the height of an object, when estimated using the horizon ratio, is scaled in eye-height units. For example, if a tree's horizon ratio is 5:1, then the tree is approximately five eye-height units tall.

There is psychophysical evidence that the visual system makes use of the horizon ratio, among other sources of information, when estimating an object's height (Mark 1987; Warren and Whang 1987; Rogers 2003) as well as when estimating its egocentric distance (Ooi et al. 2001) and width (Wraga 1999).¹² If human perceivers do make use of eye-height-scaled information, then it should, in principle, be possible to modify how far and/or how large objects appear in visual experience by altering a subject's awareness of the size of her own body.¹³ This prediction is borne out by a recent study of the perceptual effects of the so-called 'body swap' illusion.

Partial body-ownership illusions are well known. For example, in the rubber-hand illusion (Botvinick and Cohen 1998), subjects vividly experience tactile sensations as arising in a prosthetic rubber hand when it is stroked in synchrony with strokes applied by the experimenter to the subject's own, unseen hand. Using a similar visuotactile stimulation technique, however, it is also possible to induce the illusion of owning an *entire body* distinct from one's own (Lenggenhager et al. 2007; Petkova and Ehrsson 2008; Lenggenhager et al. 2009). In 2011 study, van der Hoort and colleagues sought to determine whether it is possible (a) to induce illusory ownership of an abnormally large or small artificial body and, if so, (b) whether the induced illusory experience, in turn, influences visual perception of spatial layout.

Subjects, lying on their back and looking toward their toes, were presented via a head-mounted display with real-time video of the abdomen and legs of an artificial body, varying between 30, 80, and 400 cm in height. To induce the body-swap illusion, the participant's body and the artificial body were touched in synchrony for four minutes. This resulted in a powerful body-swap illusion, even when subjects were visually presented with the torso and legs of a small, 30-cm doll. '[M]ost participants', van der Hoort et al. write, 'were not aware of the extremely small size of the doll that

¹² For a helpful overview, see Bennett (2011).

¹³ Warren and Whang (1984) showed that subjects overestimate height and width when effective eye-height is underestimated. Their experimental paradigm, however, involved use of a false floor that surreptitiously increased the target's H/E ratio rather manipulation of subjects' internal representations of body size.

they felt ownership of. Instead, they experienced themselves to be located in a “giant world” (2011: 4).

The experimenters then investigated whether the illusory change in body size also evoked changes in the perceived distances and sizes of external objects shown in the display. Subjects, when asked to walk with their eyes closed to the previously seen location of an object, it was found, walked a significantly *longer* distance during the small-body condition and a significantly *shorter* distance during the large-body condition. Verbal estimates of object size were found to be congruent with behavioral measures of distance perception. When subjects experienced themselves as owning a very large body, for example, they saw objects not only as closer but also as smaller than when they experienced themselves as owning either a very small or a normal-sized body.

In order to exclude the possibility that these effects were due to use of the body seen in the display as a visual, relative size cue, van der Hoort et al. manipulated the strength of the illusion by touching the prosthetic body in or out of synchrony with touches applied to the subject’s own body. Although the size of the body seen in display was the same in both conditions, only the synchronous touch condition lead to significant distortions in visual space perception. In other words, subjects only experienced distances as significantly shorter or longer when they had the illusory experience of *owning* the prosthetic body.

Two points are important. First, perceivers’ representations of their own body size appear to be both highly labile and rapidly modifiable under conditions of intersensory conflict: when synchronous visual and tactile cues disagree, the brain swiftly alters its representation of the body’s metric properties in deference to vision. Second, the findings reviewed here suggest that perceivers use representations of their own body size to calibrate visual estimates of the distances and sizes of objects in the external world, much as Malebranche originally proposed. Taken together, they provide further support for proprioceptive-visual SLIP. In particular, they provide support for the claim that the computations underlying conscious visual space perception make use of nonvisual, proprioceptive information about body metrics as an informational resource.

4.4 Intentions for action penetrate visual size perception

A number of studies, beginning with Aglioti et al. (1985), have reported that the Ebbinghaus illusion has a significantly greater effect on visual awareness than on visually guided grasping. Although the disk surrounded by small circles in the illusion display typically *looks* about 10% larger than the disk surrounded by large circles, the increase in maximum grip aperture when reaching for the former disk exhibits a magnitude of only 6%.

This finding in addition to various pieces of neuropsychological evidence has been used to motivate a dual-systems model of visual processing according to which ‘vision-for-perception’ and ‘vision-for-action’ are subserved by functionally and anatomically

distinct processing streams in the primate brain (Milner and Goodale 1995/2006).¹⁴ According to proponents of the dual-systems model, the illusion has a different effect on visual awareness than on visually guided grasping because the former makes use of different sources of visual size information than the latter. On this model, how the size of an object appears in conscious vision should not influence grip aperture; conversely, how the size of the object is represented by motor systems that guide grasping should not influence representation of its size in conscious vision.

At variance with this idea, however, Vishton et al. (2007) found that the act of reaching for a disk in a three-dimensional version of the Ebbinghaus illusion significantly diminished the magnitude of the effect on subsequent perceptual estimation (5.74% for perceptual estimation vs. 6.10% for grasping). Strikingly, they also found that when subjects were *merely informed* prior to engaging in the perceptual estimation task that they would subsequently be required to grasp the disk that appeared larger, the effect of the illusion on visual awareness was again significantly diminished (6.18% for perceptual estimation vs. 5.54% for grasping). 'Simply listening to a description of a reaching task', the experimenters write, 'seems to affect size perception' (2007: 718).

These findings suggest that visual phenomenal content can be cognitively penetrated: high-level information from outside of the visual system seems to alter the way an object's size appears. There are different possible mechanisms whereby such penetration might occur. Vishton et al. propose that 'intending to reach for a target changes how the reacher perceives it' and that 'action choice changes the nature of visual size perception' (2007: 318). But how does action selection have this effect? One possibility (a) is that an abstract, high-level intention to act—either a 'distal' or 'proximal' intention in the sense of Pacherie 2008—somehow exerts a direct influence on perceptual estimation, say, by changing the relative weightings assigned by vision to sources of depth information such as binocular disparity, vergence, and accommodation. Since size estimation depends, in part, on perceived distance, this could explain the influence of intention on perception. A second possibility (b) is that the effect is brought about via lower-level motor representations that implement the subject's high-level intention. This would arguably still count as a case of cognitive penetration if the lower-level motor representations carried information from the subject's high-level intention that influenced relative cue weighting or other visual computations (Wu 2013). A third possibility (c) looks to elicited motor imagery for the source of penetration. Possibility (c), however, is not entirely distinct from (a) and (b), since there is evidence that internally rehearsing the performance of an action activates representations at all levels in the motor processing hierarchy. A final possibility (d) is that the effect isn't due to motor representations at all, but rather to the subject's beliefs about the action she has been requested to perform. Future studies must investigate

¹⁴ For different philosophical assessments of the dual systems model, see Briscoe (2008), Clark (2009), and Briscoe and Schwenkler (forthcoming).

which, if any, of these explanations best accounts for the effects that Vishton and his co-authors report.¹⁵

5 Conclusion

This chapter examined two approaches to the question of whether visual phenomenal content can be influenced by information originating outside the visual system. ‘High-road’ arguments undertake to show that acquiring the capacity to categorize an object as belonging to a certain high-level kind can modify the way the object subsequently appears by causing the relevant kind property to be represented in perceptual experience. Both phenomenological and psychological considerations put forward in support of this conclusion were examined and found to be unconvincing. ‘Low-road’ arguments, by contrast, seek to show that information from outside the visual system can modulate the way an object’s low-level properties visually appear. In the previous section, I presented some compelling psychophysical and neuroscientific evidence for this view. The empirical findings scouted there suggest that visual phenomenal content is not only influenced by information imported from other peripheral input systems, such as audition and touch, but that it can also be penetrated by proprioceptive representations of the size of one’s own body as well as by computationally high-level motor intentions.¹⁶

References

- Aglioti, S., Goodale, M., and DeSouza, J. (1995). Size contrast illusions deceive the eye but not the hand. *Current Biology* 5: 679–85.
- Balcetis, E., and Dunning, D. (2010). Wishful seeing: more desired objects are seen as closer. *Psychological Science* 21: 147–52.
- Bayne, T. (2009). Perception and the reach of phenomenal content. *Philosophical Quarterly* 59: 385–404.
- Bennett, D. (2011). How the world is measured up in size experience. *Philosophy and Phenomenological Research* 83: 345–65.
- Blaser, E., Sperling, G., and Lu, Z. (1999). Measuring the amplification of attention. *Proceedings of the National Academy of Sciences* 96: 11681–6.
- Botvinick, M., and Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391: 756.
- Briscoe, R., and Schwenkler, J. (forthcoming). Conscious vision in action. *Cognitive Science*.

¹⁵ A number of psychologists in what might be called the ‘New New Look’ tradition have recently reported that anticipated motor effort and/or affective states can significantly influence subjects’ perceptions of spatial layout (Proffitt et al. 2003; Proffitt 2006; Balcetis and Dunning 2010). While many of these studies seem to provide further support for cognitive penetration of conscious visual space perception, recent counterstudies (Woods et al. 2009; Durgin et al. 2009; Durgin et al. 2010; Durgin et al. 2011) suggest that a number of the reported effects may be artifacts of experimental design.

¹⁶ I am grateful to Ruth Millikan for thought-provoking discussions of ideas in this chapter.

196 R. BRISCOE

- Bruce, V., Greene, P., and Georgeson, M. (eds) (2003). *Visual Perception: Physiology, Psychology, and Ecology*, 4th edn. London: Psychology Press.
- Burr, D., and Ross, J. (2008). A visual sense of number. *Current Biology* 18: 425–8.
- Clark, A. (2009). Perception, action, and experience: unraveling the golden braid. *Neuropsychologia* 47: 1460–68.
- Cohen, M., Cavanagh, P., Chun, M., and Nakayama, K. (2012) The attentional requirements of consciousness. *Trends in Cognitive Sciences* 16: 411–17.
- Durgin, F. (2008). Texture density adaptation and visual number revisited. *Current Biology* 18: R855–6.
- Durgin, F., Baird, J., Greenburg, M., Russell, R., Shaughnessy, K., and Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin and Review* 16: 964–9.
- Durgin, F., DeWald, D., Lechich, S., Li, Z., and Ontiveros, Z. (2011). Action and motivation: measuring perception or strategies? *Psychonomic Bulletin and Review* 18: 1077–82.
- Durgin, F., Hajnal, A., Li, Z., Tonge, N., and Stigliani, A. (2010). Palm boards are not action measures: an alternative to the two-systems theory of geographical slant perception. *Acta Psychologica* 134: 182–9.
- Ernst, M., Banks, M., and Bühlhoff, H. (2000). Touch can change visual slant perception. *Nature Neuroscience* 3: 69–73.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., and Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience* 13: 1–10.
- Fish, W. (2013). High-level properties and visual experience. *Philosophical Studies* 162: 43–55.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, Mass.: MIT Press.
- Fodor, J. (2000). *The Mind Doesn't Work That Way*. Cambridge, Mass.: MIT Press.
- Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences* 10: 14–23.
- Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience* 27: 649–77.
- Harris, C. (1980) Insight or out of sight? Two examples of perceptual plasticity in the human adult. In Charles S. Harris (ed.), *Visual Coding and Adaptability*, 95–149. Hillsdale, NJ: Erlbaum.
- Henderson, J., Brockmole, J., Castelano, M., Mack, M., Fischer, M., Murray, W., and Hill, R. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. Van Gompel, M. Fischer, W. Murray, and R. Hill (eds), *Eye Movements: A Window on Mind and Brain*, 537–62. Amsterdam: Elsevier.
- Henderson, J., and Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology* 50: 243–71.
- Javadi, A., and Wee, N. (2012). Cross-category adaptation: objects produce gender adaptation in the perception of faces. *PLOS One* 7: e46079.
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology* 4: 219–27.
- Kriegel, U. (2002). Phenomenal content. *Erkenntnis* 57: 175–98.
- Lenggenhager, B., Mouthon, M., and Blanke, O. (2009). Spatial aspects of bodily self-consciousness. *Consciousness and Cognition* 18: 110–17.

- Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317: 1096–9.
- Leopold, D. A., Bondar, I. V., and Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442: 572–5.
- Leopold, D. A., O’Toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience* 4: 89–94.
- Li, F. F., VanRullen, R., Koch, C., and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America* 99: 9596–9601.
- Loffler, G., Yourganov, G., Wilkinson, F., and Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience* 8: 1386–90.
- Mack, A. (2002). Is the visual world a grand illusion? A response. *Journal of Consciousness Studies* 9: 102–10.
- Mack, A., and Rock, I. 1998. *Inattentional Blindness*. Cambridge, Mass.: MIT Press.
- Macpherson, F. (2012). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research* 84: 24–62.
- Malebranche, N. (1997[1674–5]). *The Search After Truth*, trans. T. Lennon and P. Olscamp. Cambridge: Cambridge University Press.
- Mark, L. (1987). Eyeheight-scaled information about affordances. *Journal of Experimental Psychology: Human Perception and Performance* 13: 361–70.
- McKone, E., Kanwisher, N., and Duchaine, B. (2006). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences* 11: 8–15.
- Milner, D., and Goodale, M. (1995/2006). *The Visual Brain in Action*, 2nd edn. Oxford: Oxford University Press.
- Mishra, J., Martinez, A., Sejnowski, T., and Hillyard, S. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience* 27: 4120–31.
- O’Neil, S., and Webster, M. (2011). Adaptation and the perception of facial age. *Visual Cognition* 19: 534–50.
- Ooi, T., Wu, B., and He, Z. (2001). Distance determined by angular declination below the horizon. *Nature* 414: 197–200.
- Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition* 107: 179–217.
- Peacocke, C. (1986). Analogue content. *Proceedings of the Aristotelian Society*, Supplementary Volume 60: 1–17.
- Petkova, V., and Ehrsson, H. (2008) If I were you: perceptual illusion of body swapping. *PLoS One* 3: e3832.
- Prinz, J. (2006). Beyond appearances: the content of sensation and perception. In T. Gendler and J. Hawthorne (eds), *Perceptual Experience*, 434–59. Oxford: Clarendon Press.
- Prinz, J. (2012). *The Conscious Brain*. Oxford: Oxford University Press.
- Proffitt, D. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science* 1: 110–22.
- Proffitt, D., Stefanucci, J., Banton, T., and Epstein, W. (2003). The role of effort in perceived distance. *Psychological Science* 14: 106–12.
- Polyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22: 341–423.

- Rock, I., and Harris, C. (1967). Vision and touch. *Scientific American* 216: 96–104.
- Rosenthal, O., Shimojo, S., and Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography* 21: 185–92.
- Rubens, A., and Benson, D. (1971). Associative visual agnosia. *Archives of Neurology* 24: 305–16.
- Schwartz, R. (2006). *Visual Versions*. Cambridge, Mass.: MIT Press.
- Sedgwick, H. (1973). The visible horizon: a potential source of visual information for the perception of size and distance. Doctoral dissertation, Cornell University. *Dissertation Abstracts International*, 34, 1301B–2B. (University Microfilms no. 73–22,530).
- Sedgwick, H. (1986). Space perception. In K. Boff, L. Kaufman, and J. Thomas (eds), *Sensory Processes and Perception: Handbook of Perception and Human Performance*, vol. 1, 21.1–21.57. New York: Wiley.
- Sekuler, R., Sekuler, A., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385: 308.
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature* 408: 788.
- Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research* 14: 147–52.
- Siegel, S. (2006). Which properties are represented in perception? In T. Gendler and J. Hawthorne (eds), *Perceptual Experience*, 481–503. Oxford: Oxford University Press.
- Siegel, S. (2010). *The Contents of Visual Experience*. New York: Oxford University Press.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs* 46: 201–22.
- Simmons, A. (2003). Spatial perception from a Cartesian point of view. *Philosophical Topics* 31: 395–424.
- Stratton, G. (1899). The spatial harmony of touch and sight. *Mind* 8: 492–505.
- Underwood, G., Foulsham, T, van Loon, E., Humphreys, L., and Bloyce, J. (2006). Eye movements during scene inspection: a test of the saliency map hypothesis. *European Journal of Cognitive Psychology* 18: 321–42.
- van Beers, R., van Mierlo, C., Smeets, J., and Brenner, E. (2011). Reweighting visual cues by touch. *Journal of Vision* 11(20): 1–16.
- van der Hoort, B., Guterstam, A., and Ehrsson, H. (2011). Being Barbie: the size of one's own body determines the perceived size of the world. *PLoS ONE* 6: e20195 (1–12).
- Vishton, P., Stephens, N., Nelson, L., et al. (2007). Planning to reach for an object changes how the reacher perceives it. *Psychological Science* 18: 713–19.
- Watanabe, K., and Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception* 27: 1041–54.
- Watkins, S., Shams, L., Josephs, O., and Rees, G. (2007). Activity in human V1 follows multisensory perception. *Neuroimage* 37: 572–8.
- Watkins, S., Shams, L., Tanaka, S., Haynes, J., and Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31: 1247–56.
- Webster M., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature* 428: 557–61.
- Webster, M., and MacLeod, D. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B* 366: 1702–25.
- Webster, M., and MacLin, O. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin and Review* 6: 647–53.

- Woods, A., Philbeck, J., and Danoff, J. (2009). The various perceptions of distance: an alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance* 35: 1104–17.
- Wolfe, J. (1994). Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin and Review* 1: 202–38.
- Wolfe, J., Horowitz, T., Kenner, N., Hyle, M., and Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision research* 44: 1411–26.
- Wraga, M. (1999). The role of eye-height in perceiving affordances and object dimensions. *Perception and Psychophysics* 61: 490–507.
- Wu, W. (2013). Visual spatial constancy and modularity: does intention penetrate vision? *Philosophical Studies* 165: 647–69.
- Xu, H., Dayan, P., Lipkin, R. M., and Qian, N. (2008). Adaptation across the cortical hierarchy: low-level curve adaptation affects high-level facial-expression judgments. *Journal of Neuroscience* 28: 3374–83.