Rev.Phil.Psych.
DOI 10.1007/s13164-015-0253-4

Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems

Gary Lupyan[1]

Abstract The goal of perceptual systems is to allow organisms to adaptively respond to ecologically relevant stimuli. Because all perceptual inputs are ambiguous, perception needs to rely on prior knowledge accumulated over evolutionary and developmental time to turn sensory energy into information useful for guiding behavior. It remains controversial whether the guidance of perception extends to cognitive states or is locked up in a "cognitively impenetrable" part of perception. I argue that expectations, knowledge, and task demands can shape perception at multiple levels, leaving no part untouched. The position advocated here is broadly consistent with the notion that perceptual systems strive to minimize prediction error en route to globally optimal solutions (Clark *Behavioral and Brain Sciences* 36(3):181–204, 2013). On this view, penetrability should be expected whenever constraining lower-level processes by higher level knowledge is minimizes global prediction error. Just as Fodor feared (e.g., Fodor *Philosophy of Science* 51:23–43, 1984, *Philosophy of Science* 51:23–43, 1988) cognitive penetration of perception threatens theory-neutral observation and the distinction between observation and inference. However, because theories themselves are constrained by the task of minimizing prediction error, theory-laden observation turns out to be superior to theory-free observation in turning sensory energy into useful information.

1 Introduction: Perception as a Constructive Process

Beginning sailors are often shocked to learn that it is possible to sail into the wind. They envision wind as a pushing force and rightly wonder how such a force can pull a boat. But of course wind is just a form of energy and with appropriate tools we can shape that energy to do useful work—from pulling and pushing, to turning and singing.

✉ Gary Lupyan
lupyan@wisc.edu

[1] Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

Just as it is tempting to think of wind as a force passively captured by the sails, it is tempting to think of perception as passively capturing the information that surrounds us. Just as sails capture the wind, our eyes capture photons, our ears vibrations, our skin mechanical and thermal forces, and after transducing them into neural signals and what many textbooks refer to as "further processing" out comes a sensation of sight, sound, hotness, etc. On this view, because cognitive states are downstream consequences of perceptual causes, claims that cognitive states can affect perception are thought to be (vaguely) in conflict with basic causality.[1]

But this view is wrong. Rather than a passive process of transduction, perception is more accurately viewed as a constructive process of turning various forms of energy (mechanical, chemical, electromagnetic) into *information* useful for guiding behavior. The idea that perception is a constructive process is, at least within psychology, not especially controversial. It underpins the influential view of perception as "unconscious inference" (see Barlow 1990), pervades William James's *Principles of Psychology* (James 1890),[2] is central to the idea of perceptual gestalts (Kanizsa 1979), and also figures in the writing of David Marr e.g., "Vision is a process that produces from images of the external world a description that is useful to the viewer" (Marr 1982, p. 31).[3]

Although the *general* idea of perception as a constructive (inferential) process is widely accepted, substantial controversy surrounds the extent to which perception is influenced by factors such as goals, beliefs, desires, and expectations—factors outside of what is generally viewed as perception proper (or, in Pylyshyn's 1999 terminology, "early vision"—see below). The primary goal of this paper is to show that the controversy surrounding cognitive penetrability of perception (CPP)—the idea that perceptual processes are influenced by "non-perceptual" states—vanishes when we view perception not as a passive process the goal of which is recreation of a veridical reality, but rather as a flexible and task-dependent process of creating representations for guiding behavior. Insofar as higher-level cognitive states can *sometimes* be useful in making sense of the influx of sensory energy in the service of guiding behavior, it is manifestly adaptive for perceptual systems to be influenced by such higher-level states. Penetrable systems have the potential to be far smarter than impenetrable ones.

1.1 What is Cognitive Penetrability of Perception?

The definition of cognitive penetrability that I use here is Pylyshyn's (1999): A perceptual system is cognitively penetrable if "the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows" (1999, p.

---

[1] It is common, for instance, for my undergraduate students presented with evidence of linguistic effects on color perception to wonder how language can change the retina. This reaction wonderfully illustrates the implicit (and incorrect) assumption that perception involves sampling the retina and that changes to what we see must always track a change in input.

[2] James's influence can be glimpsed in Henri Bergson's claim that "Perception is never a mere contact of the mind with the object present; it is impregnated with memory-images which complete it as they interpret it" (Bergson 1911, p. 133).

[3] There is some irony in this because Marr's emphasis on vision as a bottom-up process heavily influenced opponents of the idea that perception is cognitively penetrable (Pylyshyn 1999).

343). Pylyshyn makes a bold and clear claim: “A major portion of vision, called the early vision system, does its job without the intervention of knowledge, beliefs or expectations, even when using that knowledge would prevent it from making errors.” (p. 414). On this view, a major part of perception works alone because it would be too slow/too laborious/too error-prone if memories, emotions, and expectations were allowed to impinge on the moment-to-moment workings of the perceptual apparatus (see also Fodor 1983, 1984).

In order to make use of perceptual information, the outputs of putatively encapsulated perceptual systems must be integrated with higher-level cognitive states. Thus those who oppose CPP accept that “Knowledge and expectations of course affect what we see something *as*” but contend that this “[happens] either in the attentional selection stage prior to the operation of early vision, or in the perceptual selection or decision stage after the operation of early vision.” (p. 414). In short, on a view like Pylyshyn’s, perception can be as constructive as one wants as long as the information used in the construction of perceptual states is restricted to the perceptual system.

My goal is to challenge the core assumptions on which this thesis rests and argue that there is no *in-principle* limit on the extent to which a given perceptual process can be penetrated by knowledge, expectations, beliefs, etc. The *actual* extent to which such penetrability happens can be understood in terms of whether it helps to lower system-wide (global) prediction error—a construct described below. I will also take on and attempt to defuse frequently used arguments against CPP from the persistence of illusions in light of contradictory beliefs, arguments about a distinction between diachronic and synchronic penetrability of perception, the relationship between attention and perception as it pertains to the CPP thesis, and the consequences of CPP for epistemology.

1.2 What is at Stake?

There is considerable interest in penetrability of perception both in philosophy (e.g., Deroy 2013; Lyons 2011; Macpherson 2012; Siegel 2012; Stokes 2011) and psychology (Firestone and Scholl 2014; Goldstone 1995; Hansen et al. 2006; Levin and Banaji 2006; Lupyan and Spivey 2008; Lupyan et al. 2009; Lupyan and Ward 2013).[4] Why should anyone care? First, the CPP thesis bears directly on the question of what to expect from theories of perception. If perception is cognitively penetrated then any comprehensive theory of perception needs to take into account how lower-level operations are influenced by higher-level cognitive constructs and opens the door to influences by culture and language (De Fockert et al. 2007; Henrich et al. 2010; Lupyan 2012a). Second, CPP threatens the distinction between observation and inference (Churchland 1988; Fodor 1984; Siegel 2012). A dissolution of the distinction between observation and inference in turn threatens theory neutral observation which is important to some epistemological frameworks (e.g., Fodor 1984).

---

[4] Levin and Banaji’s (2006) demonstration that Black faces appear darker than White faces (in particular Exp. 1) cannot be taken at face value owing to stimulus confounds. Specifically, although the photographic faces have the same mean brightness the Black face is psychophysically darker even when the faces are distorted so that they no longer look like faces (Lupyan and Lang, in prep; see also Firestone and Scholl, under review).

2 Perception as a Predictive Process; Perception as a Penetrable Process

If the goal of perception is to turn sensory energy into information useful for guiding behavior, the natural question is how this can be accomplished. It turns out that a very effective way of doing this is by attempting to *predict* the input. As argued by Andy Clark in his recent synthesis (Clark 2013), "the sheer breadth of application [of predictive coding] is striking. Essentially the same models here account for a variety of superficially disparate effects spanning perception, action, and attention" (p. 201).

The crux of this approach is that the system generates a prediction and then gets to verify it based on what happens later or what happens in a different part of the sensory field or modality. By continually seeking to lower the prediction error, the system forms representations[5] that encode the ecologically relevant aspects of sensory energy, in a task-dependent manner. Prediction comes in many guises—Bayesian inference, empirical estimation, and the free energy principle, to name a few (Friston 2010; Geisler and Kersten 2002; Howe et al. 2006; Purves et al. 2011; Rao and Ballard 1999; Yuille and Kersten 2006). These approaches, while philosophically distinct, can all be viewed as various ways of implementing Helmholtz's idea of perception as "unconscious inference."

In a hierarchical system that is the visual system, predictions are made and verified at multiple levels (Friston 2008). For example, we may have a high-level prediction that objects do not suddenly vanish into thin air. An event that violates this prediction, as observed in a magic trick, will generate a prediction error—what Clark (2013 Section 4.1) calls "agent-level surprise." If the belief that objects do not vanish is allowed to fully constrain what happens at the lower-levels of the visual system, then this error is avoided. Critically, however, the processing in lower levels of the perceptual hierarchy is also constrained by lower level (more local) predictions that would be violated if neurons in those layers "ignore" an object that suddenly disappeared. In short: the reason we see the magician's trick is that the solution producing the smallest *global* error is one that allows the lower-levels to respond roughly as they normally would and generate an error at the higher level that corresponds to agent-level surprise (and it is this error that coresponds to the knowledge that what you saw was out of the ordinary).

2.1 From Predictive Coding to Cognitive Penetrability of Perception

It is now possible to succinctly state the present thesis: Perceptual systems are penetrable to the extent that such penetration minimizes *global* prediction error. If allowing information from another modality, prior experience, expectations, knowledge, beliefs, etc., to influence perceptual processing lowers global prediction error, then this information will be used to guide processing at these lower levels. For example, if hearing a sound in certain circumstances can help disambiguate a visual input, then we should expect that sound to influence vision (e.g., Shams et al. 2002). If the predictions

---

[5] The term "representation" has varying definitions. My use is consistent with the use in contemporary cognitive psychology and cognitive neuroscience where it has come to denote an information-bearing state. Although typically applied to *neural* states, it is a promiscuous term that can be applied to, e.g., the information encoded in the immune system. Critically, my saying that something is "represented" should in no way be interpreted to mean that the information is explicit or implemented in a symbolic form.

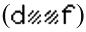
generated by hearing a word can provide the visual system with a prediction to help it interpret a signal that is too weak or noisy to be consciously perceived, then the right word should be able to make visible, what is otherwise invisible (Lupyan and Ward 2013; see Section 5).

If this "if it works, do it" attitude seems Panglossian, it is only by assuming that there is some extrinsic intelligence at work. There is no gatekeeper deciding how far down a cognitive state should penetrate perceptual processes. In evolving to minimize prediction error neural systems naturally end up incorporating whatever sources of knowledge, at whatever level, to lower global prediction error (Friston 2008; 2010; Clark 2013 for discussion). As I elaborate in the sections below, globally minimal error is sometimes achieved by allowing high-level states to penetrate lower-level processing. Other times, lower-level processing is left untouched and any conflicts between predictions and inputs can be resolved more effectively at higher-up "post-perceptual" levels.[6]

2.2 There is No Such Thing as Theory-Neutral Observation

One source of resistance to CPP has been the worry that infusing theories and knowledge into perception would somehow pollute "theory neutral observation", leading people to essentially see what they hope to see instead of what's "actually" out there (e.g., Fodor 1984). As discussed in more detail below (see esp. Section 6), this worry is misplaced. One reason is that there is actually no such thing as theory-neutral observation. Because sensory information is inherently ambiguous, only an inference-driven system can turn sensory signals into information useful for guiding behavior (e.g., Geisler and Kersten 2002; Purves et al. 2011; Yuille and Kersten 2006).

Where do inferences end and theories begin? I suggest that in the context of perception, this distinction is moot. So, for example, we can talk about the perceptual theories that govern perception of lightness. These theories take an inferential form: given the current intensity (brightness), estimates of ambient light, reflection of nearby surfaces (Kraft and Brainard 1999), 3D structure (Adelson 1993) and transparency/layer segmentation (Anderson and Winawer 2005), what is the lightness of the target object? Another example is reaching for an object that is hidden behind an occluder. Our reaching movements reflect a theory that objects do not cease to exist when they are out of sight and a prediction of its location based on the circumstances of its disappearance. As with the term "representation" (see footnote 5) the term "theory" should not be taken to mean something explicit or rule-based or symbolic.

A theory operating at a higher level may involve inferring that a form composed of meaningless lines that is surrounded by letters () corresponds to a letter that creates a meaningful word, and consequently *seeing* those meaningless forms as actual letters (Jordan et al. 1999). On the present view, the reason why word-knowledge influences our perception in this instance is that it lowers overall (global) prediction error resulting in a representation that is most likely given the input and prior beliefs/expectations.

---

[6] I will continue to use terms like "post-perceptual" because the term is descriptively useful, but I do not think it is possible to say where perception ends and cognition begins.

2.3 Prediction as a Way of Constructing Useful Representations

It is easy to see how prediction in its literal sense—knowing beforehand—is useful in perception, as when we need to know where a falling ball will land, or to anticipate the location of an occluded object. But focusing on the explicit goal of prediction risks overlooking perhaps its most powerful function: that of generating useful representations.

Consider language. Arguably, the *goal* of someone trying to understand a linguistic utterance is to understand what it means. Whereas in perception, the goal is sometimes explicitly to predict (e.g., when catching a ball), the explicit goal of predicting what someone will say is typically restricted to psycholinguistic experiments. Yet, language processing appears to reflect prediction and prediction all the way down. For example, that semantically inappropriate words generate what appear to be physiological error signals is generally taken to be a sign that a listener/reader is representing the likelihood of the occurrence of various words (DeLong et al. 2005; Kutas and Hillyard 1984). Although some error signals (e.g., the so-called N400 which signals semantic anomalies) occur post-perceptually, others operate at lower levels. For example, expecting a noun generates form-based *visual* predictions as indexed by a change in early visual processes when the expectation is violated (Dikker et al. 2009, 2010).

Why might one expect language processing to be predictive? Because making and testing prediction leads to the learning of representations that are useful in recovering (i.e., constructing) meaning from linguistic utterances (Elman and McClelland 1988; McClelland et al. 2006; Bybee and McClelland 2005). For example, by setting up a neural network to literally predict the next word, the network can learn rudimentary semantics because similar contexts lead to similar predictions yielding overlapping representations for semantically related words which tend to occur in similar contexts (Elman 2004). Indeed, *any* system whose function depends on forming rich representations may be aided by behaving in a predictive fashion.

2.4 Case Study of Prediction at Work: Cross-Modal Effects

In this section I briefly review one area in which the usefulness of a predictive framework can be plainly seen: understanding crossmodal effects. For example, seeing different mouth shapes causes us to hear physically identical speech sounds as entirely different (McGurk and MacDonald 1976). Why? Because our perception of speech sounds is under-determined by the input and it is advantageous to use whatever information is available to help constrain the system in making the correct inference. Thus, any sources that are highly predictive of a given speech sound should be incorporated and, indeed, affect what we hear. Consistent with this framing, speech perception has been found to be similarly altered by other highly predictive signals. For example, feeling a puff of air which is predictive of unvoiced but not voiced stop consonants can cause us to hear a /ba/ as a /pa/ (Gick and Derrick 2009). A theory that women and men have different voices means that speech processing can be further constrained (i.e., the posterior distribution sharpened) if the gender of the speaker is known. And it is (e.g., Strand 1999). The mere expectations that a speaker is Canadian leads Detroit residents to shift their perception of vowel boundaries (Niedzielski 1999; see also Hay et al. 2006).

Returning to vision: it is known that appropriately timed sounds can qualitatively change what one sees (e.g., Shams et al. 2002; Watanabe and Shimojo 2001). The literature on cross-modal effects commonly uses terms like “optimal cue integration” suggesting that modalities are somehow kept apart before being integrated with the integration happening downstream “true” perception. However, the loci of crossmodal effects are, at least in some cases, in classic modality-specific cortical regions (making the term “modality specific” questionable). Activity in what is typically thought of as “visual cortex” can be produced by audition (Bueti and Macaluso 2010), and touch (Blake et al. 2004). Conversely, activity in auditory cortex can be produced by visual signals (Calvert et al. 1997) and visual experiences rapidly modulate activity in primary somatosensory cortex (Taylor-Clarke et al. 2002). Whatever the neural loci of such effects, they have clear phenomenological consequences. In a recent confirmation of what Eric Schwitzgebel called the “Spelunker illusion (Schwitzgebel 2013), moving one’s hand in front of the face in complete darkness can produce visual percepts (Dieter et al. 2013). Touching a surface can change what it looks like (Meng and Zaidi 2011). Importantly, this influence of touch on vision is by no means automatic, but depends on the task-relevance of those movements (Beets et al. 2010).

It might be objected that such effects do not demonstrate CPP, but rather reflect automatic intra-perceptual modulation. For example, Pylyshyn speculates that perhaps the putative “early vision” system takes inputs from other modalities (Pylyshyn 1999; Section 7.1). However, the evidence above suggests not a fixed, reflexive system where by some inexplicable quirk vision can be modulated by sound or touch, but rather a highly flexible system where these modulations can be explained as ways of lowering the overall prediction error.

If what one sees depends on what one hears or touches, or how one moves one’s arms, how can we ever draw the line between influences of “true” theories versus just perceptual theories? I suggest that we cannot because they are but different flavors of inferences at different levels of abstraction.

3 Penetrability of Visual Processing by Categorical Knowledge and Language

Almost a century ago, Edward Sapir remarked that “Even comparatively simple acts of perception are very much more at the mercy of the social patterns called words than we might suppose (Sapir 1929, p. 210). Why might one expect words to affect visual processing and how far “down” do such effects go? Consider that language is enormously effective at conveying information useful for basic survival. It sure would be useful if hearing “watch out for the brick!” rapidly potentiated visual processes (via top-down feedback) to enable faster detection of a reddish rectangular object hurling our way. How can this happen? Among other things, learning the word “brick” involves learning associations between the word and actual bricks which happen to have certain features that are useful in distinguishing bricks from non-bricks. For example, bricks have a typical size, color, we expect bricks to form surfaces perpendicular to the ground, etc. These expectations are actively used in visual prediction (Estes et al. 2008; Oliva and Torralba

2007). In learning the word "brick" the label becomes more strongly associated with features diagnostic of bricks and dissociated from features that occur only incidentally and are not predictive of brick-ness (Lupyan 2012b). With such associations in place, activation of the label—which can occur during language comprehension or production—provides top-down activation of the associated visual properties.[7]

Although visual features corresponding to brick-ness can be activated in a variety of non-linguistic ways, labels appear to be especially effective in activating visual aspects of the label's referent in a way that allows discrimination of category members from non-members (Lupyan and Spivey 2010a; Lupyan and Thompson-Schill 2012). For instance, compared to hearing an equally informative non-verbal cue like a barking sound, hearing the word "dog" allows people both to recognize dogs more quickly and to distinguish more effectively between a canonical upright and an atypical upside-down dog (Lupyan and Thompson-Schill 2012) than hearing equally informative non-verbal cues.

Of course, object recognition and high-level discrimination (is the upright-object on the right or left?) although obviously requiring visual processing may lie outside the purview of what Pylyshyn would count as "early vision." But such effects of language on visual processing can be found at virtually every level. In a task requiring identifying the direction of moving dots, performance is made worse by hearing direction-incongruent verbs (Meteyard et al. 2007), and simply imagining or hearing a story about motion can produce motion aftereffects (Dils and Boroditsky 2010). Conversely, lexical decision times for motion words increase in the presence of congruent visual motion and decrease with incongruent motion (Meteyard et al. 2008), suggesting bidirectional influences between linguistic and visual-motion processing. Similar demonstrations have been made in the domain of contrast-sensitivity (Pelekanos and Moutoussis 2011) and face-processing (Landau et al. 2010). Newly learned information about a novel objects' functions, as well as simply learning a name for it, modulates early visual processing during subsequent passive viewing (Abdel Rahman and Sommer 2008).[8]

Another demonstration of how higher-level conceptual categories influence perceptual processing comes from a series of studies examining the role that familiar categories have on perceptual processing. These studies took advantage of a convenient dissociation between the visual and conceptual properties of the letters B, b, and p. The pairs B-b and B-p are visually equidistant,[9] but the pair B-b is more conceptually similar (i.e., both letters are members of the same class) than B-p. When tasked with performing speeded same-different judgments of physical identity (i.e., B-B=same; B-p and B-b=different), participants' judgments are *initially* equally fast for the within-category (B-b) and between-category (B-p) trials (Lupyan 2008 Exp. 2; Lupyan et al. 2010). A category-effect, measured by the RT difference between B-p and B-b stimuli emerged, however, when delays (150-600 ms.) were introduced between the presentation of the first and second letter in the pair with the first letter always visible (Lupyan

---

[7] Arguably, this activation of perceptual features *is* word recognition, but that is a subject for another paper (Lupyan and Bergen forthcoming).

[8] One explanation is that higher-level knowledge helps to integrate otherwise unrelated visual parts into coherent wholes (Lupyan and Spivey 2008).

[9] The pairs are made to be exactly equidistant by editing 'B' to have identical upper and lower loops (called 'counters' in typography jargon).

et al. 2010). These results reveal a gradually unfolding influence of a higher-level category representation on a lower-level perceptual discrimination task. One interpretation of this finding is that during the delay the perceptual representation of the first item was augmented through top-down feedback by its conceptual category with the effect that the subsequently appearing second item was now closer in representational space when it was in the same conceptual category. This account finds further support in a subsequently conducted fMRI study that showed greater representational similarity between B and b compared to B and p in extrastriate visual cortex (Lee et al. 2013).

Further demonstrations of interactions between knowledge and on-line visual processing comes from a series of studies using visual search. In a typical search task, participants are presented with an array of “distractors” among which there is occasionally a target item, the identity of which is either known ahead of time or defined by it being a “singleton” i.e., being a unique item, e.g., a single red circle among numerous green circles. It is not altogether surprising that familiar stimuli are easier to process than unfamiliar stimuli. For example, it is easier to search for a 2 among 5s than a ⊓⊔ among ⊔⊓s (Wang et al. 1994). This difference may be due to people having much more experience with the former characters as numbers, allowing them to build up over time more “robust” bottom-up representations (a kind of perceptual learning). Alternatively, what appears to be a familiarity advantage may in fact be a *meaningfulness* advantage—an instance of top-down object knowledge facilitating bottom-up processing. Indeed, in the absence of any training, simply *telling* participants that ⊓⊔ and ⊔⊓ should be thought of as rotated 2s and 5s dramatically improved visual search performance leading to search efficiency that was very near that of conventionally oriented numerals (Lupyan and Spivey 2008; Lupyan 2007; see also Smilek et al. 2006). The finding that meaningfulness affected not just overall RTs but search *efficiency* (the rate at which RTs increase as a function of the number of elements) rules out an explanation based on differences in a higher-level decision process.

Not surprisingly, in visual search tasks it helps to know ahead of time what the target looks like insofar as it helps to preactivate appropriate feature detectors (Vickery et al. 2005; Wolfe et al. 2003)—an instance of knowledge “greasing the wheels of perception”. But even with full knowledge, simply hearing a verbal label can further potentiate visual processing. For example, over and above knowing the identity of a search target (e.g., the letter “B”) hearing “bee” produces a transient facilitation when searching for a B among p’s (Lupyan 2008). Hearing similarly redundant labels also facilitates “category-based attention” which allows for more efficient processing of items from a given category (e.g., letter Bs, chairs) throughout the visual field (Lupyan and Spivey 2010b). These results are consistent with the hypothesis that processing a word activates corresponding visual representations systematically changing how incoming sensory information is processed (Lupyan 2012b for review).

Just how far “down” can language affect visual processing? The most fundamental visual task of all is simple detection. Lupyan and Ward (2013) asked participants to simply detect stimuli, responding “yes” if a stimulus—any stimulus—was present, and “no” otherwise. The stimuli to which participants had to respond were rendered invisible using continuous flash-suppression (CFS) paradigm. CFS, a variant of binocular rivalry, is known to suppress visual representations at a low level, for instance substantially reducing negative afterimages (Tsuchiya and Koch 2005), and wiping out much of the ventral stream processing associated with object categorization (e.g., Kreiman et al. 2002; Pasley et al. 2004; Sheinberg and Logothetis 1997). The results showed that simply hearing a word was sufficient to unsuppress and make visible otherwise invisible images. For example, hearing the word “zebra” enabled subjects to see a zebra that was otherwise invisible. Importantly, the effect required a match between the label and the stimulus. Labels had no effect on trials in which there was no image (hearing “zebra” did not lead people to hallucinate zebras and actually caused a further suppression when they mismatched, i.e., hearing “zebra” made it more difficult to see a pumpkin (see also Lupyan and Spivey 2010a).

Fodor’s (1983) argument of the modularity of mind—a thesis that many contemporary critics of CPP seemingly accept—focused on evidence from two domains: language, and vision. Thirty years later we know that not only are vision and language penetrable, they penetrate one another.

4 Beyond Vision: CPP in Haptics and Gustation

The debates regarding CPP have largely focused on vision. This is not surprising given the relative importance of vision in human perceptual phenomenology. But nothing about CPP limits it to vision, making it instructive to briefly consider evidence from other modalities.

4.1 Haptics: The Psychophysics of Weight Perception

Small objects feel heavier than larger objects of equal mass (Ross 1969). One somewhat simplified explanation of this so-called size-weight illusion is that because in the real world larger objects *are* heavier, we form an expectation that a larger object will be heavier than a smaller object. The violation of that expectation (by the devious experimenter who sets up a situation in which weight is uncorrelated with size) leads to a contrast effect. Although at first glance this appears to be an instance of a higher-cognitive state (expectations of weight) influencing a perceptual state (perceived weight, applied lifting force, etc.), an alternative explanation is that our apparently unitary percept of weight actually reflects the muscular inputs required to lift and handle the object which vary not just by weight but by distribution of mass, rotational inertia, etc. (Amazeen and Turvey 1996). If true, then although the size-weight illusion shows that apparently simple percepts like weight are actually produced through multi-sensory integration, it does not constitute evidence of an influence by a cognitive state as such.

Whether pure expectations alone can change weight perception was unknown until recently. Buckingham and Goodale (2010) showed participants blocks of varying sizes. Participants were wearing shutter-glasses which only allowed the block to be briefly visible. After viewing a large or small block for 1 s, the experimenter replaced it surreptitiously with a block of intermediate size. Although the block being lifted had constant size and weight, participants’ weight perception and the force they applied was systematically influenced by the size of the block they expected to lift. The block being lifted was always identical hence required all the same muscular forces. The authors conclude that "expectations of heaviness are not only powerful enough to alter the perception of a single object's weight, but also continually drive the forces we use to lift the object."

If such weight-size illusions are sensory adaptations to real-world structure, then one may expect the system to maintain some degree of flexibility. Indeed, as shown by Flanagan et al. (2008), the "fundamental expectation that weight increases with size can be altered by experience and neither is hard-wired nor becomes crystallized during development." (p. 1742). Interestingly, the authors found that although providing experience with objects with a reversed size-weight relationship affected both the lifting forces applied to the objects and subjective weight estimates, the two adapted at very different rates. One part of the system (responsible for lifting forces) adapts quickly because, the authors reasoned, its estimates are tuned to specific objects, whereas another part (responsible for weight perception) adapts more slowly because it is tuned to entire families of objects. Overall, we see in this work clear evidence of a system that allows for flexible retuning based both on moment-to-moment cognitive expectations and longer term experience with specific classes of objects. This ability to be guided by higher-level states presumably makes the system more adaptive in the real world.

4.2 Gustation: The Taste of Expectation

Imagine a mundane morning routine. You pour yourself a cup of orange juice and take a sip. Now, a variation: you pour yourself a cup or orange juice which someone then stealthily replaces with milk. Not noticing the switch, you take a sip. I have not run this experiment, but I am taking bets on its outcome. The milk will not—at least initially—taste like milk. Expectations are known to affect initial neural coding of taste (Samuelsen et al. 2012) and smell (Zelano et al. 2011), and some have argued that in these domains expectations are the primary determinant of what animals taste and smell. The "mere" expectation of what one is about to taste is sufficient to create neural activity specific to the expected taste in the gustatory cortex (Fontanini and Katz 2008 for review). Such results strongly suggest that bottom-up gustatory inputs are evaluated within the background of expectations. Could such dependence on expectations lead to the sort of epistemic disaster that some philosophers worry about (Section 6)? No, and the reason is revealed by considering a third orange juice and milk scenario: You are blindfolded and handed what you are told is *either* a glass of milk or a glass of orange juice. You take a sip. It is orange juice. The phenomenology (I posit) and neural coding (as is suggested by similar studies, Veldhuizen et al. 2011) will again be different from the case when one tastes orange juice while expecting it. In this latter case, the possibility is left open: it might be orange juice, or it might be milk and the incoming information is now appropriately processed within this greater state of uncertainty. Once again, our perception is altered by our expectations, but not arbitrarily so. By being interpreted *within the priors established by the expectation* we are able to more accurately turn sensory energy (in this case chemical) into information useful for guiding behavior.

5 Why Illusions Do Not Defeat CPP and Why Attentional Effects Should Not be Ruled Out as Cases of CPP

5.1 Argument from the Persistence of Illusions: Müller-Lyer and Beyond

One argument made by opponents of CPP concerns the persistence of illusions in the face of contradictory beliefs. A common example is the Müller-Lyer illusion (e.g., Pylyshyn 1999). In this illusion, a line that is adorned with arrow-heads looks shorter than a line adorned with arrow-tails despite the two lines being equal in length. The illusion persists even when the viewer knows that the lines are objectively equal. That people continue to see the line adorned with arrowheads as shorter in the face of the (correct) belief to the contrary is then used as evidence against CPP. I contend that this argument from the persistence of illusions unravels when we consider illusions and the putative influence of higher-level states within a framework of predictive coding.

To see why the argument from the persistence of illusions fails, one needs to understand why illusions such as the Müller-Lyer arise in the first place. The most widely cited explanation for the Müller-Lyer illusion attempts to explain it in terms of size constancy (Gregory 1963). Arrow-head adorned lines represent surfaces that are convex and thus closer than arrow-tail adorned lines that are concave and farther away. An object that is physically closer and subtends visual angle θ must be objectively smaller than an object that subtends θ but is farther away. This explanation turns out to be wrong (e.g., the illusion persists when arrows are replaced by circles, or ovals, or squares—none of which bear any resemblance to corners). An alternative is offered by Howe and Purves (2005) who empirically measured distances in a large collection of images of surfaces that have "adornments" of various forms and found that surface with adornments closer to their center, corresponding to lines with arrow-heads configuration are *actually* shorter than surfaces with adornments that are farther from the center. Thus when viewed in a broader context of correctly estimating distal sources (in the service of promoting adaptive behavior), the Müller-Lyer is not an illusion at all—indeed Purves et al. have convincingly argued that the very concept of illusions is erroneous because calling them illusions implies the goal of veridical estimation of the current stimulus rather than the *globally* optimal estimation that perceptual systems appear to engage in. Illusions on this view can be explained as "optimal percepts" (Weiss et al. 2002). The empirical estimation approach advocated by Purves et al. has been quite successful, explaining, in addition to the Müller-Lyer, the empirical basis of the Ebbinghaus Illusion (Howe and Purves 2004), Poggendorff illusion (Howe et al. 2005), and the Vertical-Horizontal illusion—the finding that vertical lines are perceived to be about 10 % longer than physically equal horizontal lines (Howe and Purves 2002).

All these illusions can be understood as operating in the service of minimizing global prediction error. Seeing vertical lines as being longer than horizontal lines (or perceiving an object as heavier simply due to expecting it to be a certain size) may appear to be maladaptive, but only insofar as we expect our perceptual systems to be certified by the Office of Weights and Measures. Rather than failures, the illusory percepts are compromises, representing *globally* optimal solutions.

An immediate response is: shouldn't it be adaptive to see the lines as truly equal given that *in this instance* they actually are? Indeed! But on no theory should this happen simply by someone telling you "Guess what? these lines are actually equal." In the case of the Müller-Lyer illusion, the prior that real-world lines in these configurations is arguably too strong, while the bottom-up input is too uncertain (you need to put a ruler to it!). However, just a bit of extra evidence weakens the illusion or makes it go away entirely (Parker and Newbigging 1963; Rudel and Teuber 1963; Selkin and Wertheimer 1957). All that's needed on the present view, is some evidence that breaks the association between the lines and real-world distal sources. Critics of CPP concede that long-term experience can allow perception to be altered by higher-level states (Fodor 1984; Pylyshyn 1999; McCauley and Henrich 2006 for discussion). This concession is often justified by positing that this kind of diachronic change (often referred to as "perceptual learning") works by altering bottom-up sensory processing providing the visual system with effectively different inputs before versus after learning. Vast experience in some domain such as letter recognition may indeed transform bottom-up processing and lead to specialized early representations (e.g., as argued by Cohen et al. 2002). But the sorts of training effects cited above fall far short of this. Performing a 100 trials worth of length estimates does not permanently warp some part of visual cortex. The changes that happen with such practice are at higher-levels (Ahissar and Hochstein 2004; Hochstein and Ahissar 2002), yet they can influence what we see.

To put it plainly: insofar as the Müller-Lyer illusion arises from the visual system attempting to represent likely real-world sources, it would be maladaptive to undo one "illusion" while *breaking the rest of vision* in the process. A bit of additional evidence in the form of training allows the system to reach a globally optimal state, making accurate local predictions while maintaining globally optimal performance.

A related argument from persistence of illusions is made by Chaz Firestone (pers. comm.) in reference to illusions such as the long-armed soccer player (Fig. 1-left). Our knowledge that people do not have six-foot long arms, does not shrink the arm or otherwise alter it to make the scene fit with our theory. Firestone's reasoning goes something like this: That we see the soccer player as having an abnormally long arm is a consequence of the visual system automatically running automatic cognitively impenetrable algorithms that output representations based on e.g., a continuity-of-form gestalt. These outputs are then fed to post-perceptual systems which, knowing what they know about bodies, interpret the percept as absurd.

Fig. 1 Illusions where low-level and high-level theories come in conflict. *Left*: The long-armed soccer player. *Right*: The Abercrombie cat

The long-armed soccer player is a better example than the Müller-Lyer because there is less uncertainty in the bottom-up input (No rulers required!). But this version of the argument fails for the same reason that we can perceive a magician making an object disappear despite knowing that this is not actually possible. In both cases the prediction error is more effectively resolved at the higher-level because the input has such low uncertainty (see Section 2).

Although one does not need a ruler to see that there is something unusual about the soccer player, consider how we would perceive the arrangement depicted in the picture in the realworld (provided we could see it in slow-motion). As all perceptual information, this scene would unfold in time. On seeing this arrangement, the most likely high-level inference is that of coincidental alignment of bodies. This leads to the prediction that the arrangement should go away as the players move. If it doesn't, we would (1) be surprised and (2) consider revising our belief that people do not have 6-foot arms. This point can illustrated more clearly by the Abercrombie cat shown in Fig. 1-right. Here, too, high-level knowledge (there are no cat-human hybrids) appears to not affect what we see. The percept can be explained by there being a cat in the bag. Altering low-level vision does nothing for lowering the global prediction error. Despite the apparent phenomenological certainty in both of these cases, we still predict that the fortuitous arm-arrangement will disappear (Fig. 1-left) and that if we were to lift our head just a little bit, the cat-hybrid percept would vanish (Fig. 1-right). No problem is solved in these cases by altering lower-level percepts.

Allowing high-level states to automatically transform low-level vision is indeed maladaptive, but not because vision works better or faster when it works alone, as assumed by e.g., Fodor (1983, 1984), but because we need to consider all sources of error. This focus on minimizing global prediction error means that we cannot make any general statements about which level of information will "win" when put in conflict (as in Fig. 1). When the low-level signal is of high quality (low uncertainty), it may dominate low-level processing but be discounted at a higher-level. Viewing the images does not lead us to believe that some soccer-players have 6-foot arms or that there are cats with human bodies. When the low-level information is more ambiguous, high-level knowledge can dominate phenomenology. For example, knowledge of body anatomy can, in fact, override low-level depth information from stereoscopic cues (Bulthoff et al. 1998).

5.2 Attention as a Change in Input: Being Misled by a Metaphor

As is now well-known, attention modulates processing at all levels of the visual hierarchy (Ghose and Maunsell 2002; Jack et al. 2006; O'Connor et al. 2002). *Prima facie*, these findings appear to be devastating for opponents of the CPP thesis because attention clearly counts as a cognitive process and its effects clearly change what we perceive (Störmer et al. 2009). Pylyshyn exempts attention from counting as evidence against CPP because "the allocation of attention to certain locations or certain properties [occurs] *prior to* the operation of early vision" (1999, p. 344). On this view, if attending to something makes it look brighter (Carrasco et al. 2004) it is equivalent to sending more of the signal into the visual system. The perceptual apparatus then proceeds to process this input exactly as it would if presented with an actually brighter image. Counting attention as an instance of CPP would, on this view, be like saying that closing one's eyes and cutting off visual input is also an instance of CPP (Fodor 1988; Brian Scholl, pers. comm.).

However, the claim that a perceptual change due to attention is just like a change in the input constitutes a reasoning error. That *X* causes *N,* and *Y* causes *N,* does not imply that *X*= *Y.* Just because some consequences of attentional shifts are similar to some consequences of physical changes to the stimulus does not mean that they are equivalent.

Here is an example of this error in action. Viewing a colored surface rapidly leads to chromatic adaptation such that (under the right circumstances) an observer will subsequently see an afterimage in the complementary color. Not surprisingly, starting at a more vivid adaptor leads to a more vivid afterimage. On the view of attention as a straightforward change of input, one would predict that attending to an adaptor would also lead to a more vivid afterimage. In fact attention does modulates intensity and duration of afterimages, but in the *opposite* direction. Attending to the adaptor weakens the resulting afterimage (Suzuki and Grabowecky 2003; Van Boxtel et al. 2010). Discussing the mechanisms responsible for this effect is not important here, but the results show the problem with assuming that attention is equivalent to a physical change in the input.

The very idea that attention is a change of input appears to rest on what has been for decades an enormously powerful metaphor: attention as spotlight. As discussed by Chun and Wolfe: "The spotlight has been a favorite metaphor for spatial attention because it captures some of the introspective phenomenology of attention – the feeling that attention can be deployed, like a beam of mental light, to reveal what was hidden in the world." (Chun and Wolfe 2001). Once one thinks of attending as shining a spotlight, it is a short leap to thinking of attention as being equivalent to literally illuminating an object[10]

Although the attentional spotlight metaphor continues to be useful in generating data (How fast? How big? Can you zoom it? Can there be two?), its conceptual underpinnings are not consistent with what is known about spatial attention (Cave and Bichot 1999) and do not accommodate the many varieties of non-spatial attention (Ansorge and Becker 2011; Egly et al. 1994; Lupyan and Spivey 2010b; Moore and Egeth 1998).

If attention is not like a change in input, how should we think about the relationship between attention and perception? A promising alternative is that attention is a surprise-reducing mechanism (Anderson 2011; Itti and Baldi 2009) that allows organisms to fine-tune on-line predictions for particular tasks (Clark 2013 for further discussion). On this view, attention is the process by which incoming sensory energy is transformed into a useful form *for a particular perceptual goal.*[11]

---

[10] The spotlight metaphor also harkens to extramissive theories of vision—that we see by emitting rays from our eyes—something that 13–67 % of adults endorse depending on the question! (Winer et al. 2002).

[11] Viewing attention in this way also avoids another common trope—often explicitly articulated at the start of talks and papers—that attentional mechanisms exist because "Reflected light carries too much information for our visual system to process at once" (Franconeri et al. 2005). Such a view implies that if only humans had higher processing capacities, there would be no need for attention. The fallacy of this argument can be seen by taking it to its extreme: If humans with their relatively large memories and sensory processing capacities are in need of attention to cope with all that information out there, then consider how much attention an earthworm would need to make sense of all that information! Responding that the perceptual system of worms are far simpler and do not have this problem, begs the question. Why should our own perception outstrip our processing capacities by such a large margin? On the present view, attention is needed because the optimal perceptual representation for one task is suboptimal for another. Attention is the process by which perceptual represntations are transformed to make them more useful for guiding for a specific task.

For example, consider viewing a scene with some cars and pedestrians. How might the perceptual representation of that scene change when we attend to the pedestrians versus attending to the cars? Cars and people differ visually in various ways but one salient difference is that people tend to be vertically oriented and cars horizontally oriented. (In fact the very same cloud of pixels can look like a car when horizontally oriented and like a person when it is vertically oriented, Oliva and Torralba 2007). An effective way of transforming a perceptual representation in the service of attending to people therefore might be to accentuate processing of vertically oriented features. The vague phrase "accentuating processing" corresponds in a predictive-coding framework to using the prior (pedestrians are likely to be vertically-oriented) to more accurately model the posterior distribution corresponding to locations in the scene with high-power in vertical orientations. Some existing behavioral evidence is consistent with the conclusion that attending to people accentuates processing of vertical features while attending to cars accentuates the processing of horizontal features (Oliva and Torralba 2006).

There is, in theory, no limit to the complexity of perceptual transformations that such attentional modulations can achieve, and indeed, people are highly effective in attending to arbitrarily high-level concepts such as images that depict birthday parties (Potter and Fox 2009; Potter 1975). Viewed this way, attention functions through continued modulation of perceptual states (Çukur et al. 2013; Ghose and Maunsell 2002; Jack et al. 2006) and calling an effect "merely" attentional as distinct from "truly" perceptual has little empirical basis. In short: there is lots of evidence that attention involves modulation of perceptual states and no solid evidence that attention is something that happens before "genuine" perceptual processing. Hence, there is no a priori reason to rule out attentional effects as evidence of CPP.

6 Consequences for Epistemology: Will Cognitive Penetrability Kill You?

A concern of allowing perception to be influenced by higher level states, that has been voiced by some philosophers, is that doing so would threaten the distinction between observation and inference. Fodor's "granny" narrator, for example, is "particularly aroused about people playing fast and loose with the observation/inference distinction … We may not have prayers in the public schools, [but] by G-d we will have a distinction between observation and inference" (Fodor 1984). Fodor's concern (one that is shared by Pylyshyn) is that seeing what we expect seems to defeat the purpose of vision: "[an organism] generally sees what's there, not what it wants or expects to be there. Organisms that don't do so become deceased" (Fodor 1983).

Siegel (2012, 2013) articulates a similar concern. Suppose Jill believes Jack is angry at her and that this belief makes Jack's (neutral) face look actually angry. This percept would then provide Jill with additional evidence that Jack is angry. This creates a circularity from Jill's point of view such that Jill starts out with "penetrating belief, and end[s] up with the same belief, via having an experience." (2012, p. 202).

On the present view, these worries may be unwarranted. First, just as percepts are not random, but depend in a principled way on the outside world, beliefs too are not random, but are constrained by an organism's history of interactions (see Lyons 2011 for a similar argument). Suppose I am walking through the jungle, trying not to get eaten. My perceptual systems are doing their best to transform the sensory input into a form that best distinguishes between things that can eat me and those that can't (while simultaneously keeping me from tripping, ensure I don't run into a tree, etc.). Fodor assumes that allowing expectations to influence this process would make it more likely that I get eaten. But that assumes that my expectations are unconstrained. In reality, my knowledge that e.g., leopards have stripes, that snakes slither, that certain sounds presage being pounced on are manifestly useful in *improving* detection, discrimination, and categorization.

Similarly, if Jill confabulates Jack's angry appearance due to a belief that is not supported by any evidence, this would be bad news indeed. But more likely is that Jill detects something about Jack's posture or tone of voice that leads her to believe that Jack is angry. Just as hearing "pumpkin" did not lead people to confabulate pumpkins, but rather improved the observers' ability to detect pumpkins (Lupyan and Ward 2013), simply believing or expecting Jack to be angry should not cause him to look angry. A hierarchical predictive system may respond to a belief of Jack's anger by, e.g., sharpening the representation of facial features that distinguish angry from non-angry faces thereby improving the likelihood of detecting anger (insofar as anger is something that can be accurately ascertained from the face).

Of course, beliefs and expectations are sometimes wrong. If, while I am on my dangerous jungle walk, someone tells me that in these regions leopards are actually purple, my survival might be threatened by expecting purple things and thereby missing the orange leopard that I otherwise would be able to spot. Similarly, being told that someone is angry at us can cause us to mis-judge their actual emotional state by initiating an active search to confirm our hypothesis (confirmation bias). But so it goes.

6.1 Epistemic Downgrading and CPP

We do not believe everything we see. To take an extreme case, individuals with Charles-Bonnet syndrome can experience vivid hallucinations of e.g., floating disembodied heads and colorfully clothed dwarves without actually believing that such things exist (Schultz and Meizack 1991). Ingesting certain substances may cause one to see (really see!) the walls moving, but it does not, at least in the short-term, lead to our abandoning the belief in the stability of walls. On the present thesis, this is because the link between observation and belief revision depends on the certainty of the perceptual input as well as the certainty of the belief.

How much a percept is to be epistemically privileged depends on a host of both bottom-up factors (e.g., how clearly is it perceived?) and top-down factors (e.g., does it violate previously held theories? did I take drugs?). Moreover, the extent to which a higher-level state should influence lower-level representations depends on the *content* of those lower-level representations. So, the reason referring to ⊔⊓ and ⊓⊔ as rotated numbers immediately improves visual search (Lupyan and Spivey 2008) is that doing so allows the visual system to code arbitrary collection of lines in terms of larger units thereby improving discrimination. But this improvement depends on our having prior higher-level categories (2 and 5) into which the novel symbols could be sensibly incorporated; referring to the symbols as "persimmons" and "koalas" would not and should not cause us to see them as such. Similarly, the reason that hearing "pumpkin" makes an otherwise invisible pumpkin, visible, but does not cause people (at all!) to confabulate pumpkins where there aren't any is that in the former case the pumpkin hypothesis is supported by bottom-up evidence which is too weak/fragmented in this case to lead to conscious perception on its own (Lupyan and Ward 2013).[12] To reiterate: the way in which expectations are theorized to affect perception (at least on the present view) is *not* through a simple bias. The modulations of perceptual states by cognitive states are not confabulations or wishful seeing and so there is no danger in epistemically privileging them.

It is worth noting that the precise way in which knowledge augments perception depends the details of the situation. For example, the first time we see a heat mirage we may genuinely mistake it for a body of water. Later, even though we may know there is no water there, we may still see mirages. This kind of top-down override is too drastic given the quality of the input that the mirage constitutes. However, knowledge of mirages may sensitize us to noticing, e.g., how the mirage moves with respect to our perspective in a way that real water would not, helping to distinguish it from real water.

7 Conclusion

The goal of perception is to convert sensory energy into information that can guide behavior. An especially effective way to achieve this goal is to have the perceptual system to engage in prediction. In so doing, *all* sources of information both low and high-level are *potentially* useful. The empirical evidence reviewed throughout this paper shows just how pervasive penetrability of perception is.

The degree to which high-level predictions are used to adjust lower-level representations depends on the circumstances. To the extent that higher-level states can lead to more accurate lower-level predictions, they will influence the corresponding lower-level processes. If a mismatch between bottom-up sensory data and top-down predictions is more effectively resolved at a higher level, then conflicts can be resolved at these higher levels leaving lower-level representations less affected. No one needs to "decide" where the influence of a higher-level state should terminate. This is an emergent property of a hierarchical predictive system, examples of simply performing computations that reduce global prediction error, much as water finds the shortest path downhill.

I have argued that the predictive coding framework can provide a unified way of understanding where we should and shouldn't expect penetrability and of what sort. I have also attempted to defuse the arguments frequently used by critics of CPP—namely, the argument from the persistence of illusions, and distinctions between effects of perceptual learning and attention from "true" penetrability. Finally, I have argued that perception is penetrable because penetrability is adaptive. Penetrable perceptual systems are simply better and smarter in fulfilling their function of guiding behavior in a world that is at once changing and predictable.

---

[12] However, if the cost of a false alarm is much lower than the cost of a miss, a simple change in bias may be sufficient. For example, being thirsty appears to bias people to see things as being more transparent possibly because seeing water where there is none is better than missing water entirely (Changizi and Hall 2001).

Acknowledgments Preparation of this manuscript was in part supported by NSF#1331293 to G.L. I thank Emily Ward, Chaz Firestone, Zoe Jenkin, and Jonathan Lang for useful discussion.

References

Abdel Rahman, R., and W. Sommer. 2008. Seeing what we know and understand: how knowledge shapes perception. *Psychonomic Bulletin & Review* 15(6): 1055–1063. doi:10.3758/PBR.15.6.1055.

Adelson, E.H. 1993. Perceptual organization and the judgment of brightness. *Science* 262(5142): 2042–2044.

Ahissar, M., and S. Hochstein. 2004. The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences* 8(10): 457–464.

Amazeen, E.L., and M.T. Turvey. 1996. Weight perception and the haptic size–weight illusion are functions of the inertia tensor. *Journal of Experimental Psychology: Human Perception and Performance* 22(1): 213–232. doi:10.1037/0096-1523.22.1.213.

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

Anderson, B.L., and J. Winawer. 2005. Image segmentation and lightness perception. *Nature* 434(7029): 79–83. doi:10.1038/nature03271.

Ansorge, U., and S.I. Becker. 2011. Automatic priming of attentional control by relevant colors. *Attention, Perception, & Psychophysics*.. doi:10.3758/s13414-011-0231-6.

Barlow, H. 1990. Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research* 30(11): 1561–1571.

Beets, I.A.M., B.M. 't Hart, F. Rösler, D.Y.P. Henriques, W. Einhäuser, and K. Fiehler. 2010. Online action-to-perception transfer: only percept-dependent action affects perception. *Vision Research* 50(24): 2633–2641. doi:10.1016/j.visres.2010.10.004.

Bergson, H. 1911. *Matter and memory.* Cambridge: The MIT Press.

Blake, R., K.V. Sobel, and T.W. James. 2004. Neural synergy between kinetic vision and touch. *Psychological Science* 15(6): 397–402. doi:10.1111/j.0956-7976.2004.00691.x.

Buckingham, G., and M.A. Goodale. 2010. Lifting without seeing: the role of vision in perceiving and acting upon the size weight illusion. *PLoS ONE* 5(3), e9709. doi:10.1371/journal.pone.0009709.

Bueti, D., and E. Macaluso. 2010. Auditory temporal expectations modulate activity in visual cortex. *NeuroImage* 51(3): 1168–1183. doi:10.1016/j.neuroimage.2010.03.023.

Bulthoff, I., H. Bulthoff, and P. Sinha. 1998. Top-down influences on stereoscopic depth-perception. *Nature Neuroscience* 1(3): 254–257.

Bybee, J., and McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, *22*(2–4). doi:10.1515/tlir.2005.22.2-4.381.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., … David, A. S. 1997. Activation of Auditory Cortex During Silent Lipreading. *Science*, *276*(5312): 593–596. doi:10.1126/science.276.5312.593.

Carrasco, M., S. Ling, and S. Read. 2004. Attention alters appearance. *Nature Neuroscience* 7(3): 308–313.

Cave, K.R., and N.P. Bichot. 1999. Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin & Review* 6(2): 204–223.

Changizi, M.A., and W.G. Hall. 2001. Thirst modulates a perception. *Perception* 30(12): 1489–1497. doi:10.1068/p3266.

Chun, M.M., and J.M. Wolfe. 2001. Visual Attention. In *Blackwell handbook of perception*, ed. B. Goldstein, 272–310. Oxford, UK: Blackwell Publishers Ltd.

Churchland, P.M. 1988. Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor. *Philosophy of Science* 55(June): 167–87.

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

Cohen, L., S. Lehéricy, F. Chochon, C. Lemer, S. Rivaud, and S. Dehaene. 2002. Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain* 125(5): 1054–1069. doi:10.1093/brain/awf094.

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

De Fockert, J., J. Davidoff, J. Fagot, C. Parron, and J. Goldstein. 2007. More accurate size contrast judgments in the Ebbinghaus illusion by a remote culture. *Journal of Experimental Psychology, Human Perception and Performance* 33(3): 738–742. doi:10.1037/0096-1523.33.3.738.

DeLong, K.A., T.P. Urbach, and M. Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8): 1117–1121. doi:10.1038/nn1504.

Deroy, O. 2013. Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies* 162(1): 87–107. doi:10.1007/s11098-012-9989-1.

Dieter, K.C., B. Hu, D.C. Knill, R. Blake, and D. Tadin. 2013. Kinesthesis can make an invisible hand visible. *Psychological Science*. doi:10.1177/0956797613497968.

Dikker, S., H. Rabagliati, and L. Pylkkänen. 2009. Sensitivity to syntax in visual cortex. *Cognition* 110(3): 293–321. doi:10.1016/j.cognition.2008.09.008.

Dikker, S., H. Rabagliati, T.A. Farmer, and L. Pylkkänen. 2010. Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science* 21(5): 629–634. doi:10.1177/0956797610367751.

Dils, A.T., and L. Boroditsky. 2010. Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences* 107: 16396–16400. doi:10.1073/pnas.1009438107.

Egly, R., J. Driver, and R.D. Rafal. 1994. Shifting visual-attention between objects and locations - evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology-General* 123(2): 161–177.

Elman, J.L. 2004. An alternative view of the mental lexicon. *Trends in Cognitive Sciences* 8(7): 301–306. doi:10.1016/j.tics.2004.05.003.

Elman, J.L., and J.L. McClelland. 1988. Cognitive penetration of the mechanisms of perception - compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27(2): 143–165.

Estes, Z., M. Verges, and L.W. Barsalou. 2008. Head up, foot down: object words orient attention to the objects' typical location. *Psychological Science: A Journal of the American Psychological Society / APS* 19(2): 93–97. doi:10.1111/j.1467-9280.2008.02051.x.

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

Flanagan, J.R., J.P. Bittner, and R.S. Johansson. 2008. Experience can change distinct size-weight priors engaged in lifting objects and judging their weights. *Current Biology: CB* 18(22): 1742–1747. doi:10.1016/j.cub.2008.09.042.

Fodor, J.A. 1983. *The modularity of mind*. Cambridge: MIT Press.

Fodor, J.A. 1984. Observation reconsidered. *Philosophy of Science* 51: 23–43.

Fodor, J.A. 1988. A reply to Churchland's 'perceptual plasticity and theoretical neutrality'. *Philosophy of Science* 55: 188–98.

Fontanini, A., and D.B. Katz. 2008. Behavioral states, network states, and sensory response variability. *Journal of Neurophysiology* 100(3): 1160–1168. doi:10.1152/jn.90592.2008.

Franconeri, S.L., A. Hollingworth, and D.J. Simons. 2005. Do new objects capture attention? *Psychological Science* 16(4): 275–281. doi:10.1111/j.0956-7976.2005.01528.x.

Friston, K. 2008. Hierarchical models in the brain. *PLoS Comput Biol* 4(11), e1000211. doi:10.1371/journal.pcbi.1000211.

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

Geisler, W.S., and D. Kersten. 2002. Illusions, perception and Bayes. *Nature Neuroscience* 5(6): 508–510. doi:10.1038/nn0602-508.

Ghose, G.M., and J.H.R. Maunsell. 2002. Attentional modulation in visual cortex depends on task timing. *Nature* 419(6907): 616–620. doi:10.1038/nature01057.

Gick, B., and D. Derrick. 2009. Aero-tactile integration in speech perception. *Nature* 462(7272): 502–504. doi:10.1038/nature08572.

Goldstone, R.L. 1995. Effects of categorization on color-perception. *Psychological Science* 6(5): 298–304.

Gregory, R.L. 1963. Distortion of visual space as inappropriate constancy scaling. *Nature* 199(4894): 678–680. doi:10.1038/199678a0.

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

Hay, J., Nolan, A., and Drager, K. 2006. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3): doi:10.1515/TLR.2006.014.

Henrich, J., S.J. Heine, and A. Norenzayan. 2010. The weirdest people in the world? *The Behavioral and Brain Sciences* 33(2–3): 61–83. doi:10.1017/S0140525X0999152X. discussion 83–135.

Hochstein, S., and M. Ahissar. 2002. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36(5): 791–804.

Howe, C.Q., and D. Purves. 2002. Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Sciences of the United States of America* 99(20): 13184–13188. doi:10.1073/pnas.162474299.

Howe, C.Q., and D. Purves. 2004. Size contrast and assimilation explained by the statistics of natural scene geometry. *Journal of Cognitive Neuroscience* 16(1): 90–102. doi:10.1162/089892904322755584.

Howe, C.Q., and D. Purves. 2005. The Müller-Lyer illusion explained by the statistics of image–source relationships. *Proceedings of the National Academy of Sciences of the United States of America* 102(4): 1234–1239. doi:10.1073/pnas.0409314102.

Howe, C.Q., Z. Yang, and D. Purves. 2005. The Poggendorff illusion explained by natural scene geometry. *Proceedings of the National Academy of Sciences of the United States of America* 102(21): 7707–7712. doi:10.1073/pnas.0502893102.

Howe, C.Q., R. Beau Lotto, and D. Purves. 2006. Comparison of Bayesian and empirical ranking approaches to visual perception. *Journal of Theoretical Biology* 241(4): 866–875. doi:10.1016/j.jtbi.2006.01.017.

Itti, L., and P. Baldi. 2009. Bayesian surprise attracts human attention. *Vision Research* 49(10): 1295–1306. doi:10.1016/j.visres.2008.09.007.

Jack, A.I., G.L. Shulman, A.Z. Snyder, M. McAvoy, and M. Corbetta. 2006. Separate modulations of human V1 associated with spatial attention and task structure. *Neuron* 51(1): 135–147.

James, W. 1890. *Principles of psychology Vol. 1*. New York: Holt.

Jordan, T.R., S.M. Thomas, and K.C. Scott-Brown. 1999. The illusory-letters phenomenon: an illustration of graphemic restoration in visual word recognition. *Perception* 28(11): 1413–1416.

Kanizsa, G. 1979. *Organization in vision: essays on gestalt perception*. New York: Praeger.

Kraft, J.M., and D.H. Brainard. 1999. Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America* 96(1): 307–312.

Kreiman, G., I. Fried, and C. Koch. 2002. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences* 99(12): 8378–8383. doi:10.1073/pnas.072194099.

Kutas, M., and S.A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307(5947): 161–163.

Landau, A., L. Aziz-Zadeh, and R.B. Ivry. 2010. The influence of language on perception: listening to sentences about faces affects the perception of faces. *The Journal of Neuroscience* 30(45): 15254–15261. doi:10.1523/JNEUROSCI.2046-10.2010.

Lee, Y. L., Lupyan, G., Roberts-Kedes, G., Mattar, D., Aguirre, G. K., & Thompson-Schill, S. L. (2013). Automatic categorization of letter identity in extrastriate visual cortex. Presented at the Human Brain Mapping, Seattle, WA.

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

Lupyan, G. 2007. Reuniting categories, language, and perception. In *Twenty-Ninth Annual Meeting of the Cognitive Science Society*, ed. D.S. McNamara and J.G. Trafton, 1247–1252. Austin: Cognitive Science Society.

Lupyan, G. 2008. The conceptual grouping effect: categories matter (and named categories matter more). *Cognition* 108(2): 566–577.

Lupyan, G. (2012a). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Cognition*, *3*(54). doi:10.3389/fpsyg.2012.00054.

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123942937000078.

Lupyan, G., and M.J. Spivey. 2008. Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology* 18(10): R410–R412.

Lupyan, G., and M.J. Spivey. 2010a. Making the invisible visible: auditory cues facilitate visual object detection. *PLoS ONE* 5(7), e11452. doi:10.1371/journal.pone.0011452.

Lupyan, G., and M.J. Spivey. 2010b. Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics* 72(8): 2236–2253. doi:10.3758/APP.72.8.2236.

Lupyan, G., and S.L. Thompson-Schill. 2012. The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology-General* 141(1): 170–186. doi:10.1037/a0024904.

Lupyan, G., and E.J. Ward. 2013. Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences* 110(35): 14196–14201. doi:10.1073/pnas.1303312110.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2009). The penetration of visual representations by conceptual categories. In *Meeting of the Vision Sciences Society.*

Lupyan, G., S.L. Thompson-Schill, and D. Swingley. 2010. Conceptual penetration of visual processing. *Psychological Science* 21(5): 682–691.

Lyons, J. 2011. Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues* 21(1): 289–311. doi:10.1111/j.1533-6077.2011.00205.x.

Macpherson, F. 2012. Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research* 84(1): 24–62. doi:10.1111/j.1933-1592.2010.00481.x.

Marr, D. 1982. *Vision: a computational approach*. San Francisco: Freeman & Co.

McCauley, R.N., and J. Henrich. 2006. Susceptibility to the Muller-Lyer Illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology* 19(1): 79–101.

McClelland, J.L., D. Mirman, and L.L. Holt. 2006. Are there interactive processes in speech perception? *Trends in Cognitive Sciences* 10(8): 363–369.

McGurk, H., and J. MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264(5588): 746–748.

Meng, X., and Q. Zaidi. 2011. Visual effects of haptic feedback are large but local. *PLoS ONE* 6(5), e19877. doi:10.1371/journal.pone.0019877.

Meteyard, L., B. Bahrami, and G. Vigliocco. 2007. Motion detection and motion verbs: language affects low-level visual perception. *Psychological Science* 18: 1007–1013. doi:10.1111/j.1467-9280.2007.02016.x.

Meteyard, L., N. Zokaei, B. Bahrami, and G. Vigliocco. 2008. Visual motion interferes with lexical decision on motion words. *Current Biology* 18(17): R732–R733. doi:10.1016/j.cub.2008.07.016.

Moore, C.M., and H. Egeth. 1998. How does feature-based attention affect visual processing? *Journal of Experimental Psychology. Human Perception and Performance* 24(4): 1296–310.

Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1): 62–85. doi:10.1177/0261927X99018001005.

O'Connor, D.H., M.M. Fukui, M.A. Pinsk, and S. Kastner. 2002. Attention modulates responses in the human lateral geniculate nucleus. *Nat Neurosci* 5(11): 1203–1209. doi:10.1038/nn957.

Oliva, A., and A. Torralba. 2006. Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research* 155: 23–36. doi:10.1016/S0079-6123(06)55002-2.

Oliva, A., and A. Torralba. 2007. The role of context in object recognition. *Trends in Cognitive Sciences* 11(12): 520–527. doi:10.1016/j.tics.2007.09.009.

Parker, N.I., and P.L. Newbigging. 1963. Magnitude and decrement of the Müller-Lyer illusion as a function of pre-training. *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 17(1): 134–140. doi:10.1037/h0083262.

Pasley, B.N., L.C. Mayes, and R.T. Schultz. 2004. Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42(1): 163–172.

Pelekanos, V., and K. Moutoussis. 2011. The effect of language on visual contrast sensitivity. *Perception* 40(12): 1402–1412. doi:10.1068/p7010.

Potter, M.C. 1975. Meaning in visual search. *Science (New York, N.Y.)* 187(4180): 965–966.

Potter, M.C., and L.F. Fox. 2009. Detecting and remembering simultaneous pictures in a rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance* 35(1): 28–38. doi:10.1037/a0013624.

Purves, D., W.T. Wojtach, and R.B. Lotto. 2011. Understanding vision in wholly empirical terms. *Proceedings of the National Academy of Sciences of the United States of America* 108(Suppl 3): 15588–15595. doi:10.1073/pnas.1012178108.

Pylyshyn, Z. 1999. Vision and cognition: how do they connect? Response. *Behavioral and Brain Sciences* 22(3): 401–423.

Rao, R.P., and D.H. Ballard. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience* 2: 79–87.

Ross, H.E. 1969. When is a weight not illusory? *Quarterly Journal of Experimental Psychology* 21(4): 346–355. doi:10.1080/14640746908400230.

Rudel, R.G., and H.-L. Teuber. 1963. Decrement of visual and haptic müller-lyer illusion on repeated trials: a study of crossmodal transfer. *Quarterly Journal of Experimental Psychology* 15(2): 125–131. doi:10.1080/17470216308416563.

Samuelsen, C.L., M.P.H. Gardner, and A. Fontanini. 2012. Effects of cue-triggered expectation on cortical processing of taste. *Neuron* 74(2): 410–422. doi:10.1016/j.neuron.2012.02.031.

Sapir, E. 1929. The status of linguistics as a science. *Language* 5: 207–214.

Schultz, G., and R. Meizack. 1991. The Charles Bonnet syndrome: "phantom visual images.". *Perception* 20(6): 809–825. doi:10.1068/p200809.

Schwitzgebel, E. (2013). Striking Confirmation of the Spelunker Illusion. *The Splintered Mind*. Retrieved from http://schwitzsplinters.blogspot.ca/2013/11/striking-confirmation-of-spelunker.html.

Selkin, J., and M. Wertheimer. 1957. Disappearance of the mueller-lyer illusion under prolonged inspection. *Perceptual and Motor Skills* 7(3): 265–266. doi:10.2466/pms.1957.7.3.265.

Shams, L., Y. Kamitani, and S. Shimojo. 2002. Visual illusion induced by sound. *Cognitive Brain Research* 14(1): 147–152.

Sheinberg, D.L., and N.K. Logothetis. 1997. The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences* 94(7): 3408–3413.

Siegel, S. 2012. Cognitive penetrability and perceptual justification. *Noûs* 46(2): 201–222.

Siegel, S. 2013. The epistemic impact of the etiology of experience. *Philosophical Studies* 162(3): 697–722. doi:10.1007/s11098-012-0059-5.

Smilek, D., M.J. Dixon, and P.M. Merikle. 2006. Revisiting the category effect: the influence of meaning and search strategy on the efficiency of visual search. *Brain Research* 1080: 73–90.

Stokes, D. (2011). Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philosophical Studies*, 1–16. doi:10.1007/s11098-010-9688-8.

Störmer, V.S., J.J. McDonald, and S.A. Hillyard. 2009. Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences* 106(52): 22456–22461. doi:10.1073/pnas.0907573106.

Strand, E.A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1): 86–100. doi:10.1177/0261927X99018001006.

Suzuki, S., and M. Grabowecky. 2003. Attention during adaptation weakens negative afterimages. *Journal of Experimental Psychology: Human Perception and Performance* 29(4): 793–807. doi:10.1037/0096-1523.29.4.793.

Taylor-Clarke, M., S. Kennett, and P. Haggard. 2002. Vision modulates somatosensory cortical processing. *Current Biology: CB* 12(3): 233–236.

Tsuchiya, N., and C. Koch. 2005. Continuous flash suppression reduces negative afterimages. *Nat Neurosci* 8(8): 1096–1101. doi:10.1038/nn1500.

Van Boxtel, J.J.A., N. Tsuchiya, and C. Koch. 2010. Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences* 107(19): 8883–8888. doi:10.1073/pnas.0913292107.

Veldhuizen, M.G., D. Douglas, K. Aschenbrenner, D.R. Gitelman, and D.M. Small. 2011. The anterior insular cortex represents breaches of taste identity expectation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31(41): 14735–14744. doi:10.1523/JNEUROSCI.1502-11.2011.

Vickery, T.J., L.-W. King, and Y. Jiang. 2005. Setting up the target template in visual search. *Journal of Vision* 5(1): 81–92. doi:10.1.1.329.5562.

Wang, Q.Q., P. Cavanagh, and M. Green. 1994. Familiarity and pop-out in visual-search. *Perception & Psychophysics* 56(5): 495–500.

Watanabe, K., and S. Shimojo. 2001. When sound affects vision: effects of auditory grouping on visual motion perception. *Psychological Science* 12(2): 109–116.

Weiss, Y., E.P. Simoncelli, and E.H. Adelson. 2002. Motion illusions as optimal percepts. *Nature Neuroscience* 5(6): 598–604. doi:10.1038/nn858.

Winer, G.A., J.E. Cottrell, V. Gregg, J.S. Fournier, and L.A. Bica. 2002. Fundamentally misunderstanding visual perception. Adults' belief in visual emissions. *The American Psychologist* 57(6–7): 417–424.

Wolfe, J.M., S.J. Butcher, C. Lee, and M. Hyle. 2003. Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology-Human Perception and Performance* 29(2): 483–502.

Yuille, A., and D. Kersten. 2006. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences* 10(7): 301–308. doi:10.1016/j.tics.2006.05.002.

Zelano, C., A. Mohanty, and J.A. Gottfried. 2011. Olfactory predictive codes and stimulus templates in piriform cortex. *Neuron* 72(1): 178–187. doi:10.1016/j.neuron.2011.08.010.