

# Is Aesthetic Experience Evidence for Cognitive Penetration?

Daniel C. Burnston

Tulane University, Philosophy Department, and Tulane Brain Institute

[dburnsto@tulane.edu](mailto:dburnsto@tulane.edu)

**PENULTIMATE DRAFT: Please cite final version, available online:**

**<https://doi.org/10.1016/j.newideapsych.2017.03.012>**

## 1. Introduction

There are many theoretical choices involved in whether one argues for against cognitive penetration (CP)—the thesis that perception is sensitive, in an “intelligible way” (Stokes, 2015, p. 75) to our beliefs and other cognitive states. These choices include how to distinguish cognition and perception, how strong a relationship between them needs to be in order to count as CP, and what kinds of cases provide evidence for one view over the other.

*Aesthetic perception* is one case that has recently been argued to provide very good evidence for CP (Nanay, 2014; Stokes, 2014). I will construe aesthetic perception very broadly—as perception of aesthetically relevant *qualities* such as an artworks’ being *somber* or *dynamic*, and perception of artistic *categories*, such as being a Monet or being Mannerist or Cubist. On this broad view, aesthetic perception is what occurs when we use perceptual mechanisms in particularly aesthetic contexts. This assumes that there is *some* connection between what goes on in “normal” (non-aesthetic) and aesthetic perception, but leaves open precisely what that connection is.<sup>1</sup>

The issue of CP stands at a crossroads of philosophical and cognitive scientific analysis of aesthetic experience. Many models within cognitive science make no place for CP, claiming that perceptual analysis of artworks occurs prior to beliefs about them (Bullot & Reber, 2013; Leder, Belke, Oeberst, & Augustin, 2004; Ramachandran & Hirstein, 1999). Some philosophers, on the other hand, claim that CP is vital for understanding aesthetic appreciation. According to McMahan, for instance, aesthetic perception is “concepts all the way out” (2012, p. 417). If this view is right, then there are major revisionary consequences for empirical and philosophical theorizing about aesthetics (Nanay, 2016).

Some factors involved in aesthetic perception make it a good test case for CP. If aesthetic perception genuinely occurs, it is one possible case of “high-level” perceptual content, which is often taken to provide *prima facie* evidence for CP. Moreover, aesthetic perception can be shaped by *expertise* that is mediated by *social instruction* and learning. These combined features can make aesthetic perception look like a likely case for CP—but *only*, I will argue, given some highly questionable assumptions and pragmatic choices about how to construe the debate. These include the tendency to systematically weaken the kinds of relationships that are taken to be evidence for CP, and the (historically well-precedented) tendency to severely underestimate the power of perceptual association and perceptual learning. Together, these assumptions result in

---

<sup>1</sup> Alas, I have nothing to say about the perennial debate about what defines the aesthetic. I’ll settle for discussing paradigm cases.

what I will call the *scaling argument*—the claim that the more high-level, categorical, socially-mediated, and learning-dependent a percept is, the more likely it is to be the result of CP.

I will argue that the scaling argument is invalid. Given the best empirical and theoretical picture of how perception and perceptual learning work, CP is not required to explain what goes on in aesthetic perception. Rather, we can account for even expertise-mediated percepts through purely perceptual processes and more mundane roles for cognition—such as instructing us where and how to look—that are too weak to imply revisionary theses such as CP.

My strategy will be as follows. First, in the next section, I will discuss the pragmatics surrounding debates about CP, with the goal of arguing that we should not straightforwardly sign on for the assumptions mentioned above. Then, in sections 3 and 4 I will discuss two recent arguments in favor of aesthetic CP given by Nanay (2014) and Stokes (2014), which focus on the roles of *association* and *learning*, respectively, in aesthetic experience. I will show that neither argument provides strong support for even a weak view of CP. Section 5 concludes with a discussion of the consequences of social learning for perception. I will argue that these consequences are important, but that we don't need CP to account for them.

## 2. Pragmatics

The version of the CP debate I'll engage with is this: are there causal-informational interactions between perception and cognition, which show that the former is sensitive, in a meaningful way, to what goes on in the latter?<sup>2</sup> “Sensitive” and “meaningful,” of course, require considerable cashing out. Moreover, given that we're analyzing causal relations, we need to have at least a relatively firm grip on the relations we're discussing. I'll spend the rest of this section first describing the different kinds of CP theses that are on the table, and then discussing how to make calls about what falls on what side of the perception/cognition border.

A number of definitions of CP have been posited, which we can identify as “strong,” “weak,” and “consequentialist.” Strong views posit that there is not only a causal relationship between perception and cognition, but that cognitive contents have a semantic influence on perceptual processing. The strong view is articulated slightly differently by different theorists: perceptual outcomes have a “logical relationship” to cognitive states; cognitive states provide an “informational resource” to perception; cognitive states “partly determine” the content of perception (Macpherson, 2012; Z. Pylyshyn, 1999; Siegel, 2012; Wu, 2013). The differences all coalesce, however, on the view that cognitive influence on perception is semantic and computational (Burnston, forthcoming).

Weaker views do not explicitly invoke semantic or computational relations. According to Stokes' (2012) early formulation, an interaction between cognition and perception counts as CP as long as that link is *internal* and *mental*. This rules out cases in which the cognition-perception relation is mediated by overt movement. For instance, if one's belief that there is beer in the

---

<sup>2</sup> A distinct, but not always clearly delineated thesis is that there simply is no clear distinction between perception and cognition (Churchland, 1988; Lupyan, 2014; and Vetter & Newen, 2014, I believe, all come close to endorsing this thesis). Construing the debate in terms of causal relations, of course, requires that they can in fact be distinguished. I believe that such a distinction can be drawn (Burnston, forthcoming), and I say more about it below.

fridge, combined with a desire for beer, cause an action (opening the fridge) which leads a beer perception, this wouldn't be sufficient. However, internal processes such as covert attentional effects might be enough to enact CP, since they are mental processes by which a cognitive state affects a perceptual outcome.

Another proposal, represented in some of Stokes' more recent work (Stokes, 2013, 2015), is to analyze whether cognition-perception interactions count as CP based on their *consequences*. Instances of CP are supposed to have upshot for views about the epistemic role of perception, about theory-neutrality, or about the modularity of perceptual systems. Stokes argues that consequentialist considerations speak in favor of weak views of CP—i.e., if an interaction meets the weak definition, it is likely to have the relevant consequences and thus be worth calling CP.

To avoid begging any questions, I don't want to commit to a particular rendering of CP at the outset. What I *do* want to focus on is the fact that CP is supposed to be a revisionary thesis: accepting it is supposed to have, as Stokes puts it, "important philosophical and scientific consequences" (Stokes, 2014, p. 1). We thus have a kind of litmus test for considering CP: weak views should not be so weak that they demand no significant changes to models of aesthetic perception that make no explicit place for CP. Philosophically, we can ask whether more commonplace relations, which all parties would acknowledge independently CP, account for the nature of a proposed instance of CP. If so, then the instance does not force revisionary theorizing. I'll argue that, given appropriate views of perception and perceptual learning, interactions between cognition and perception can be accounted for in such more commonplace ways.

The other major question to be considered in this section, then, is about how to construe perception and cognition. As we will see at several points, the less sophisticated one thinks perception is, the more likely one considers CP to be. This is one of the key assumptions behind the scaling argument. Rather than attempting to firmly define the line between cognition and perception, I will satisfy myself with characterizing some of the resources that perception brings to the table.<sup>3</sup> We can then ask whether, given these resources, we need to posit CP in particular instances, or whether less substantial relationships can suffice.

On the view of perception that inspires the scaling argument, perception is a rather brute low-level device, comprising some strictly bottom-up feature detectors, and perhaps (at most) some mechanisms for combining their outputs into bound objects. Anything *categorical*, *associative*, or *top-down*, or anything involving *inference* or *assumptions*, must be due to cognition. However, there is strong reason to question this kind of view. Consider the following two examples.

---

<sup>3</sup> In other work (Burnston, forthcoming), I argue that the best way to distinguish cognition from perception is via a distinction in representational *form*—perceptual representations have a structure that maps to the structure of their referents, whereas cognitive representations are propositional and language-like (cf. Dretske, 1981; Isaac, 2013; Kulvicki, 2007). I think the form-distinction is a good way to explain why the broader distinctions I make in this section hold, but I don't want my argument here to be beholden to it.

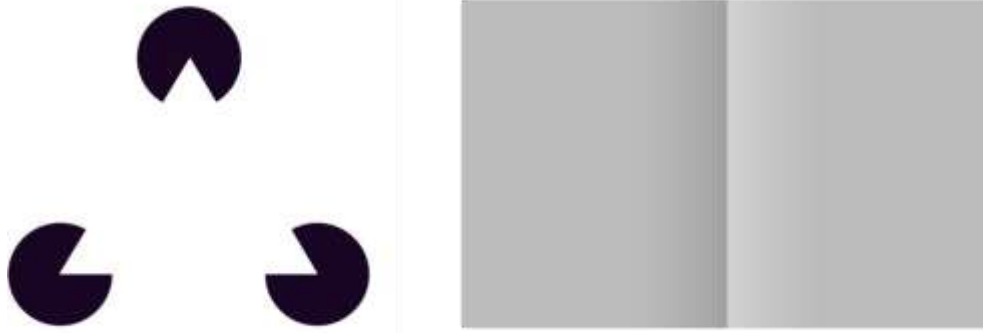


Fig 1. The Kanizsa triangle and the Cornsweet illusion.

In the Kanizsa triangle to the left, the alignment of Pacman-shapes suggests the corners of a triangle sitting atop three circles. Subjects have the distinct impression of seeing the boundaries of the triangle even against the purely white background. Indeed, this impression correlates with activity in appropriately orientation-selective V1 cells whose receptive fields are in the area of the illusory border. One story for how this effect comes about is that top-down connections from higher levels of the visual cortex (for instance, V2) activate cells selective for orientations in the area of the illusory border (Albright & Stoner, 2002).

In the right panel, the “Cornsweet illusion,” subjects almost universally see the patch to the left as darker than the patch on the right. However, the two patches have exactly the same luminance values, other than the light-to-dark gradient to the left of the center line and the dark-to-light gradient to the right of it. The standard story regarding this illusion is that perception assumes that the discontinuity flanked by the gradients is extended towards the perceiver in depth. The left side, as cued by the dark-to-light gradient, is in shadow, and the right side is directly lit. It thus represents the left side as darker than the right (Purves, Shimpf, & Lotto, 1999).

These cases involve many of the properties discussed above. They are associative, relating distinct perceptual features. They involve top-down influence. And they are *abductive*—that is, they go beyond the given stimulus values based on assumptions about the scene. However, it turns out that we only need to grant perception one relatively basic ability in order to account for all of these properties. That is the ability to *internalize regularities*. These associations between different cues—luminance, depth, and shape—reflect assumptions about how these kinds of features are normally correlated. According to Shepard (2001), perception itself operates by making assumptions about how perceptual features mutually constrain each other in the kinds of objects and spatial environments that we encounter in the world.

Contra those who define these properties as cognitive (e.g., Churchland, 1988), there are some strongly motivated reasons to count these effects as perceptual. The first is their dependence on *configuration and structure*. Each of these effects requires particular arrangements of cues in the scene. In the Kanizsa stimulus, if there are no black circles (or other relevant cues) to provide depth-indicating contrast, or if the pacmen are not in a “coincident alignment” (Stoner & Albright, 2002, p. 351), then the effect will not occur. The Cornsweet effect depends on the two

luminance gradients being on either side of a discontinuity (the “Cornsweet edge”; Purves et al., 1999). On the plausible assumption that vision has as its primary function representing stimulus values in spatial configurations, then these kinds of properties strongly speak to a perceptual reading.<sup>4</sup>

A second reason involves a rather large gulf between the kinds of associations being implemented here and the kinds of beliefs subjects are likely to hold. For one, these effects are much more *fine-grained* than most of our beliefs. Consider: in the Cornsweet case, the effect can be enhanced by “bowing” or “foreshortening” the two panels, thus increasing the depth cue. It can be eliminated by reducing the contrast of the panels with the background, thus eliminating the depth cue. These effects reflect very specific mappings of certain cue arrangements to certain perceived feature values. It is unclear at best where we would have come by such specific beliefs about how to relate these perceptual features.

Moreover, there are some extreme differences between these perceptual abductions and stereotypical cognitive inferences. Cognition operates on testimony, considers arguments, and explicitly represents reasoning rules such as consistency, avoidance of fallacies, etc. There is no reason to read these properties into the perceptual abductions at stake here, especially given that, as Fodor (1984) stressed long ago, many of these effects are *mandatory*—i.e., not sensitive to changes in our occurrent beliefs.<sup>5</sup> Now, Fodor took such properties as mandatoriness to entail the impenetrability of perception. That is a stronger claim than I want here. For the rest of the paper, I will assume that perception *can* internalize regularities in its processing. This is only a modal claim, which leaves open the possibility that in *some* cases cognition influences these processes in a way worth calling CP—I’ll assess this possibility in aesthetic cases below.

Before doing that, however, let’s consider *categorical* perception. Take, for example, the category of *faces*. We can recognize many facial categories, including individuals, genders, and races, even when viewing novel exemplars with which we have no prior experience. On several well-supported current approaches, this recognition is itself a perceptual process (Tsao & Livingstone, 2008). The basic idea is that a perceptual category is a location in a *perceptual space*. When one perceives something as belonging to a category, one perceives it as occupying the space typical of the category, or close to it. A space is composed of dimensions, which can correspond to simple features or to higher-order relationships between them. Basically, a dimension is any parameter that varies systematically. In faces, for instance, one has simple features of lightness and color which can vary in systematic ways, but also face-specific dimensions, such as the ranges of shape and size in noses, eyes, ears, etc. A perceptual category is a higher-order set of dimensions that captures relationships between these simpler features.

---

<sup>4</sup> At least in principle, this kind of view can be scaled up considerably, to account for complex scene recognition, biological motion (Burnston & Cohen, 2015), and even forms of social perception (Toribio, 2015). If a temporal aspect is added to the regularities that perception can internalize, then phenomena such as perceived *chasing* can be counted as purely perceptual (Burnston & Cohen, 2015; Gao, Newman, & Scholl, 2009).

<sup>5</sup> Churchland’s (1988) argument against mandatoriness implying impenetrability is based heavily on the idea that *learning* has to occur in order for perception to, eventually, mandatorily interpret the stimuli in these ways. And Churchland defines any learned process as cognitively penetrated. As we’ll see in section 4, however, this view can’t be assumed without argument, and there is compelling reason to question it.

Importantly, these categories are *holistic*, and allow for *generalization*. Higher-order dimensions both correlate variations among individual parameters, and represent *configural* spatial relationships between them. In faces there are correlations between, for example, facial lightness and nose shape, but there are also stereotypical spatial relationships—male faces tend to have larger spaces between the nose and the upper lip than female ones, etc. Possessing a perceptual category involves representing associational and configural relationships within the relevant perceptual space. Importantly, perceptual categories allow one to *generalize*—once one can represent a dimensional space constituting a category, one can recognize novel exemplars (e.g., a male face one has never seen before, or an individual’s face at a new angle) in virtue of their falling in the characteristic space. Similar dimensional accounts have been offered for perceptual categories corresponding to objects such as cars and birds (Folstein, Gauthier, & Palmeri, 2012), as well as for categories of complex shapes (Schyns & Rodet, 1997), made-up but vaguely “biological-looking” creatures (Chua, Richler, & Gauthier, 2015; Folstein, Gauthier, & Palmeri, 2010), and so on.

Why should this kind of category representation be counted as perceptual? Notice that it exhibits the properties discussed in the simpler cases given above. It is dependent on the spatial configuration of perceptual features. It exhibits inference (in this case, inference about category membership), based on internalized regularities about how features are correlated. Moreover, it is often mandatory—it is hard to not see a male face *as* male, your brother’s face *as* your brother, just like it is hard not to see the Cornsweet patches as differing in brightness, or certain patterns of dot-motion as a person walking. Finally, as I will explain in more detail below, subjects are often unaware of the dimensional structure that informs their categorization—i.e., in certain cases it can be shown that subjects can recognize faces even when it is very implausible that they have beliefs about the dimensional structure they recognize. Again, this is only meant to establish a modal claim. Perception *can* represent categorical dimensional structures, in a way continuous with its ability to internalize regularities in simpler cases. This does not settle the question about CP writ large.

However, if all I’ve said in this section is granted, some very significant pragmatic consequences follow, including for the debate about aesthetics. For instance, Nanay (2014) suggests that aesthetic perception involves top-down processing. But as long as it is possible for top-down processing to occur within perception (as, for instance, in the explanation of the Kanisza effect given above), simply making that suggestion is no argument for CP (cf. Deroy, 2013). Both Stokes (2014) and Seeley (2013) suggest that aesthetic perception is mediated by tacit knowledge. But if perception can embody assumptions about stimuli, then this itself is not an argument for CP. Similarly, the bare fact that we can perceive aesthetic qualities or categories is not sufficient—it requires further argument to establish that these instances *in fact* involve CP.

I will now assess two arguments that purport to establish CP in the aesthetic case. The first is Nanay’s claim that CP is implemented through *conceptual association* with artworks. The second is Stokes’ argument that CP is implemented through *learning* to perceive aesthetic categories and properties. I will argue that in each case, there is a modal claim about perception related to the one I’ve made here and that, given the modal claim, CP is not required to explain the effects discussed. That is, the scaling argument is a bad one, and even perception of highly social categories like artworks does not involve CP. In the following two sections, I will only

consider the strong and weak views of CP. In section 5 I will take up the consequentialist position.

### 3. Conceptual Association and CP

Nanay (2014) claims that our associations with particular concepts influence how we perceive artworks. Nanay's target is Danto's (1981) "gallery of indiscernibles" thought experiment, which was intended to show that "aesthetic value is only loosely (if at all) related to perception" (Nanay, 2014, p. 1). On Danto's thought experiment, two perceptually identical red squares are given distinct titles: "Red Square" and "Israelites Crossing the Red Sea." Danto claims that while one's overall evaluation of the works will differ when given the distinct titles, one's perceptual experience will not change. Hence, aesthetic valuation does not supervene on perceptual experience.

Nanay points out that Danto's argument goes through only if there is no CP involved in these aesthetic experiences. If CP occurs, then the two experiences *can* differ, despite the identity of perceptual input. In particular, he suggests that our respective associations with the Soviet flag and the Old Testament anecdote can change our visual experience. Despite the critical tenor of his argument, Nanay draws some positive conclusions from it. He suggests that "in general, one's experience is not determined in a bottom-up manner by the perceptual stimulus" (Nanay, 2014, p. 2), and thus that CP is a general mechanism by which aesthetic evaluation and perception might be joined.

As pointed out above, the simple fact of top-down processing is not sufficient to show CP. It has to be, specifically, top-down influence from *cognitive* representations on perceptual ones. Conceptual association, for Nanay, is the cognitive influence that exerts the needed top-down influence.

To make his argument, Nanay cites purported evidence for CP in non-aesthetic cases. These include heart-shapes looking redder than non-heart shapes (Delk & Fillenbaum, 1965), and faces labeled 'black' looking darker than faces labeled 'white' (Levin & Banaji, 2006). He further suggests that if CP occurs in these cases, then it is likely to occur in artistic cases as well. That is, if association with the concept 'black' is likely to make a face look darker, then association with the concepts 'Israelites' or 'Soviet flag' is likely to also modify our percepts. The claim here is that, to the extent that a perceptual scenario involves higher-level and more socially mediated categories, that scenario is more likely (or at the very least, as likely as its less-sophisticated counterparts) to involve CP. Hence, Nanay's claims rest on the scaling argument. I won't discuss the heart and face cases in detail here (see Burnston, forthcoming). I want to question whether the scaling argument is valid—i.e., regardless of whether CP occurs in these other cases, does conceptual association bring about CP in the aesthetic case?

The argument goes in the way outlined in the previous section. First, I'll attempt to articulate a modal claim—in particular that perception, as well as direct (i.e., not cognitively-mediated) association between perceptual, motor, and affective states, can provide descriptions of how we perceive aesthetic qualities like *dynamism* and *intentionality* in an artwork, as well as our

*emotional and evaluative* responses to it. Once this is admitted, I'll then argue that even when conceptual association *does* occur, we don't need to read it in terms of CP.

First, let's consider purely perceptual processes. Livingstone (2002) has suggested that the *dynamic vitality* of the Mona Lisa is partly due to spatial frequency and the interaction between peripheral and central vision. That is, her smile is more pronounced at low-spatial contrast, which is processed more heavily by peripheral than foveal vision. Hence as one scans the painting, and the smile shifts from central to peripheral vision, it seems ever so slightly to change. This is a representation of an aesthetic quality achieved purely through the manipulation of the visual system—dynamism in particular is a property that Stokes (2014) thinks perception, on its own, has no means to access. If this *kind* of explanation generalizes, then even perception of seemingly distinctive aesthetic properties can fall under purely perceptual explanation.

And there is reason to think it can generalize, given the dimensional view of categorical perception outlined in the last section. Ramachandran and Hirstein (1999) suggest that artworks work partially by stretching relevant perceptual dimensions. Given a dimensional perceptual category for faces or bodies, artists can manipulate these dimensions in order to produce specific effects—accentuating or exaggerating certain dimensions in ways that are intended to affect perception of the work. This is particularly powerful when combined with direct association between dimensional perceptions and motor or affective states.

For instance, we seem to have baseline preferences for certain postures and movements (Orgs, Hagura, & Haggard, 2013). In the perception of dance, subjects initially rate symmetric movements as more pleasing than asymmetric ones. These associations *can* be built or modified without cognitive influence. Ratings for asymmetric movements go up simply through viewing them, an example of what is called a “mere exposure” effect (Orgs et al., 2013, and citations therein). Here we have visual recognition of a certain kind of movement in a body-movement space developing an affective association, namely going from negative to positive, and underlying an evaluative representation of the movement. Importantly, subjects in the study were not informed that symmetry was being tested, and as all subjects were novices in dance perception, they likely had no prior beliefs about their preferences for dance movements. Hence, mere exposure effects are best explained in terms of modulated associations between posture-perception and affect.

In some cases, perceptual-to-affective association is mediated by *motor emulation*—automatic mapping of perceived bodily states onto one's own somatic and motor representations—and these associations can provide explanations of a wide range of cases (Freedberg & Gallese, 2007). Consider the following three artworks, two representational and one non-representational.





Figure 2. Artworks evincing sensorimotor association.

Here are interpretations of these artworks from the perspective I've been advocating. El Greco's *The Vision of St. John* (left) and Rodin's *The Age of Bronze* (middle) involve perceptual exaggerations of the dimensions of the figures—more pronounced in El Greco's case, through the lengthening and curvature of the figures; more subtle in Rodin's by the lengthening and torsioning of the torso—which trigger embodied effects that underlie the emotional reaction to the artworks. We feel the emotion of the figures partially due to emulating their postures with our own motor systems (Freedberg & Gallese, 2007). Production of these emotional responses, presumably, underlies at least some of our evaluative response to the artwork. Importantly, this is not restricted to representational artworks. Umiltà et al. (Umiltà, Berchio, Sestito, Freedberg, & Gallese, 2012) have shown motor activation in subjects' brains when they view non-representational works like the Fontana to the right, and interpret this as simulation of the act required to produce the artwork. These kinds of responses are important, because they offer a sensorimotor association account of perceived intentionality in artworks—a key aesthetic property that McMahon (2012), for instance, thinks can only be generated through cognitive appraisal. The modal claim for this section is that dimensional perception plus motor and affective association *can* result in the representation of properties that, if the scaling argument is assumed, must be due to cognitive influence.

I want to stress firmly here: I am *not* arguing that *all* aesthetic responses can be accounted for in this way (cf. Seeley, 2013). Conceptual association certainly occurs. However, if the modal claim is granted, we can ask whether cognitive association working with this baseline system should be described in terms of CP. I suggest the following alternative to Nanay's account: conceptual association helps form associations between perceptual representations and emotional reactions that are hard to implement directly, or through the medium of the motor system. This can be *an* important role for conceptual association in producing overall aesthetic responses and evaluations, but it is not one that includes CP.

To see how this might work, start with a non-aesthetic case. My grandmother loved butterflies, monarch butterflies particularly. And I miss my grandmother. So when I see a butterfly, a monarch butterfly particularly, this conjures up recollections of my grandmother, and inevitably

inspires a wistful emotion. Arguably, this abstract association requires conceptual abilities.<sup>6</sup> I have to know that my grandmother adored butterflies before the emotional association can come about. But though the association plays a role in my overall experience, it does so through conceptualization-emotion association, not by a causal influence on perception.

So far, this is simply an alternative. Why should we believe it rather than extending purported results of CP from non-aesthetic cases and inferring CP in the aesthetic case? There is an important difference between the kinds of conceptual associations purported to be at work in non-aesthetic cases, and those that Nanay cites in the aesthetic ones. The category ‘hearts’ has some *stereotypical* perceptual properties associated with it, namely being heart-shaped and red. The category ‘black’ has canonical perceptual associations as well, namely being dark. But in the case of associations with artistic objects, there are often no such stereotypical consequences. Take the Soviet flag. Other than being red, which is already perceptually represented, the flag does not have any one specific set of *either* conceptual or perceptual properties associated with it. To someone who lived in the U.S. during the Cold War, it is likely to be ominous. To a Leninist (if any remain), it is likely to be a symbol of hope. To someone who lived through Perestroika, maybe it is a symbol of disillusionment. All of these associations with the concept are possible, and there are no obvious stereotypical perceptual properties associated with any of them.

Now, maybe a story could be told on which ominous things should be perceived as darker, symbols of hope brighter, symbols of disillusionment paler, or some such. But as we build more complex associations, the CP story in fact looks *less and less* likely. Consider: a Christian is likely to have particular associations with the El Greco above, due to its biblical content. A historian is likely to have particular associations with the Rodin, due to its thematic connection to the aftermath of the Franco-Prussian war. Both of these sets of associations will differ from whatever ones I had as a first-year college student seeing these pieces at the Met for the first time. But it isn’t obvious what the perceptual ramifications of these assorted associations should be, and this renders aesthetic association quite distinct from standard (purported) cases of CP. In the Fontana case, presumably some of the associations will be mediated by beliefs about the aesthetic value of representational versus non-representational art, Fontana’s historical role in the respective art-theoretic debates, etc. It is difficult to even guess at an appropriate perceptual consequence to this association.

The point is made even more strongly when *multiple* associations are taken into account. Consider McMahon’s (2003) phenomenological report about her associations on viewing Vermeer’s *The Milkmaid*: “thoughts that flood my mind when contemplating this painting relate

---

<sup>6</sup> I will assume that it does for current purposes, but it’s not clear that that’s correct. My grandmother often had images of butterflies about her—on brooches, on decorations in her Coney Island apartment, etc.—so it is possible that I constructed a purely perceptual association between them. In general, it is a very interesting question exactly how complex perceptual and sensorimotor associations can get. Ramachandran and Hirstein (1999), for instance, suggest that part of the power of metaphor is that it taps into complex associations that operate outside of our explicit beliefs. I will assume that conceptualization is sometimes needed, however, and argue that even in these cases it doesn’t work by implementing CP.

to servitude, resignation to one's lot, idealizations of honest physical toil, simple rewards, the solace of hard work, traditional female roles, and the security and certainty of older cultures weighed against their rigidity and emphasis on conformity" (p. 269). It makes sense that reflection on servitude, gender roles, and traditionalism could have emotional association for people, and indeed, different emotional associations depending on different individuals' views on these topics, but I have no idea what perceptual consequences should be enacted in response to this litany. And I highly doubt that cognition does either.

If these considerations are reasonable, then things look bad for the scaling argument, which assumes that socially mediated categories such as artworks are likely candidates to implement CP. Reflection on the kinds of conceptual associations involved show that this is very unlikely to be the case.

So, there is an alternative story to CP in the aesthetic case, and one that has some pull behind it. Note that the alternative I've proposed is compatible with this view that there are no causal interactions *at all* between concepts and perception. While this might be too strong a view in general (Burnston, forthcoming), it at least shows that the role of concepts in aesthetic appreciation needn't operate in terms of modulating perceptions, and thus it seems that neither the strong nor weak views of CP are required. In response, it might be argued that, in this case cognition at least has *close access* to the workings of perception and emotion, if it is able to mediate associations between them. If this is important for aesthetic experience, is it not worth calling CP, at least in some weak sense?

However, calling this kind of influence CP requires no revisions or modifications to standard models of aesthetic experience. Bullot and Reber (2013), for instance, suggest that beliefs and associations with the artwork come into play *after* perception of it. While these may affect the overall evaluation, they don't affect perception. Calling the kind of view I've offered an instance of CP would be simply verbal; it adds or subtracts nothing from extant models. Now, I don't want to weigh in here on any of Danto's larger claims about individuating artworks or the limits of aesthetic theory—I simply want to question the validity of the scaling argument as applied by Nanay. That is, if Danto is wrong in his larger agenda, it is not *because* conceptual association in aesthetic experience implements CP.

#### 4. Learning and CP

Stokes (2014) argues that it is during *learning* that aesthetic concepts and beliefs are likely to implement CP. The idea is that socially-mediated cognitive experience and artistic instruction convey to us beliefs about what constitutes certain artistic categories, and that these beliefs eventually enable us to perceive the categories. Stokes takes this view to be in support of theories of art appreciation that rely on the art-historical knowledge of the viewer. For instance, on Walton's (1970) view, we can only perceive the aesthetic qualities of an artwork relative to a category for that artwork. Whether we perceive a work as (e.g.) dynamic or somber depends on a contrast class, since something that is somber for a Dali may not be somber for a Corot, etc. Hence, we must perceive aesthetic objects in relation to a "gestalt" that defines the artistic category to which they belong. McMahan (2003) suggests that the perception of artistic *style* requires a "canonic schemata," stored in long term memory, against which an occurrently viewed

aesthetic object can be compared. Each of these positions relies, at least implicitly, on the idea that expertise is shaped by learning, and learning is cognitively and socially mediated. McMahan suggests that since these schemata are learned during the observer's lifetime, they are "determined by the ideas, theories, conceptions of art history, and so on to which they have been exposed" (2003, p. 260; cf. McMahan, 2012). Stokes suggests that learning and expertise mediate the attentional capacities of art-viewers, and that these capacities implement CP of their experience.

I will assume the holistic dimensional structure for perceptual categories discussed in section (2), and that this is the way to interpret talk of "schemas" or "gestalts." As in the other cases, I will attempt to establish a modal claim, this time regarding perceptual learning. I will then attempt to argue that, given the modal claim, CP is not a good way to describe the role of cognition in socially mediated perceptual learning. The modal claim is very important here, because Stokes thinks it is simply *impossible* for perceptual learning to implement percepts corresponding to socially constructed categories without cognitive intervention. He says "there is no account to be given about the evolution or plasticity of perception for the Pink Panther or the Coca-Cola icon" (Stokes, 2014, p. 16). This is still another instance of the scaling argument—it is simply assumed that high-level perceptual learning must be due to CP. Establishing a modal claim regarding perceptual learning of higher-level categories would show that this assumption is wrong, at least when stated categorically in this way.

The modal claim is established by showing cases of perceptual learning that (i) correspond to categories, and (ii) are not plausibly due to beliefs. Here is a case, which shows that we can learn *novel* dimensions through *mere exposure*. Folstein et al. (2010) constructed stimuli consisting in "creatures": objects with recognizable heads, arms, wings, etc., but which don't closely resemble any real animal. In the examples, certain lower-order dimensions were correlated with each other (e.g., a type of head correlated with a type of wing). Subjects were instructed simply to look at the stimuli during a training period. That is, they were given no explicit instructions about how to view the stimulus, what to look for, etc. Despite the lack of instruction, they still picked up on the higher-order dimensional structure. This was shown through a secondary task. One experimental group had to learn a category that matched the dimensions present in the original training set, while a second group had to learn a category based on a different dimensional structure. The first group learned the new classification much more quickly, suggesting that they had learned the original dimensional structure through mere exposure.

In this case, subjects have no prior beliefs about the category or the dimensions to be learned—both are entirely novel. They had no category-specific strategies for viewing or attending to the stimulus, for the same reason. But they picked up the dimensions anyway. Folstein et al. (2010) call this "statistical perceptual learning." On this view, perceptual learning itself seeks for dimensions that differentiate categories, and begins to form holistic representations on the basis of these dimensions, even with no prior categorical knowledge to inform it.<sup>7</sup>

---

<sup>7</sup> Mere exposure effects such as this are well-recognized in the perceptual learning field. See the citations in Folstein et al. (2010).

One possible response is to just define this dimensional learning as cognitive. But going this route would require presuming an a priori limitation to perception, and thus beg the question. The dimensional structure that is learned here is best described in perceptual terms. The process that learns to represent the dimensions operates prior to any beliefs about the category on the part of the learner. And, if the previous arguments that perception can internalize regularities are granted, then representing this kind of dimensional correlation is already within the purview of perception, supposing we admit it some plasticity in its ability to do so. All that is required to see this learning as perceptual is to see it as an extension of what perception is already established to do. Calling this process cognitive would then have to rest on a stipulation than anything worth calling “learning” can never be perceptual. But this is exactly the kind of claim that is under question here. So, the modal claim for this section is that statistical perceptual learning can occur without cognitive influence.

If the modal claim is correct then Stokes’ blanket statement that there is “no account” of perceptual learning that can explain how we come to recognize novel, socially constructed (e.g., constructed by an experimenter) perceptual categories is simply false. We can get some considerable distance in recognizing novel, arbitrary dimensions just with statistical learning on the part of perception. Now, as with the other discussions, the modal claim doesn’t tell the whole story. Often instructions and prior beliefs *do* contribute to how learning comes out. But, I will now argue, given the modal claim there is no motivation to interpret these interactions in terms of CP.

First, it must be noted that the simple fact that some bit of learning is instructionally mediated doesn’t mean that it is an instance of CP. There are many types of instructions that might be causal precursors to a piece of learning, but in entirely commonplace ways. When I was a kid, I wasn’t too good at remembering what type of wrench was a socket wrench. When my dad wanted a 3/8 socket, he had to remind me to *look in the black wrench case*. Given this information, I was able to find the case, open it, and grab the one labeled 3/8. The instruction told me *where to look*. Given the instruction, I could find the item, and thus perceive it by looking at it. Stokes admits that this kind of “indirect” effect of cognition on perception is not sufficiently robust to compel interpretation in terms of CP. There are corollaries to these kinds of instructions in the art case. If it’s my aim to learn about Mannerism, then your instruction to go to the Mannerism wing of the museum will be quite helpful. You might even tell me it’s on the third floor, past the lobby, etc. But, according to everyone, this is not CP.

Similarly, there are a range of behavioral instructions involving *how to look* at something that are not robust enough to count. You have to learn that, to see a picture in a Magic Eye, you must start with it close to your face and move it outward. Obviously, the content you actually see bears no deep or interesting relation to the behavioral skill you’ve learned, and hence I take it that no one wants to call Magic Eye perceptions cases of CP—certainly, no theorist worth her salt would have denied the phenomenon of Magic Eye (!), independently of her feelings about CP.

I think that where-to-look instructions are a subset of how-to-look instructions (looking *left* is one way to look), and will thus use “how-to-look” instructions to cover both. Notice that these instructions are about *behavioral precursors* to perceptions. They tell you to *do* something, so

that you can then perceive the item you're supposed to look for. In order to show that socially-mediated learning institutes CP, we'd need to have a relationship that is categorically distinct from these more mundane cases. I'll argue that even in cases of highly mediated perceptual learning, cognitive influence is not more than a behavioral precursor. In particular, instructions and beliefs tell us *which objects to focus on*, so that we can *then* learn something perceptually about these categories. These kinds of instructions are also a version of how-to-look instruction, and needn't be read either in terms of strong or weak CP.

To make the case, let's consider two types of instructions that are employed in perceptual learning studies—*indexical* instructions and *descriptivist* instructions. The first simply says the category to which a presented exemplar belongs. The second explicitly describes some perceptual properties of a given category. I will assume that these are good models for the beliefs that subjects have prior to learning, and indeed this seems apt. One way that you can instruct me to learn a category is to attempt to inculcate in me either appropriate indexical beliefs about category exemplars, ('this is a Mondrian') or beliefs that describe some features relevant to the category ('Mondrians consist in flat colored shapes').

Take indexical instructions first. Goldstone and Steyvers (2001; cf. Gureckis & Goldstone, 2008), constructed arbitrary face stimuli by “melding” four distinct faces along two dimensions. That is, each face was placed at a corner of a dimensional space, and other exemplars were constructed by blending these faces towards the ones at the other corners. The faces are shown below in figure 3. The experimenters then defined two categories of faces, determined by the line in the middle of the space. Faces on the left of the line were called type ‘A’, those on the right type ‘B’. Subjects, during a training period, were shown exemplar faces from the left or right, and were subsequently told the category to which that face belonged, without any further information about the category. They were then given new exemplars and asked to classify them, which they did with a high degree of success. Interestingly, after learning, subjects *perceive* objects differently along the relevant dimension. They will perceive objects in class A as *closer together* along the horizontal dimension than those of type B, even if (i) they compare an exemplar from type A and one from type B that are very close along this dimension (e.g., in the highlighted section on either side of the vertical line), and (ii) the exemplars are ones that they did not see during training.<sup>8</sup> This is an instance of what has subsequently been called “morphspace stretching” (Folstein, Gauthier, & Palmeri, 2012)—during learning, the perceptual system not only determines which dimensions are diagnostic, but tweaks the representation of those dimensions to group category-members closer together along them. Hence, this is an example of dimensional representations being used to generalize.

---

<sup>8</sup> Similarity can be measured either in terms of increased/decreased discriminability in same-different judgments (Folstein et al., 2012) or by similarity ratings (Goldstone, Lippa, & Shiffrin, 2001).

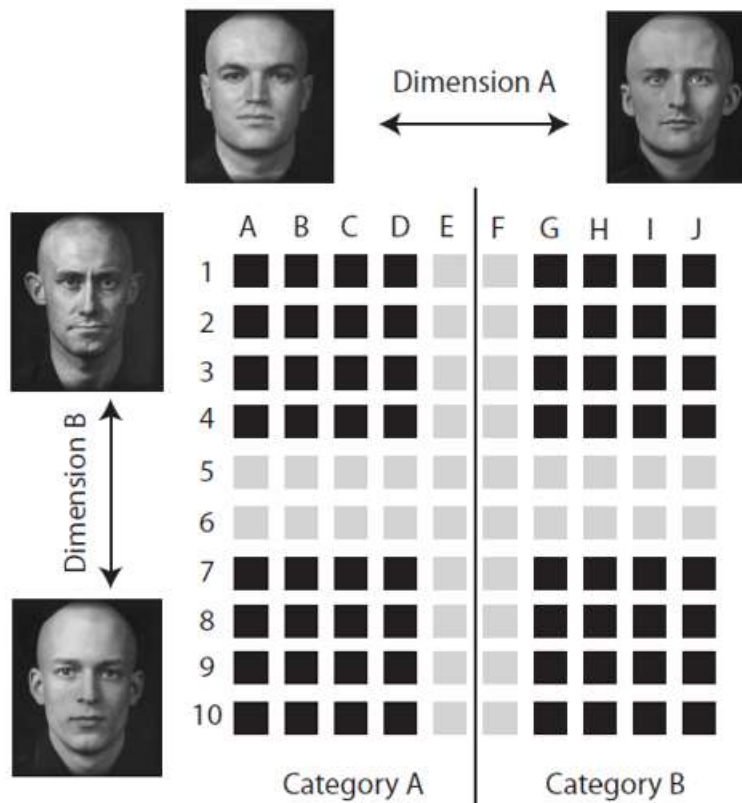


Figure 3. Morphed faces (from Gureckis & Goldstone, 2008).

I suggest that this effect should not be viewed in terms of CP, either strong or weak. The strong view of CP is simply a non-starter here. At best, the beliefs that subjects could attain on the basis of the instructions are a series of indexical beliefs (‘This is an A’, ‘This is a B’, etc.) Since the beliefs that subjects actually hold contain *no content* about the perceptual dimensions of the category, the beliefs have no resources to provide an informational resource to perception about what the category should look like. Importantly, in post-study interviews, subjects did a very *poor* job of describing the dimensions that were actually informing their decisions. While some mentioned particular facial features (lower-order dimensions) as what they looked for in judging cases, none described the dimensional structure that they actually learned. So, here we have a learned dimensional structure that underlies categorical perception but whose content considerably outstrips any beliefs that the subjects could reasonably have been taken to have prior to learning.<sup>9</sup>

A proponent of the weak view might contend that, since the indexical beliefs are *a* causal precursor to the percept, this should count. But consider a simplified case. Suppose you have

<sup>9</sup> One possible alternative would be to suggest that subjects, without instruction, employ a hypothesis-and-test method, forming provisional beliefs about dimensions are relevant and how to group examples along them. However, this kind of view would not explain morphspace stretching. Confirming a belief about the relevant dimensions would not entail either (i) *that* the morphspace should be stretched, or (ii) that it should be stretched by any particular amount.

two boxes in front of you, along with a set of objects. I tell you to lift each object successively. Each time you lift a green one, I say “left box”; each time you lift a red one I say “right box.” The instruction has told you how to lump the objects together, but it hasn’t said *anything* about the particular perceptual properties of the distinct categories. You might learn the fact that all of the items in the left box are green, but you learn it perceptually, without any help from the instruction. What the modal claim suggests is that, so long as there is *a* dimensional structure to be learned about the objects, your perceptual system will begin to pick up on it. In essence, the precursor is *behavioral*. One is told to group *this* face in the A box, but this doesn’t provide any content about the actual perceptual category—indeed *any* learnable perceptual structure amongst the “A box” items could be picked up on the same way.

The defender of the weak view might object that in perceptual case the relationship is *internal*, and this makes the difference. The “A box” here, after all, is a mental box, and one must lump the A exemplars together to discern their structure. What I suggest, however, is that if a kind of relationship is too weak in an external sense to implement CP, then it doesn’t make much difference if that relationship is internal. Suppose we had an experiment with different stimuli, organized into a different (but discernible) perceptual space, but with exactly the *same* setup and indexical instructions. Here, the *very same* behavioral structure, and the very same beliefs, lead to very different percepts. But if this is possible, then there just is not a very close relationship between the belief and the perceptual categories that are developed—it is just like the box case above.

Note the similarity here between this situation and the more mundane relationships discussed earlier. If my dad had told me to look in the black wrench case, and it turned out there were marbles (or whatever) in there instead of wrenches, I would have perceived those. If you tell me how to do the Magic Eye, and there’s a unicorn there rather than a hippo (or whatever), the instruction isn’t informative of that difference. Here the situation is the same. The content of the instruction/belief doesn’t say anything about “A-type” faces rather than “B-type” ones. It provides a behavioral instruction to lump some exemplars together, so that the category can then be learned. Given its similarities to more mundane cases, there is no reason to consider this a case of CP.

The second type of instruction, which is a better *prima facie* case to implement CP, involves descriptions of the category, including descriptions of certain of its perceptual characteristics. Indeed, the direction of attention to diagnostic features does seem important for learning categories, as Stokes (2014) suggests. However, reading this importance in terms of CP faces a problem: in order to be useful to learning, attention needs to be cued to features that *subjects can already perceive*. Otherwise, they cannot help the learner fixate on the appropriate type of object in order to learn about it. But if the beliefs they form beforehand are about perceptual content they can *already* represent, and the category they subsequently learn is a category that they *could not previously* represent, then there is a significant gap between the content of the belief and the content of the learned perception. I will argue that this gap undermines any significant interpretation in terms of CP.

To flesh out the problem take an aesthetic case: suppose you’ve never seen a Mannerist work of art before, and I send you into the museum with some instructions for how to find examples. I



might tell you that figures in Mannerist works have *elongated* features, which often occur in *contorted positions* arranged in *dramatic* scenes (consider, for instance, Giambologna's *Rape of the Sabine Women* as a paradigm Mannerist work). The point is that these features will be of precisely zero use to you unless you can already recognize them. Given the list of features, you can of course look for objects that instantiate the conjunction. But this isn't the whole story: you need to not only focus on objects that possess the right features, but develop, and possibly stretch, a dimensional structure that will allow you to generalize. Perceiving Mannerist works involves grouping Cellini and Giambologna together, despite that they are not in the same place on either the elongated-feature or contorted-position dimensions. Same goes for Pontormo and Parmigianino. Hence, the content learned in the perception is distinct from the content named in the instructions and beliefs one has prior to learning.

There are scientific cases that show this gap as well. Take Biederman and Shiffrar's (1987) classic example of teaching novices how to perceive the sex of day-old chickens. Chicks can be told apart by some distinguishing characteristics of their genitals—there is a small bulb on the genitals which is concave in females and convex in males. Interestingly, training in this field normally consisted in indexical instructions—novices would pick up a chick, say what sex it was, and then get feedback from an expert. Biederman and Shiffrar showed, however, that there are some advantages to descriptivist instructions. They showed novices (i) a diagram with caricatured male and female genitals, (ii) told them *where to look* on the chicken for these features, and (iii) told them that *convex* genitals are male and *concave* ones female. Novices' performance improved *immediately* upon getting these instructions.

In another example, Sowden et al. (Sowden, Davies, & Roling, 2000) trained novices to perceive abnormalities in radiographic images. Subjects were told that abnormalities showed up as dots. After training on a large number of example images, subjects' *sensitivity* to dots increased—they could identify dots at smaller contrast to the background than they could originally. In both cases we have changes to perception that seem semantically coherent with the beliefs subjects form on the basis of instruction. Thus, descriptivist beliefs seem to have a rather direct effect on learning. So why not interpret these according to CP?

Take chicken-sexers first. There are two reasons not to interpret this as an instance of CP. First, as Biederman and Shiffrar note, the visual system is *already* good at distinguishing convexity from concavity. Given that the visual system can already do this, the instruction of where to look, along with the propositional knowledge about which feature leads to which category, can explain the increased performance. This is just a where-to-look instruction, which as we saw above is not an instance of CP.

Secondly, there is also a gap between these instruction-mediated judgments and *full* expert performance. Experts interviewed by Biederman and Shiffrar stated that the full learning curve for expert perception involves the ability to quickly perceptually categorize an ever greater number of *subtypes*, including those that are very subtle—i.e., a dimensional sensitivity for finer-grained convexity/concavity judgments needs to be developed, and subtypes generalized over along this dimension. (Note the similarity between this and the dimensional discrimination in the Goldstone studies above.) But *this* enhanced ability is not entailed by the simple instruction

“look for convexity/concavity.” Here again, there is a gap between the explicit instruction and the ability that is eventually developed.<sup>10</sup>

Similarly Sowden et al. compared two distinct groups. One group was told to look for dots, then trained on exemplars with positive contrast (dots lighter than background). Another group was told to look for dots, then trained on negative contrast (dots darker than background) exemplars. Interestingly, the group trained on positive contrast dots did not gain increased sensitivity to negative contrast dots, and vice versa for the group trained on negative contrast. So again, the actual perceptual learning that occurs is not settled by the instruction—subjects in the distinct groups get the *same* instructions (“look for dots”) but the perceptual abilities they develop differ (increased positive contrast sensitivity versus negative contrast). Hence, the subjects’ perceptual category for “abnormalities” outstrips their prior beliefs about the category (namely, its consisting in dots).

Given the gap, strong views of CP are ruled out. These views say that the content of the belief provides information that perception uses to compute its novel content. However, the gap shows that the beliefs don’t have the novel content that perception develops. They have content that describes what perception can *already* perceive, but this is not the content that is at stake. *If* perceptual categories were simply conjunctions of features, then this would not be an issue. But perceptual categories are not simple conjunctions—they are holistic dimensional structures. Hence, descriptive beliefs contain no resources to inform the particular perceptual content that is developed. And if that is the case, then the strong view is off the table.

What about the weak view? Here, again, we have some previous cognitive processing (of the instructions) and belief formation (encoding the content of the instructions) that is a causal precursor to the perception, even if it lacks the content to determine the actual category. Is this a strong enough relationship to count as CP?

Yet again, however, when we consider the relationship to the mundane cases, it’s clear that the CP interpretation is not required. Let’s say you are walking through the museum with my instructions in mind. You wander into a wing and notice that all of the sculptures and paintings involve contorted, elongated figures arranged in dramatic poses. You might conclude, “Ah, this must be the Mannerist wing; *these are the objects I should study*.” In their effect on your perceptual learning, my instructions have not been different than if I had taken you to the third floor, pointed to the wing, and said “Mannerism’s over there.” In each case, I’ve just allowed you to focus on the right objects so that you can *then* study the exemplars to construct the perceptual category. This is the same as with dots and genitals—the instruction tells you to do something you can already do (recognize dots or concave shapes), and thus points you towards

---

<sup>10</sup> Pylyshyn (2003) notes the where-to-look nature of these instructions of the Biederman and Shiffrar case, but seems to think that this accounts for the entire phenomenon—i.e., that learning is not required here at *all*. This isn’t right, but it is perceptual learning, not cognitive meddling, that’s required to fill the gap. Interestingly, Biederman and Shiffrar do suggest a method for enhancing this subtype learning, but it *isn’t* by giving more explicit instructions. They suggest having one binder with many pictures of male subtypes, and a different one with female subtypes, and simply forcing novices to study them. This is a reversion to indexical instruction, which we saw above is not CP. Similarly, Sowden et al. suggest that descriptivist instruction is not sufficient, and must be combined with *extensive training on exemplars*—i.e., extensive opportunity for perception itself to learn the novel category.

the right objects, at which point perceptual learning *does just what it would otherwise do*—namely dimensional association and morphspace stretching.

I think Brooks (1998) explains the issue quite correctly when discussing perceptual training in the medical field: “Verbal labels are not in any sense exhaustive. A conjecture we could make is that the vocabulary actually used is a compromise between a level of generality that makes communication feasible ... but sufficiently specific that it points adequately to the referent. In this sense, *the technical vocabulary is a signal for the student to begin learning*” (p. 20, emphasis mine). Instruction, that is, and the beliefs formed on the basis of it, are a *starting point* for learning. We are directed what to look at, and how to look at it, by instruction, and it is important that we have these instructions in becoming perceptual experts, whether in X-rays or artworks. But we can explain this fact with a more sophisticated view of perception and perceptual learning, along with the mundane (and universally acknowledged) notion that instruction sets up behavioral precursors. It doesn’t require CP of any sort.

Before moving on, it is important to focus on the role of attention. Stokes (2014) suggests that the pathway between beliefs and percepts is mediated by attention—that beliefs influence attentional capacities, and these in turn influence how we categorically perceive objects. Stokes is right that attentional abilities are important parts of perceptual categories, but he is wrong in thinking that categorically *distinctive* attentional abilities are shaped by cognition. Chua et al. (2015) have studied *holistic attentional effects*, which emerge in perceptual learning. For instance, *once* subjects have learned a holistic perceptual category for types of random objects (“Greebles”) there are attentional costs in substituting parts—one cannot *ignore* a part of an object that is part of a holistic perceptual representation, even if that part is irrelevant for a particular task. The important point is that these characteristic attentional effects only emerge *after* the development of the perceptual category. Any instructions for how to identify particular categories are available equally to both novices and “experts,” but the effects don’t occur for novices. Hence, the instructions don’t explain the attentional differences that are distinctive to the learned category.<sup>11</sup>

Of course, none of this is to deny the importance of the instruction. Instructions are important because, especially in perceptually complex and culturally rich environments, there are just a *lot* of potential categories floating around. Instructions for how, where, and what objects to look at provide an important precursor to learning by focusing our attention on some sets of objects at the expense of others. But all this does is constrain what is taken as an exemplar of the category-to-be-learned. It doesn’t explain the novel perceptual categories. Perceptual learning does.

Suppose a novice art viewer wanders into a wing of 19<sup>th</sup> c. paintings, but all of the informational placards have been removed for cleaning. What statistical perceptual learning suggests is that, with enough study, the novice can pick up on some dimensional regularities in the set of

---

<sup>11</sup> It is possible to try to say that the instructions or beliefs are not exhaustive of the cognitive content. Maybe there are expectations, assumptions, or whatnot, going on beneath the level of propositional awareness. However, given the discussion in section 2, these claims are simply murky. If we admit the empirical fact of statistical learning in perception, it’s unclear what more we need besides how-to-look instructions and perceptual learning. And, on the view I’ve been urging, CP is unnecessary to explain these effects.

paintings *even without* any particular prior beliefs about the category.<sup>12</sup> This is not to say that beliefs and instructions couldn't be helpful. Consider what simply having the names of the painters can do. The name gives a categorical instruction to *lump these paintings together* for study. This might facilitate perceptual recognition of broader dimensional regularities across items with some superficial dissimilarities. Despite the differences in coloring and subject matter between Seurat's landscapes and his theater scenes, one can come to notice that these works similarly have figures with *flat features* and the foregrounds are generally *shaded*. It's not that statistical perceptual learning *can't* recognize these similarities or differences. It's that it might not if exemplars are not grouped in the right way. This kind of facilitation doesn't require CP—just focusing on the right objects.<sup>13</sup>

It might be the case that finer-grained perceptual distinctions are facilitated by finer-grained instructions. It might help if someone told me, for instance, how to look at a painting in order to reliably discriminate a Corot from a Daubigny. (Even having spent a fair bit of time looking at Barbizon school paintings, I doubt I would do much better than chance at this discrimination.) Being told what to do, though, is only a facilitation. *Actually* being able to *perceive* the difference requires appropriating the features into a holistic representation that reliably generalizes over Corots but excludes Daubignys, and vice versa. The instruction has no resources to inform this process.

So, neither the strong nor weak views of CP are needed to account for learning effects. Entirely mundane claims—namely, that instructions and prior beliefs inform us where, how, and what to look at—plus a recognition of the role of statistical perceptual learning, can account for socially mediated learning. I will conclude with a discussion of consequentialist approaches to CP.

## 5. Consequences

Stokes offers two related meta-theoretic considerations in favor of CP. The first is that the role of cognition in perception has some *consequences* that merit calling it CP. Occurrence of CP may problematize the way that perception generates knowledge, or the notion of theory-neutral observation. It may undermine claims of modularity for perceptual systems. The second consideration is that, if we treat CP as a kind of inference to the best explanation, we get a unifying explanation for the role of expertise and skill in perception, and for the importance of cognition in generating it.

---

<sup>12</sup> Interestingly, Stokes seems to admit this very possibility in countenancing the idea of a “natural appreciator,” who can gain some level of expertise in viewing artworks even with no explicit knowledge at all. But since he just *defines* expertise as cognitive, this comes out as CP for him. See further discussion in section 5.

<sup>13</sup> As an anonymous reviewer points out, this isn't quite sufficient. One has to see both the commonalities *and* the differences within a category in order to appreciate it. The perceptual learning approach offered here has the resources to account for this. In the Gureckis and Goldstone case above, subjects not only made dimensional discriminations between the type-A and type-B faces (the vertical line in the middle of the space), they also—and with *no feedback at all*—formed subcategories within the type-A and type-B faces that differentiated along the horizontal dimension (the horizontal line). This suggests that perceptual learning can account for both between-category and within-category discrimination, and there is some evidence that how this interplay works can be mediated depending on the training set (Tanaka, Curran, & Sheinberg, 2005). Much more would need to be said and investigated here, however, and I won't pursue it further.

As to the first consideration, unfortunately, none of the consequences *obviously follow* from the kind of view I've articulated. As Lyons (2011) has compellingly noted, the type of causal relationship involved matters a lot for putative epistemic consequences of CP. And the kind of relationship I have proposed above seems (ahem) neutral on theory-neutrality. On my view, a set of beliefs can point you in a particular direction, but you won't learn to perceive dimensional structure if it isn't there. Even morphspace stretching is relatively innocuous—sure, it warps our perceptions, but it seemingly does so to accentuate genuinely relevant dimensions, *outside* of direct instructions by cognition. It's not clear that this could never lead us in to error, but it also isn't clear what the epistemic upshot is, or whether this has anything to do with the theory-neutrality debate as traditionally discussed. Even with modularity the waters are murky. Suppose, simplistically, that beliefs put certain sets of objects into a mental “box” so that we treat their exemplars as of a kind and look for perceptual regularities within them. If perception just does *what it would otherwise do* (namely attempt to learn dimensional regularities), except over a new set of exemplars, then it arguably isn't a strong enough (penetrating!) influence to challenge modularity. It's an “external” effect on perception (Burnston, forthcoming): perhaps an interesting fact about cognitive architecture, but again one that doesn't necessarily compel interpretation in terms of CP.

There is a lot to be said about these issues on the view I've proposed, and I don't want to decree anything here. The point is that, if the strong or weak views aren't established, the consequentialist approach doesn't offer any *further* resources, on its own, for deciding the question.

Still, there are important issues at stake here. Here are two facts everyone should recognize and want to explain: (i) our cultural and social environments play a huge role in structuring our minds; (ii) perception, especially expert perception, is indeed a *skill* that takes considerable training and development. I think the scaling argument is motivated by what we can call *enabling* counterfactuals. *If it weren't* for the presence of social-cultural institutions, categories, publically available beliefs and theories, etc., we would not end up with the perceptual categories we have. One might argue that perceptual learning relying, even without the strong or weak views, on this cultural environment is a consequence we should dignify with the honorific of CP.

I fully grant the truth of many of these counterfactuals. However, there are a variety of ways to cash them out, and again there are many instances of similar counterfactuals that are obviously not instances of CP. The social institution of zoos, and my knowledge about how to exploit that institution, allow me to see giraffes. Given my everyday environment, it is highly unlikely that I'd ever see one otherwise. But my knowledge in this case is knowledge of how to *put myself in position* to see a giraffe—how to get to the zoo, get the zoo map, read the label ‘giraffe’ etc. If it weren't for both the social institution and my knowledge of how to exploit it, I wouldn't have seen the giraffe. The fact that my knowledge is an important precursor to my perception doesn't mean that CP occurs.

On the view I've been espousing, perceptual learning is a hungry process, searching for generalizable structural regularities in the environment. This is perfectly compatible with the *environment we find ourselves in* being highly culturally mediated, as it obviously is. Culturally structured regularities *are* regularities after all—artworks contain regularities that have a causal

history in the tradition of art that produces them, and, on the view I've offered, it's possible for perception to find these regularities (cf. Bullot & Reber, 2013). If we combine this kind of process with how-to-look knowledge, we can potentially generate a massive amount of categorical perceptual learning. If I want to know what a platypus looks like, I know to go to the zoo (or, more likely, Google Images). If I want to know what a Mannerist painting looks like, I know I have to go to the art museum. I have suggested that even "descriptivist" cases, where our beliefs about the to-be-learned categories include descriptions of perceptual features, can be treated as more specific how-to-look instructions, rather than requiring a fundamentally different type of description invoking CP. Importantly, this kind of knowledge allows for *scaffolding*—for the intentional employment of existent abilities (looking for contorted, elongated figures) in the generation of new ones (the ability to recognize Mannerist works). If this is the case, then a combination of perceptual learning and how-to-look knowledge can account for the massive dependence of our extant perceptual capabilities on social facts. The truth of enabling counterfactuals, along with their very significant consequences, can be accounted for without resorting to CP.

What about skill and expertise? Stokes, unfortunately, simply *defines* expertise as cognitive (see, e.g., Stokes, 2014, p. 2, fn. 1), and thus employs it in invoking the scaling argument—he suggests that CP is the best abductive inference we can make to account for the role of learning and skill in developing perception. It "unifies" these complicated processes under a single phenomenon.

Viewed through the pragmatic and empirical lens I've been pushing, the purported unifying power of CP is achieved at the expense of some pretty significant conflations and mistakes. To start, we simply miss the empirical facts about perceptual learning. Furthermore, we misplace the role of attention in generating perceptual categories (see the discussion of holistic attentional effects above). Furthermore again, we risk conflating more mundane notions, like instruction telling us how to look at something, with deeper claims that are supposed to come with a range of weighty scientific and philosophical ramifications. Further still, we risk missing out on dialectical options. Nanay (2016), for instance, asserts that the debate over whether an "innocent eye" in aesthetic perception is possible just is a debate about CP. On the view I'm proposing, perception may not be innocent of our social institutions, but this might still not be due to CP.

Lastly, and most importantly, we unify at the expense of having a detailed explanation of how the relevant expertise comes about. On the account I've offered, we have an *explanation* for both the importance of socially mediated learning for the generation of expertise, and the *difficulty* of that expertise despite all the resources cognition and culture have to offer. As Stokes admits, cognition can't simply dictate expertise to perceivers—perceptual expertise is *difficult*. It takes the right enabling conditions *and* huge amounts of practice. Practice involves *looking*, soaking it in, immersing oneself, letting one's vision develop. Unless perceptual learning plays a large role in this process, relatively independently of what cognition decrees, then we have no account for this other, more intuitive side of expertise. Compared to an explanation based on perceptual learning and how-to-look instructions, positing CP in these cases comes off as a vague and uninformative account.

And here we find ourselves in an interesting situation—one wherein the broad, seemingly compelling philosophical thesis does much less work than some mundane facts (e.g. that instructions can tell us how to look) and a better empirical understanding of the categories at issue (e.g., perceptual learning). I cannot hope to have done otherwise than to have offered a package of theoretical commitments and assumptions that differ from those who find the scaling argument compelling. I just happen to think that it's the right package—i.e., the one that philosophers of psychology should be selling to those more primarily focused on aesthetics.

## REFERENCES

- Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual review of neuroscience*, 25, 339-379. doi: 10.1146/annurev.neuro.25.112701.142900
- Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 640.
- Brooks, L. R. (1998). The development of new functional features by instruction: The case of medical education. *Behavioral and Brain Sciences*, 21(01), 20-21.
- Bullot, N. J., & Reber, R. (2013). The artful mind meets art history: Toward a psycho-historical framework for the science of art appreciation. *Behavioral and Brain Sciences*, 36(02), 123-137.
- Burnston, D. C., & Cohen, J. (2015). Perceptual integration, modularity, and cognitive penetration *Cognitive Influences on Perception: Implications for Philosophy of Mind, Epistemology, and Philosophy of Action*. Oxford: Oxford University Press.
- Burnston, D. C. (Forthcoming). Cognitive Penetration and the Cognition-Perception Interface. *Synthese*. doi:10.1007/s11229-016-1116-y
- Chua, K.-W., Richler, J. J., & Gauthier, I. (2015). Holistic processing from learned attention to parts. *Journal of Experimental Psychology: General*, 144(4), 723.
- Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science*, 55(2), 167-187.
- Danto, A. C. (1981). *The transfiguration of the commonplace: a philosophy of art*. Cambridge, MA: Harvard University Press.
- Delk, J. L., & Fillenbaum, S. (1965). Difference in perceived colour as a function of characteristic colour. *The American Journal of Psychology*, 52(2), 255-265.
- Deroy, O. (2013). Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies*, 162(1), 87-107.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge: MIT Press.
- Fodor, J. A. (1984). Observation reconsidered. *Philosophy of Science*, 51(March), 23-43.
- Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2010). Mere exposure alters category learning of novel objects. *Frontiers in psychology*, 1, 40.
- Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object representations: Not all morph spaces stretch alike. *Journal of Experimental, Learning, Memory & Cognition*, 38(4), 807-820.
- Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, 11(5), 197-203.
- Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, 59(2), 154-179.

- Goldstone, R. L., Lipka, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, 78(1), 27-43.
- Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, 130(1), 116.
- Gureckis, T. M., & Goldstone, R. L. (2008). *The effect of the internal structure of categories on perception*. Paper presented at the Proceedings of the 30th Annual Conference of the Cognitive Science Society.
- Isaac, A. M. (2013). Objective similarity and mental representation. *Australasian Journal of Philosophy*, 91(4), 683-704.
- Kulvicki, J. V. (2007). Perceptual Content is Vertically Articulate. *American Philosophical Quarterly*, 357-369.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4), 489-508. doi: 10.1348/0007126042369811
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: the role of race categories. *Journal of Experimental Psychology: General*, 135(4), 501.
- Livingstone, M. (2002). *Vision and art: The biology of seeing*. New York: Harry N. Abrams.
- Lupyan, G. (2014). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*.
- Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues*, 21(1), 289-311.
- 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.
- McMahon, J. A. (2003). Perceptual Constraints and Perceptual Schemata: The Possibility of Perceptual Style. *Journal of Aesthetics and Art Criticism*, 61(3), 259-273. doi: 10.1111/1540-6245.00111
- McMahon, J. A. (2012). The aesthetics of perception: form as a sign of intention. *Essays in Philosophy*, 13(2), 2.
- Nanay, B. (2014). Cognitive penetration and the gallery of indiscernibles. *Frontiers in psychology*, 5.
- Nanay, B. (2016). *Aesthetics as philosophy of perception*: Oxford University Press.
- Orgs, G., Haggard, N., & Haggard, P. (2013). Learning to like it: aesthetic perception of bodies, movements and choreographic structure. *Consciousness and cognition*, 22(2), 603-612.
- Purves, D., Shimp, A., & Lotto, R. B. (1999). An empirical explanation of the Cornsweet effect. *The Journal of neuroscience*, 19(19), 8542-8551.
- Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(03), 341-365.
- Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*: mit Press.
- Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7), 15-51.
- Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(3), 681.
- Seeley, W. P. (2013). Art, Meaning, and Perception: A Question of Methods for a Cognitive Neuroscience of Art. *British Journal of Aesthetics*, 53(4), 443-460.
- Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences*, 24(04), 581-601.



- Siegel, S. (2012). Cognitive Penetrability and Perceptual Justification. *Noûs*, 46(2), 201-222.
- Sowden, P. T., Davies, I. R., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 379.
- Stokes, D. (2012). Perceiving and desiring: a new look at the cognitive penetrability of experience. *Philosophical Studies*, 1-16.
- Stokes, D. (2013). Cognitive penetrability of perception. *Philosophy Compass*, 8(7), 646-663.
- Stokes, D. (2014). Cognitive penetration and the perception of art. *Dialectica*, 68(1), 1-34.
- Stokes, D. (2015). Towards a Consequentialist Understanding of Cognitive Penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception: New Philosophical Perspectives* (pp. 75-100). Oxford: Oxford University Press.
- Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological science*, 16(2), 145-151.
- Toribio, J. (2015). Visual experience: rich but impenetrable. *Synthese*. doi: 10.1007/s11229-015-0889-8
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annu Rev Neurosci*, 31, 411-437. doi: 10.1146/annurev.neuro.30.051606.094238
- Umiltà, M. A., Berchio, C., Sestito, M., Freedberg, D., & Gallese, V. (2012). Abstract art and cortical motor activation: an EEG study. *Frontiers in human neuroscience*, 6, 311.
- Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and cognition*, 27, 62-75.
- Walton, K. L. (1970). Categories of art. *The Philosophical Review*, 79(3), 334-367.
- Wu, W. (2013). Visual spatial constancy and modularity: Does intention penetrate vision? *Philosophical Studies*, 165(2), 647-669.