Social-eyes:
Rich Perceptual Contents and Systemic Oppression

1. Introduction

Over the last decade, there have been a disturbing number of high-profile incidents where unarmed Black individuals were unjustly killed by North American police officers\(^1\). One common explanation is that the fatal behaviours of these officers were caused by the perception of Black individuals as threatening (e.g. Skinner & Haas 2016). Examples of biased misrepresentation are readily available in a wide variety of other social contexts, and can affect different marginalized communities in similarly troubling ways. These phenomena pose difficult and pressing problems for philosophical theories of perception. The apparent influence of problematically biased representations (e.g. racist, sexist, homophobic, ableist) on perceptual processes should put pressure on researchers in the cognitive sciences to investigate the underlying mechanisms.

One relevant debate in the philosophy of perception concerns those who hold “thin” or “shallow” views (e.g. Brogaard 2013, Prinz 2013, Mandelbaum 2017) and those who hold “rich” views (e.g. Bayne 2009, Siegel 2017) of perceptual content\(^2\). A central issue here is determining which prior “assumptions,” or previously formed representations (e.g. evolutionarily or developmentally learned regularities, socially constructed categories, valenced associations, etc.) directly modulate to the contents of visual perception. The hypothesis that systemic oppression influences perceptual content has a long history in feminist philosophy and social epistemology (e.g. Peirce 1970, Alcoff 2007, Mills 2007), and demands much more explicit theoretical and empirical investigation in the cognitive sciences.

Rich accounts typically appeal to top-down modulation (TDM) in order to explain the relevant underlying processes. The idea is that assumptions stored more centrally in the system feedback into perceptual processing and thereby structure perceptual information, creating higher-level, or more informationally rich, representational content in perception itself. TDM is highly controversial though (see Firestone & Scholl 2016), and the failure to provide irrefutable evidence of its occurrence is one primary motivation for adopting thinner accounts of perceptual content. In what follows, I will show that systemic oppression and its corresponding

---

\(^1\) The fatal shooting of Treyvon Martin by George Zimmerman, for example, garnered much media attention due to the appearance that the incident was racially motivated (Lee 2013). The Ontario Human Rights Commission also recently released a report suggesting that Black individuals are “grossly overrepresented” in violent police interactions: https://www.cbc.ca/news/canada/toronto/ohrc-interim-profiling-report-1.4939242.

\(^2\) I’ll be focussing on visual perception in this paper, but similar cases could be made for other sensory modalities, and even for multi-modal perceptual experiences.
representational properties lead to a richer account of perceptual content that does not appeal to TDM, but rather appeals to neurobiological facts about the amygdala and its role in the visual system. In doing so, I avoid the need to settle deep issues concerning the modularity of perception. In section 2, I’ll briefly sketch the relevant issues in the philosophy of perception. In section 3, I’ll show that working out the details of how these problematic assumptions influence the contents of visual perception provides a unique conceptual route to genuinely rich perceptual contents that does not appeal to TDM.

2. The Contents of Visual Perception: Rich or Thin?

Determining just how rich or thin perceptual content is has a range of important implications for the philosophy of perception, and for cognitive science more generally. Most obviously, formulating an accurate list of the kinds of properties that are represented perceptually would contribute significantly to debates about the modularity of perception and the border between perception and cognition (e.g. Mandelbaum 2017). Drawing a clear distinction between perceptual and cognitive processes will be more plausible if we can cleanly pick out and sort which kinds of representational contents belong to which processes.

The standard story on thin accounts is that perceptual systems output proprietary information about shape, color and motion, for example, and perhaps some very basic categorization3, and this information is subsequently processed or further conceptualized or abstracted by some downstream cognitive process. In this way, thin views often imply perceptual modularity and a strict perception-cognition border. In contrast, identifying genuinely rich perceptual properties is typically taken as evidence in favour of top-down modulation or cognitive influence on perception, and thus against the modularity and informational encapsulation of perceptual systems. Here, the common assumption is that rich properties enter into perceptual representations by way of feedback from higher cognitive capacities to sensory (e.g. visual) processing mechanisms. The idea is that perceptually processed information about shape, color and motion is subsequently structured by assumptions stored downstream that infuse perceptual content with richer properties. In this way, rich views of perceptual content typically go hand-in-hand with doubts about modularity or informational encapsulation and a strict perception-cognition border. That is, genuinely rich perceptual contents are generally thought to implicate good old-fashioned TDM. This means that anyone skeptical about TDM (e.g. Firestone & Scholl 2016) seems to already have prima facie reasons to be skeptical about rich perceptual contents. The logical connection between TDM and rich contents has recently been called into question, however (e.g. Brogaard & Chomanski 2015), and alternate explanatory models are starting to be developed.

Regardless of how rich or thin an account of perception one endorses, it is important to note that every account allows that some prior assumptions work their way into the contents of perception. One recurring and illuminating example is the Light-From-Above (LFA) prior and its influence on visual perception. It is generally acknowledged that the particular shading of a

---

3 On Mandelbaum’s “shallow” account, discussed below, perceptual systems output basic-level categories and so basic-level categorization is genuinely perceptual. Fodor’s original account of modularity also assumed that visual modules output basic-level categorizations (Fodor 1983).
two-dimensional circle, for example, will reliably cause perceivers to represent the image as either concave or convex given the learned, encoded and behaviourally reliable assumption that light typically comes from a single source from above. Although deeply rooted and apparently phylogenetically ubiquitous, it is interesting to note that this prior assumption can be overridden when different light cues and other contextual information are made salient to the perceiver, changing how they perceive the world to be (Morgenstern et al. 2011). Thus, everyone agrees that at least some assumptions (i.e. LFA) work their way into the very contents of perception as genuine perceptual properties (i.e. convexity).

The LFA prior provides a baseline case for thinking about the contents of perception. On thin views, LFA is representative of the small class of prior assumptions that influence the contents of visual perception. The idea is that the modular processing of incoming visual stimuli is reliably biased (e.g. weighted) in such a way as to construct and subsequently output visual representations according to LFA (Orlandi 2014). Note that we have a feed forward story here; the only assumptions that modulate perceptual contents are those that are contained in the visual system module itself, and the only properties that find their way into the contents of perception are those that are traditionally thought to be in the jurisdiction of visual processing (luminance, shape, color, motion, etc.).

On rich accounts, we typically get a much more complicated story. There are reasons to think that more informationally robust assumptions stored more centrally in the system can influence perceptual contents as well, such as beliefs about the world that have been acquired from previous experiences (e.g. that individuals who drive expensive cars are wealthy). As visual information is processed in the visual system proper, the particular low-level properties of the stimuli activate more centrally-stored prior assumptions, which then feed back into the visual system to modulate those contents. That is, the visual system appears to output information about particular shapes, colours and motions that activate prior assumptions about, for instance, expensive cars and the people who typically own them. These prior assumptions then feed back into the visual content itself, allowing us to perceive the stimulus as a wealthy person. Accordingly, the border between perception and cognition gets blurred given this sort of top-down, or recurrent, processing. It is important to note that, while each of the proposed rich contents needs to be defended independently, allowing for some rich contents can provide the logical and mechanistic framework needed to explain other rich contents.

So, the debate seems to hinge on the order of processing from visual input, to perceptual representation, to the influences of cognitive processing. Luckily, this is an empirical question that can be settled by taking a closer look at the relevant science. One sort of evidence that can be marshalled here is information about processing speed, which ultimately provides insight into processing sequence (Block 2010, Fish 2013). Mandelbaum (2017) points out that feedback from central cognition to the visual system, and any post-perceptual processing for that matter, should operate on longer time scales than the feedforward processing of the visual system proper. Thus, if representational properties emerge at a significantly short time scale, the implication is that these properties should be the result of processing in the perceptual system proper, and not the result of TDM or any other downstream post-perceptual process.

To be clear, I do think there is a case to be made for genuine TDM. Many theorists have responded to Firestone and Scholl’s methodological challenges, providing convincing evidence of genuine top-down influences on the content of perception (e.g. Witt 2017). Also, given the
architectural organization of the visual system, and its relationships to other neural structures, it is likely that prior assumptions and stored regularities feedback from many stages of processing to structure the contents of perception in complex ways. Wilson & Wilkinson (2015) for instance, identify extensive feedback connections from higher to lower structures in the ventral visual stream that could plausibly underwrite TDM processes. However, TDM is only one possible path to richer perceptual contents, and one that is unpalatable for many precisely because it implies giving up on the modularity of perception and a strict perception-cognition border.

If we can build a plausible account of rich perceptual contents without invoking the TDM story, it ought to be persuasive regardless of one’s intuitions about modularity and informational encapsulation. The assumptions rooted in systemic oppression (e.g. racist, sexist, or ableist assumptions), and the possibility that they contribute a certain type of rich evaluative content to visual perception, provide illuminating cases for just such an account.

3. Rich Perceptual Contents Without TDM

The internalized assumptions that result from systems of oppression, and the possibility that they generate evaluative representational contents—for example, perceiving an individual from a particular social group as threatening—provide an ecologically valid, pervasive, and practically important hypothesis for determining just how rich perceptual content is. Once again, on the traditional formulation of this debate, if this sort of evaluative information is indeed part of the very content of perceptual representation, then we should expect top-down influences to be implicated.

Mandelbaum (2017) assumes as much in his recent argument for a “shallow” or thin view of perceptual content. His model relies heavily on empirical facts about processing speeds, which provide fairly convincing evidence about particular processing sequences. Because we have evidence of “basic-level” categorization of visual input at extremely short time scales, as fast as 13ms, Mandelbaum thinks we have reason to doubt that any top-down influence or post-perceptual processing could have been involved. Thus, basic-level categorization must be genuinely perceptual. Moreover, this marks the upper limit of how rich perceptual content can be, because only this sort of basic-level categorization, and not more sophisticated conceptual or evaluative capacities, are thought to temporally precede the possibility of any post-perceptual processing. This picture seems to support the idea that vision is a modular perceptual system and the influence of genuine TDM is doubtful, and that basic-level categorization is carried out in the visual system itself. This means that any other representational properties, including socially-rich evaluative properties, would have to be constructed as a result of post-perceptual processing, and thus fail to count as genuinely perceptual. On this account, socially-rich evaluative contents are properly understood as belonging to cognitive processes, and the perception-cognition border is maintained. However,

---

4 This refers to the sorts of basic object categories that Rosch (1978) postulated as the categories in folk taxonomies that facilitate our most common cognitive tasks. They include categories like chair (as opposed to the subordinate category kitchen chair or superordinate category furniture) because the former is thought to be more useful for everyday cognitive tasks, and thus more ubiquitously employed.
using Mandelbaum’s own argumentative strategy, we can show that visual perception represents richer contents without having to settle the issue of whether downstream cognitive processes subsequently feedback in such a way as to genuinely modulate perceptual contents.

What makes the internalized assumptions stemming from socially maintained systems of oppression a particularly illuminating case study for working out the contents of perception is that the psychological and neurological mechanisms of social bias have recently been the subject of increased scientific interest, and a few distinct underlying processes have been isolated (e.g. Amodio 2014). There are processes underlying the rapid valencing of incoming sensory information, for instance, that appears to susceptible to these social influences. Specifically, there is a wealth of literature on the role that the amygdala plays in emotional assessment thanks to its underlying neurological properties (e.g. Beyeler 2018). Assessments of valence (i.e. positive or negative value or relevance for the organism) and intensity (i.e. how significant that value or relevance is) of incoming stimuli are widely thought of as the psychological primitives that form the core of all emotional processes (e.g. Barrett and Wager 2006). There is also compelling evidence that this function of the amygdala is involved in the biased evaluative judgments that are systematically made about members of particular social groups (Phelps et al. 2000, Amodio 2014).

The relevant processing speeds are important here. Mandelbaum cites evidence that he takes to show that rapid semantic categorization can occur as fast as 13 ms, almost a tenth of the time of the shortest recorded blinks (100 ms). Recurrent processing within the visual system is estimated to take at least 50 ms, and so if vision is in fact modular, we can assume that 50 ms is the upper limit for any property to be constructed genuinely in perception, and not post-perceptually. We can also expect TDM and any other cognitive processing of perceptual information to be slower than this. One study places the top-down influence of cognitive processing on visual stimuli in the Implicit Association Test in the range of 200-350 ms (Healy et al. 2015). Again, processing speed is being used as a window into processing order, based on the straightforward logic that earlier processes happen more quickly, and TDM and post-perceptual processing takes significantly longer than even the most elaborate recurrent processing within the visual system.

Crucially, the evidence suggests that the amygdala performs evaluations of incoming visual stimuli at extremely rapid speeds. Studies combining MEG and MRI methods have shown amygdala reactivity at 20-30 ms post stimulus onset (Luo et al. 2007), while other studies using intracranial electrophysiological recording techniques show greater relevant activity to fearful than to neutral faces as early as 50ms (Sato et al. 2011). Researchers have generally taken this as evidence that the amygdala is “an early detector of the relevance of stimuli for the individual” (Bonnet et al. 2015). In this way, the amygdala plays a direct role in shaping the representational content processed in the visual system.

The speed of this rapid evaluation is further supported by the underlying neuronal architecture. The amygdala receives direct input from subcortical visual pathways (Diano et al. 2017). Information from the retina follows a feedforward subcortical pathway via the superior colliculus and pulvinar inferior, ultimately to the amygdala before reaching V1 and higher visual cortical centres, presumably in the service of the rapid evaluation of visual input (see Figure 1). This architecture makes sense evolutionarily speaking; the assessment of the positive or negative value of a visual stimulus needs to be rapid. If the stimulus is a predator, for example,
there is no time to allow for TDM or post-perceptual processing of the incoming visual stimulus. The basic idea is that this rapid evaluation can quickly alert the organism to the potential threat or benefit of some incoming visual information, which can either quickly activate reflexive behaviour or draw attentional resources for further visual processing and conceptualization, depending on the intensity or significance of the information. The amygdala’s connectivity within the visual system further supplements the argument that its assessments are genuinely part of perception.

Figure 1 (Diano et al. 2017)

So far, we have good evidence that evaluations by the amygdala are part of perceptual processes, but why think of them as producing rich contents? The final piece of the puzzle is to get clear on how amygdala reactivity is “tuned” in the first place. There is evidence that amygdala reactivity is tuned by a complex range of processes throughout the organism’s life in ways that calibrate its rapid evaluations of valence and intensity. Recent developmental studies have shown that greater emotional responsiveness in an individual’s upbringing reliably increases amygdala reactivity to interpersonal threat (Farber et al. 2018). This is evidence that complex emotional information derived from one’s social contexts over time is reflected in how the amygdala generally responds to visual input and contributes to the visual system’s output. Studies by Phelps et al. (2000) also suggest that amygdala reactivity is reliably correlated with indirect measures of racial evaluation. They conclude that amygdala reactivity reflects “cultural evaluations of social groups modified by individual experience” (Phelps et al. 2000). In this way, complex social information, like the temporally extended narratives rooted in systems of oppression, tunes the amygdala, which determines the complex social information that will be represented in the very contents of visual perception.

The sort of information carried by socially-rich evaluative contents reflects rich social input that cannot be captured by lower-level information about shape, color and motion—and this is typically how rich content is defined (e.g. Seigel & Byrne 2017). Indeed, assigning valence to certain stimuli in general seems to be paradigmatic of rich representational content: these evaluations are generally thought of as involving the addition of a different sort of information (e.g. sometimes described as flavouring existing representations with “hedonic tone”) to perceptual content (e.g. Barrett 2006, Zhang & Li 2018). When these evaluations result from the temporally-extended intake of complex social information, there is overwhelming reason to think of them as rich contents.
4. Conclusion

Systemic oppression and the evaluative perceptual contents that it generates make the case for rich visual contents without appealing to TDM, and hence without needing to settle more general debates about the modularity of perception and the existence of a strict perception-cognition border. The socially-rich evaluations of visual information carried out by the amygdala operate at such rapid timescales that it is plausible that the resulting evaluative contents are indeed genuinely perceptual. This is further supported by the nature of the connectivity of the amygdala within the visual system.

References


