Cognitive Penetrability: Modularity, Epistemology, and Ethics

Zoe Jenkin[1] · **Susanna Siegel**[1]

If you are picking blueberries and you think that fat berries are best, you'll be less likely to overlook them. Here, perception is influenced in a humdrum way by beliefs and desires. Your desire for fat berries makes you search for them, and your beliefs about where they are likely to be influences where you look when you search.

Can perception be influenced in more radical ways by what we expect, fear, want, or know? Could your desire for fat berries cause you to see the berries as fatter than they actually are? Could your expectation that more berries are to be found on the bush lead you to hallucinate berries where there are none, or to mistake nibs of the bush for berries?

If such influences are possible, they give us reason to question whether there is a distinct division in the mind between the perceptual systems and the cognitive system. In *The Modularity of Mind* (1983), Jerry Fodor made the case that there is such a division. He proposed that perceptual input systems and the language processing system are informationally encapsulated, whereas cognition is not. Fodor argued that the mind contains *modules* – informational systems that take in a specific kind of information and no other kind, in response to a highly specific set of triggering cues.[1] For instance, the visual module (or a color module within vision, if vision includes multiple modules) takes as input the ambient light array, and outputs color representations. The way that modules process information is encapsulated, in the sense that it cannot be influenced by information stored outside the system. Since it was published in 1983, Fodor's book has spurred debate about whether the mind contains any

---

[1]Fodor originally outlined nine criteria for modularity: domain specificity, mandatory operation, limited central accessibility, fast processing, information encapsulation, shallow outputs, fixed neural architecture, characteristic and specific breakdown patterns, and characteristic ontogenetic pace and sequencing (Fodor 1983). We focus on information encapsulation and domain specificity here, although some challenges to modularity can be cashed out in terms of mandatory operation.

✉ Zoe Jenkin
zoejenkin@g.harvard.edu

Susanna Siegel
ssiegel@fas.harvard.edu

[1]   Harvard University, Cambridge, MA, USA

modules at all, and if so, which information is processed in modules, (Churchland 1988, 2006; Pinker 1997; Prinz 2006; Firestone and Scholl 2014, 2015) and how modular processing relates to learning (Carruthers 1996, 2006; Carey 2010).

The idea of a division in the mind between perception and cognition is crucial not only to psychology, for establishing our accounts of cognitive architecture and processing systems, but also to our philosophical theorizing about the contours of the mind, and how it is related to the world. If perception is informationally encapsulated and operates in a solely "bottom-up", or stimulus-driven, way, it seems to provide a direct channel from the world to our experiences. The directness of this channel would allow our experiences to provide a check on our beliefs, as they face the "tribunal of sense experience" (Quine 1951), as Quine put it.[2] If in contrast perception is subject to "top-down" influences from previously held cognitive states, we no longer have this secure and unmediated link to the external world, because our perceptions themselves may be reflecting facts about our own cognitive make-up, even when they seem to be reporting on the world around us. Perception is often thought to be the ultimate source of epistemic and ethical justification, and so cognitive penetrability calls the foundation of some normative theories into question as well.

Top-down influences are possible in principle. Whether they actually happen is a question for psychology. But there are in fact multiple possible phenomena in the vicinity, and so there are multiple questions for psychology to answer. The phenomena differ depending on what aspect of perception is influenced. Answering these psychological questions is of particular importance for philosophy, because the details of the stages, mechanisms, and sources of influence may determine what kinds of challenges they pose to theories in the philosophy of mind, as well as to normative theories including ethics, epistemology, and aesthetics.

The papers in this special issue of the *Review of Philosophy and Psychology* explore many of the philosophical facets of the varied phenomena known as 'cognitive penetration'. To orient the reader, we offer a guide to research questions in this area. Psychological research is making it both more possible and more pressing to address these questions. We distinguish between different kinds of influences on perception in section 1, and then we outline major questions about cognitive architecture discussed by some of the authors in section 2, and provide an overview of the challenges from putative cases of cognitive penetration to theories in epistemology (section 3) and ethics (section 4).

1 Varieties of Extra-Perceptual Influences

We can distinguish among various kinds of extra-perceptual influences on perception using two independent parameters. First, we can distinguish between potential influences on the mechanisms through which influences operate, such as direction of focus,

---

[2] Of course, perceptual error is possible on this picture, due to low-level malfunctions of parts of the input systems. Information encapsulation does not guarantee that every perception we have is perfectly accurate. However, the sorts of errors that would be possible on this picture would not preclude perceptions from playing the role of providing a check on beliefs, because the source of the perceptions is something entirely external to the cognitive system.

or attention of various kinds. Second, we can distinguish different stages at which influence occurs – regardless of which kind of attention (if any) mediates it.

Mechanisms

- **Focus:** influences on where you direct your gaze.
- **Object-based or spatial attention:** influences on which parts of objects or regions you consciously or unconsciously attend to, holding the direction of gaze fixed.
- **Feature-based attention:** influences on which features you attend to, holding the objects and regions fixed.
- **Other:** influences via mechanisms other than focus or attention.

The types of influence listed here differ according to the potential mechanisms by which extra-perceptual factors can influence what a subject perceives. Visual attention is a necessary part of the functioning of the visual system, and visual perception would not be possible without some sort of allocation of attention. Perceptual attention by itself is not an extra-perceptual influence on perception. But distributions of attention can be determined by factors external to the perceptual system, as well as by mechanisms internal to perception. To fix ideas, here are some potential examples of each kind of influence.

When influence operates via focus, where you direct your gaze helps determine which inputs are selected for perceptual processing in the first place. An example of this first type of effect might be subjects' tendency to follow the gaze of other people in social in-groups, but not out-groups (Adams and Kveraga 2015). If different inputs are selected for processing by this sort of attentional mechanism, the content of your perceptual experience will also turn out to be different.

In another sort of attentional effect, while your gaze is held fixed, spatial attention is allocated in different ways across your visual field. For example, some perception of ambiguous figures can be influenced by expectations via spatial attention. If you expect to see a duck when you look at a drawing of a Duck-Rabbit, you may focus on parts of the image so that it looks like a duck. In contrast, if you expect to see a rabbit, you may attend to certain parts of the image so that it looks like a rabbit (Macpherson 2012). There may be other factors that help explain our perceptual shifts when looking at ambiguous figures as well, but allocation of spatial attention plays an important role.

When influences operate via feature-based attention, an extra-perceptual state influences which features of objects or regions one attends to. For instance, one could focus on an object's color instead of its size – as you might if you wanted the bluest berries, regardless of their size.

If you attend to an object, region, or feature, something is there to be attended to. This list gives us no way to describe inaccurate perceptual states that result from such influences. A perceptual state can be inaccurate because it represents objects that aren't there, or because it attributes to an object a feature that the object doesn't have. To allow for influences that result in falsidical perceptual states, we need a different set of distinctions.

Instead of focusing on attentional mechanisms by which influences can operate, we can abstract away from any mechanisms by which the influence occurs, and distinguish between stages at which perceptual experience or judgment can be influenced. The

second list divides into stages the processes up to and including conscious perception (which we call 'perceptual experience'), and responses to perceptual experience.

Stages of processing

Perceptual processing

- influences on early vision.
- influences on the contents of unconscious or pre-conscious perceptual states and processing at stages after early vision.
- influences on the contents of perceptual experience at stages after early vision.

Responses to perceptual experience

- influences on what you introspectively judge the contents of your perceptual experience to be.
- influences on non-introspective conclusions you draw from perceptual experiences.

The entries on this list leave unspecified what the mechanism of influence is. At any of these stages, influences could in principle operate through attention.[3]

Effects on early vision consist of influences on the course that visual processing takes at the initial stages, when generating basic outputs such as shape, location, motion, and color (Pylyshyn 1999).[4] This sort of influence, as well as the sort of influence in all the subsequent types, might be manifested by changes in how feature-based attention is allocated—that is, how the various aspects of the stimulus inputs are weighed in processing. This sort of attentional effect is a type of modulation of perceptual processing via attention that is not as simple or as early as a selection effect (See Lupyan's and Macpherson's articles in this issue for further discussion). Influences on early vision might also include ones that do not operate via attention, but instead by adding new information into the system. For example, if you expect that banana-shaped objects will be yellow, you may classify a grey banana as slightly yellow (Witzel et al. 2011. For discussion of whether this experiment establishes an effect on perceptual experience, see the papers in Zeimbekis and Raftopoulos (2015)). It seems unlikely that this is an influence on how much attention is given to particular features of the input, because there is no yellow input at all to be had. Instead, information that is plausibly external to your perceptual system[5] is relied upon in order to generate a perceptual experience with a content that is different from what it would have been, absent that influence.

---

[3] If influences on introspective judgments operated via attention, a further question would be whether it is attention to experience, or to external things. On whether it is possible to attend to one's own experiences, see Smithies and Stoljar (2012).

[4] 'Early vision' is used in a variety of ways. It can denote (i) processing that occurs in a particular neuroanatomical area, such as the primary visual cortex, (ii) processing with outputs of a certain basic type, or (iii) a part of vision that is functionally isolated from cognition (as in Pylyshyn 1999). Here, we use the term in way (ii), to refer to the initial stages of visual processing that produce representations of visual properties such as shape, location, motion, and color.

[5] Although the influence that is responsible for these color memory effects is also plausibly from information that is stored within the perceptual system itself. See Deroy (2013) for further discussion of this possibility.

Other influences may occur in processing stages that are subsequent to early vision. These include any influence on the processing of features that are beyond the outputs of early vision. For example, your fear of snakes might cause you to see a rope as a snake. If being a snake and other natural kind properties can be part of the contents of perception (Siegel 2010), then such influences might affect the contents of experience, and they might occur at stages in perception, after basic features such as the long, coiled shape and rapid movement pattern are processed. In contrast, when influences at these later stages affect states that are not conscious, we do not directly experience the effects of this type of cognitive penetration. They might, however, have implications for other aspects of our mental lives that we do eventually experience. If you unconsciously perceive a rope as a snake due to a fear of snakes, this might lead to a behavioral response such as flinching away, even if you cannot introspectively tell why you moved. One could also experience the results of the influence directly and immediately, even if one is not aware of its source or even of the fact of its occurrence. If you consciously perceive a rope as a snake due to your fear of snakes, it will seem to you as if there is a snake before you, even though you may not be aware that this perceptual experience is caused by anything outside your perceptual system at all.

In contrast to influences on perceptual processing, there are also potential influences on transitions from perception to other states, including perceptual beliefs. Returning to the blueberry picking case, here is example of an influence on what we take our experience to be: you see the berries as just as green and tiny as they actually are, but when reflecting on your own experience you judge that you saw them as blue and fat. In contrast, here is an example of an influence on external-world beliefs made in response to our experiences: you see the unripe blueberries accurately as green and tiny, but conclude that they were ready for pie-baking nonetheless. In both cases, the influence occurs exclusively in the response to the experience, rather than in the production of the experience.

We can now see what makes the phenomena on both lists belong to the same family. For any pattern of behavior that is prima-facie well-explained by effects on perceptual experience, judgments made in response to experience (either introspective judgments or judgments about the external world) are potentially competing explanations, while effects on earlier stages of perceptual processing, or on mechanisms, or both, are potential supplementary explanations.

So far, we have focused on the type of state that is influenced by extra-perceptual factors, and on the mechanism of influence. We could also ask, for any of these phenomena, what type of psychological state does the influencing? When the influencing state turned out to belong to another sensory modality, the effect is often referred to as a cross-modal effect, and in one sense it would not be an 'extra-perceptual' effect after all.[6] When the influencing state is cognitive, and influences perceptual experience or pre-conscious perceptual processing, the phenomenon is often known as *cognitive penetration*. Since 'cognitive' is often used loosely, and since 'influence' comes in many forms, the exact contours of cognitive penetration are hostage to further precisification.

---

[6] In another sense it might count as an extra-perceptual effect, if 'perceptual' were being used to refer to the workings of one dedicated sensory input system. However, if 'perceptual' is used to refer to the various perceptual systems taken together, cross-modal effects would count as intra-perceptual.

The papers in this special issue collectively discuss all of the phenomena in the family. Taken together, the papers collected here address the psychological evidence favoring all of these possibilities, the implications these influences have for cognitive architecture, and the significance of top-down effects on perception in ethics and epistemology.

2 Cognitive Architecture

In different ways, the contributions to this volume by Gary Lupyan, by Reginald Adams and Kestutis Kveraga, and by Anya Farennikova challenge the architecture of the mind proposed by Fodor in *The Modularity of Mind*. Lupyan draws on an alternative architecture (called 'predictive coding'), and all three contributions discuss specific phenomena that the authors take to challenge Fodor's picture. Lupyan focuses on interactions between language and perception as well as interactions among sensory modalities, Adams and Kveraga focus on social perception, and Farennikova focuses on the perception of absences.

In "Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems", Lupyan makes the case that information stored anywhere in the mind can bear on the contents of perception, depending on how relevant that information is to a particular task. If this picture is right, then the implications for cognitive architecture are vast. It would imply not only that cognitive states can on occasion alter perceptual processing, but that they always do, because the contents of perception are generated based on these system-wide priors. On this picture, there is no proprietary store of information that is distinctive of perception, or of any particular sensory modality. The same body of information can bear on the formation of perceptual and cognitive states. Since many ways of drawing the distinction between perception and cognition invoke precisely this kind of difference, Lupyan's conclusion would entail that no such difference between perception and cognition can be found. Any attempt to distinguish between perceptual and cognitive processes would need to draw on some criterion other than the information those processes take in.

Even if perception and cognition are not distinguished by proprietary stores of information, other versions of the distinction may track explanatorily important divisions in the mind. One option for this distinction is the format of a representational state. On some views, perceptual representations are iconic, whereas cognitive representations are sentential in format.[7] Distinguishing perception and cognition in terms of format seems compatible with rejecting a distinction in terms of information processing. On this picture, the informational content of all mental states might be generated by the same sort of mechanism, as Lupyan suggests (see also Hohwy 2013 and Clark 2013), but that information might be presented or stored in a different format, depending on whether it was perceptual or cognitive.

---

[7] Burge (2014) states that all perceptual states are iconic, and all conceptualized cognitive states are propositionally structured (although he leaves room for some iconic cognitive states, such as mental imagery, or pre-conceptualized states). Block (2014) endorses Burge's picture. Carey (2010) says that perception is generally iconic and cognition is generally propositional, and that the outputs of the object module, which lie at the border of perception and cognition, are likely iconic.

Lupyan's main conclusion, if true, would erase one distinction between perception and cognition, replacing it with a single kind of processing. But it could leave other versions of the distinction intact, and relative to those distinctions, Lupyan's conclusion could be adapted by saying that although cognition and perception are distinct, the perceptual states one ends up with can depend on cognitive factors. In her commentary on Lupyan's paper, Fiona Macpherson discusses further what kind of top-down effects Lupyan purports to give evidence for. She examines the relationship between the predictive coding model of visual perception that Lupyan endorses and top-down effects on perception.

Adams and Kveraga challenge Fodorian modularity from the domain of social perception. In "Social vision: Functional forecasting and the integration of compound social cues", they explore the impact of emotion and social categorization on face processing. They summarize experimental results that suggest that we are more likely to have fear responses to certain facial expressions or face-types than others. For example, they discuss how race can have a major influence on the ways in which other social cues are processed and interpreted. It has been widely documented that humans and other primates follow the gaze of their conspecifics to direct their attention across all sorts of scenarios. Adams and Kveraga present experimental evidence that in the United States, the extent of gaze-following behavior is often sensitive to race and social power. Not always, but often enough to be interesting, Americans of European descent (White participants) followed the gaze of White faces but did not follow the gaze of Americans of African descent (Black participants), whereas Black participants followed the gaze of both groups of faces. To the extent that gaze-following indicates confidence that the followed-person's object of attention, or experience of it, is epistemically valuable, it is reasonable to hypothesize that this result reflects an underlying pattern of social valuation, and specifically the epistemic under-valuation of Black adults by White adults.

In the family of top-down effects, the cultural effect on gaze-following that Adams and Kveraga describe is an effect on focus, via direction of gaze. By itself, an effect of this type does not challenge modularism, because the system may be entirely encapsulated, yet still generate different outputs, simply due to the different initial inputs. For that reason, some researchers would exclude influences on focus via gaze direction from the class of cognitive penetration as they define it. They want to preserve the label "cognitive penetration" for phenomena that threaten modularism. The social significance of the effect shows that the kinds of top-down effects that matter for understanding politics and culture do not always coincide with those that matter for understanding debates about modularism.

However, Adams and Kveraga argue that the best explanation of the gaze-following results is that systems previously thought to be informationally encapsulated are in fact informationally integrated. They argue that top-down integration of socially relevant information such as race and gender can modulate the course of visual processing at the earliest stages. For example, when a face is categorized as belonging to a member of a group with privileged status, visual attention is oriented in a different manner than it would be absent this categorization. Similar effects have been found for other aspects of face processing—detection of emotion, fear responses, and quality of memory all seem to be dependent on our cognitive make-up. Even feed-forward processing, they argue, may rely on

stored associations that map incoming information to social categories, integrating cognitive information with perceptual input.

Based on a wide array of experimental results, Adams and Kveraga propose a comprehensive top-down model of social perception according to which contextual effects and biases can modulate the contents of perceptual experience by three different mechanisms: 1) bottom-up integration of social cues, 2) unconscious integration of divergent cues, and 3) top-down influences on shared associations across cues. They argue that their model entails that perception is cognitively penetrable. In his commentary on Adams and Kveraga, Francesco Marchi argues that Adams and Kveraga's own experiments and arguments leave open the possibility that cognitive factors make a difference to post-perceptual judgments but not to perceptual experience. (Using the list of top-down effects at the start, he is arguing that the evidence does not decide between effects on stages of processing.) Marchi tries to argue directly for cognitive influences on perceptual experience, by interpreting experiments by Carroll and Russell (1996) and Levin and Banaji (2006), in which subjects' judgments about the race or emotional expression of an ambiguous face vary depending on what look to be cognitive factors.[8] Marchi argues that Adams and Kveraga's proposed mechanisms could explain these sorts of effects.

Josefa Toribio's commentary on Adams and Kveraga focuses on their key assumption that perceptual experiences can represent high-level social properties such as race, gender, and emotional expressions, in addition to low-level properties like shape, color, and motion. Adams and Kveraga claim that experiences represent high-level social properties on the grounds that processing proceeds differently in vision at very early stages depending on which social features are present. Toribio argues that this evidence does not settle the question of whether the high-level contents are contents of perceptual experience, as opposed to contents of judgments downstream of perception.

How can we figure out which type of representation has the high-level content? Toribio suggests that a property is part of the contents of perception only if the representation of that property is directly controlled by features of the visual input. Applying this criterion to Adams and Kveraga's results, she argues that the best explanation of their data is that the contents of subjects' perceptual experiences include some high-level emotional and social properties. Toribio thinks those properties change in response to changes in the stimulus in a way that is the hallmark of visual processing.

In "Perception of Absence and Penetration from Expectation", Farennikova situates the modularity debate with respect to perceptual experiences of absences. She argues that we can perceive the absences of objects, in addition to perceiving objects and their properties. For example, she argues that it is possible to see the absence of your keys on the shelf, if you expected them to be there. She makes the case that absence perception is implemented by visual search mechanisms involving templates for perceptual scenes. The templates are constructed from past experiences. When we detect a mismatch between the expected template and the scene before us, we perceive an absence.

How does absence perception relate to debates surrounding modularity? Because absence perception relies on expectations, Farennikova argues, background beliefs play

---

[8] However see Firestone and Scholl (2015) for an argument that the apparent effect in Levin and Banaji (2006) is entirely explicable in terms of differences in low-level features of the stimuli, and so does not entail the cognitive penetrability of visual processing.

a role in visual processing, and so visual processing cannot be wholly informationally encapsulated. She distinguishes the sort of cognitive penetration that is involved in absence perception from the sort that is traditionally discussed in philosophy of mind and psychology, highlighting how expectations facilitate object perception in successful searches, and create absence perception in failed searches. Farennikova's research on absence perception provides a new angle on the classic modularity debates.

Lupyan, Adams and Kveraga, and Farennikova's papers begin to explore novel questions for the philosophy of mind that are raised by psychological results on cognitive penetrability. Questions in philosophy of mind that are ripe for further research include:

- If there is cognitive penetration, what are its limits, if any?

– Even if cognitive penetration does occur, could we start with a largely modular mental architecture and specify conditions under which information encapsulation is violated to come up with an accurate picture of the mind? Or is a modular picture fundamentally mistaken?
– What are the rules, if any, that govern when cognitive states influence perception and when perception operates independently from cognition?

- What is the relationship between cognitive penetration and cross-modal influence? Is cross-modal influence always cognitively mediated, or can it operate independently of cognition?
- What is the relationship between cognitive penetration of perception and cognitive penetration of language processing? Could one system be encapsulated and the other not, or is perception such a crucial part of language processing that the penetrability of both systems stand and fall together?
- What exactly are the mechanisms by which cognitive states influence perceptual states? How is information transferred from one system to the other, especially on views on which perceptual and cognitive states have different sorts of representational formats (iconic and discursive, respectively)?

3 Epistemology

When Fodor argues that perception is modular in *The Modularity of Mind*, a center-piece of his argument is the idea that the role of perception is to take in information without regard for what one knows or believes. Jona Vance examines this idea in his contribution to this volume, "Cognitive Penetration and the Tribunal of Experience." Vance observes that if one takes this idea about perception as a fixed point, one might be led to think that any influence of background beliefs on perceptual experience could reduce the epistemic power of experience. In particular, one might think that if background beliefs about what colors or shapes things are likely to have influenced the colors and shapes we perceived them as having, then we'd be cut off from perceiving the actual colors and shapes of things. Perception would then be epistemically less powerful, in a sense that Vance develops, on behalf of proponents of this line of thought.

Vance sets out to show that this line of thought is mistaken. He describes a Bayesian framework for understanding how background beliefs could regulate perceptual experiences, and argues that there are no general grounds for finding epistemic fault with the resulting perceptual experiences.

Vance's arguments bring into focus a larger question in epistemology: what kinds of mental states can be rationally evaluated? For instance, if a person S believes that Obama was born in Kenya, we can evaluate how rational S is in holding that belief, if we find out more about what led S to have it. In contrast, a hiccup does not reflect in any way on the rationality of the subject. Which mental states belong with beliefs to the class of states that bear on the subject's rationality? A long strand of 20th century Anglo-American analytic philosophy (with some precedents in the early modern period in Hume, Descartes, Locke, and Reid) held that transitions from perception to belief were merely causal, and only transitions beginning with beliefs were evaluable as reasonable or unreasonable. Later on in the 20th century, a more liberal picture emerged. It held that transitions from perception to belief could be reasonable or unreasonable (Pollock and Cruz 1999). But it stopped short of allowing that perceptual experiences themselves, or the processes that generate them, could be reasonable or unreasonable.

The idea that prior psychological states can influence our perceptual experiences raises the question of whether those influenced experiences could be rationally evaluable due to the relationship with these prior states. Vance's conclusion suggests that they can be (see also Vance (2014)). In cases like the blueberry picking one described at the start, the pattern of wishful seeing seems structurally analogous to wishful thinking. Since wishful thinking is often epistemically irrational, wishful seeing that arises in analogous ways plausibly is, too. More generally, once a perceptual experience is an output of a process that would make a belief unreasonable, we can ask whether the epistemic profile of the perceptual experience becomes more like that of belief.

Other epistemological issues raised by the possible influences on perception outlined at the start raise questions in moral and legal philosophy. In the law, the reasonableness of one's assessments of one's personal safety helps distinguish illegal aggression from legal forms of self-defense. When would be it reasonable to feel fearful in a social situation? On the face of it, how reasonable it is to feel afraid is modulated by how things look. But if how things look is in turn influenced by what one already fears, then does the reasonableness of fear depend only on the actual situation, or only on how the subject perceives the situation, or both? If the way a subject perceives the situation is itself epistemically evaluable, then the reasonableness of fear can depend on how the subject perceives the situation, and that perception can itself be reasonable or unreasonable.

Questions regarding the epistemic impacts of cognitive penetrability include:

- What are the criteria for rational assessability of a mental process? Can cognitive penetration bring the etiology of a perceptual process into the realm of rational evaluability?
- Is there a difference in the epistemic impact of cognitive penetration of perception and of influence on perception by information stored within the perceptual system, either crossmodal or intramodal? If so, why?

- Is cognitive penetration always epistemically detrimental, or can it ever improve an individual's epistemic position?
- How do the epistemic impacts of cognitive penetration influence legal and social issues about reasonable action or belief on the basis of perception?

4 Ethics

If perception can be influenced by any of the top-down effects we listed at the start, can it also include representations of values? Can moral properties such as rightness and wrongness belong to the contents of perceptual experience? For example, if a person is watching a pickpocket grab a wallet from another person's pocket, she might see this action as wrong, in the same way that she sees the wallet as brown and square.

The idea of moral perception traces back as far as Aristotle, who at least on some interpretations held that virtuous people can perceive the good, through interactions of their perceptual and imaginative faculties (Moss 2012). Such Aristotelian ideas also feature prominently in the work of thinkers such as Iris Murdoch (1970) and John McDowell (1979), who discuss moral perceptual concepts such as the salience of relevant moral properties to an agent, and our ability to see another agent's character. Moral perception has been discussed in recent years from a variety of angles (Sturgeon 2002; Audi 2013).

As we attribute more complex types of contents to perception, the question of how our perceptual systems manage to represent such contents on the basis of their inputs and processing mechanisms becomes more pressing. Top-down effects on perception provide a natural explanation of how perceptual contents come to include such rich properties, including moral ones, as part of the contents of perception. If producing representations that have moral properties among their contents is beyond the reaches of the initial workings of our perceptual systems, we might come to perceive such moral properties through the intervention of beliefs, fears, desires, concepts, and other sorts of cognitive states. If cognitive penetration is not the correct explanation for how perceptual experiences get moral content, some other psychological story needs to be told. Experimental work in psychology would not determine how we ought to act in response to our experiences of moral properties, if there is any moral perception. But by helping us assess the plausibility of moral perception, psychology can constrain the available frameworks for our normative theorizing.

In "Cognitive penetrability and ethical perception", Robert Cowan reviews the different implications that cognitive penetration has for a variety of ethical theories that make use of the notion of moral perception. Cowan labels this family of views 'perceptualism', and he takes it to include virtue ethics, moral realism, and ethical intuitionism. These views all hold that at least sometimes, we can simply see the right or wrong in a given action, without having to perform any deliberate, conscious inferences in order to arrive at a moral assessment. Cowan argues that cognitive penetration would provide a good explanation for how moral perception might work—our beliefs may make us more or less sensitive to certain ethically relevant features in our perceptual experience, and could help to produce experiences with ethical contents. This would vindicate a literal sort of interpretation of moral perception, on which

moral properties are the outputs of the visual system, in just the same way that shape, color, and location are.

However, Cowan also discusses the possibility that the use of the term 'perception' in theories of moral perception is not quite so literal, and so need not be psychologically explained by cognitive penetration in order to play a legitimate theoretical role. On this more metaphorical sort of view, moral perception might simply involve influences on how we allocate our attention and on how we respond to our perceptions. Cowan notes that in order to capture the theoretically important aspects of ethical 'seeing', such as its immediacy and automaticity, it need not actually be perception in a technical, psychological sense.

For example, in his discussion of virtue ethics, Cowan considers whether ethical perception taken literally is really crucial to this view, or whether it would do just as well understood metaphorically. On a literal interpretation, the special ethical ability that a virtuous person has is the ability to perceive moral properties such as rightness and wrongness. The contents of her visual experience when she sees a pickpocket stealing a wallet, for example, would include the wrongness of the action, and perhaps also that it is to be stopped. If ethical perception involves cognitive penetration, then on this literal interpretation the special abilities of the virtuous person will be explained by the influences of her beliefs on her perceptions. On the alternative metaphorical interpretation, the virtuous person has the same sort of perceptual capacities that anyone else does, but she is able to infer from her perceptual experience to the right sorts of ethical judgments. The metaphorical model need not admit of any cognitive penetration, because all the differences between the virtuous person and the non-virtuous person are in their reasoning capacities, not in their perceptual systems. Cowan argues that a literal model of ethical perception in virtue ethics does not offer a clear advantage over a metaphorical model, so while cognitive penetration provides a nice explanation of ethical perception, it does not particularly tell for or against virtue ethics.

Cowan also considers how the epistemic relationship between cognitively penetrated perceptual experiences and the beliefs that cause them impacts different versions of perceptualism. Cognitively penetrated perceptual experiences are at least on some views (Siegel 2013) epistemically dependent on the cognitive states that cause them, in the sense that their justificatory role in part derives from the epistemic status of those beliefs. In this way they inherit their epistemic power from the states they were derived from, much in the same way that beliefs arrived at through inference do. Cowan argues that because ethical intuitionism holds that the justification that ethical experiences provide is theory-independent and non-inferential, a cognitive penetration model of ethical perception would not support intuitionism. Cornell realism, on the other hand, embraces the theory-dependence of ethical perception, so the epistemic dependence of cognitively penetrated ethical experiences would not be problematic.

Cowan's paper covers much of the terrain as to how cognitive penetrability may impact our existing ethical theories. Further questions in this area include:

• How might cognitive penetration modulate whether one perceiver sees another as a moral agent, or deserving of respect?
• How might it expand the potential contents of perception, especially to include ethical concepts? What are the psychological mechanisms by which this might work?

- What exactly are the features of perception that make it so appealing as a foundation for some ethical theories? Are these features apparent in any other sorts of cognitive processes?
- What kinds of cognitive penetrability would lend support to the idea central to virtue ethics that which features of a situation are salient to a subject has at least as much normative significance as how the subject acts in the situation?

In addition to the implications for traditional ethical theories, many other pressing moral and political issues arise from the possibility of top-down effects on perception. Racism has well-known perceptual extensions. Some forms of racism[9] are found in a wide range of subjects in the United States, including both White and Black Americans.[10] Jennifer Eberhardt (2004) found that White American subjects who are primed with images of Black faces are more likely to detect guns in blurred images than subjects who are primed with images of white faces, and that when they were primed with a crime-related objects, their attention is drawn to Black men more than to White men. Just as the attention of anxious subjects is drawn more to threatening than to non-threatening things, the attention of subjects primed with guns, police badges, or handcuffs, is drawn to Black male faces more than to White male faces.[11] Here, the psychological state triggered by either a Black prime or a crime-related prime may influence where we allocate our attention, how we perceptually process stimuli, the judgments we make in response to what we see, or some combination.[12] Perception becomes a handmaiden to behavior, decisions, and judgments that are congruent with racism. Whether these effects challenge modularism or not is interesting in its own right, but that issue does not need to be settled to see that writ large, perceptual behavior and judgment are deeply integrated into the social world. Studying the interactions between perceptual beliefs and society can help us illuminate the most intimate ways in which our social structures permeate everyday life.

References

Adams, R., and K. Kveraga. 2015. Social vision: Functional forecasting and the integration of compound social cues. *Review of Philosophy and Psychology*. In press.

---

[9] The chief form of racism that we are discussing here is implicit bias, which may be manifested through individuals' behavioral, perceptual, and reasoning mechanisms even if they do not explicitly endorse racist claims. (For more on implicit bias, see M. Brownstein's entry in the *Stanford Encyclopedia of Philosophy* on Implicit Bias). Some people might deny that implicit bias constitutes racism, but for our purposes the term is appropriate in that it highlights the connection between at least unconscious belief states and perceptual experience.

[10] Studies providing evidence that black Americans harbor anti-black biases include Ashburn-Nardo et al. (2003), Nosek et al. (2002) and Dasgupta (2004).

[11] Eberhardt's study was done only with (self-identified) White subjects. These results do not reflect controlling for the race of the participants.

[12] See for example Richeson et al. (2005) for research on effects on attentional distribution that correlate with implicit racial attitudes.

Ashburn-Nardo, L., M. Knowles, and M. Monteith. 2003. Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition* 21(1): 61–87.

Audi, R. 2013. *Moral perception*. Princeton: Princeton University Press.

Block, N. 2014. Seeing-as in the light of vision science. *Philosophy and Phenomenological Research* 89(1): 560–572.

Brownstein, Michael. "Implicit Bias", The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2015/entries/implicit-bias/>.

Burge, T. 2014. Reply to block: Adaptation and the upper border of perception. *Philosophy and Phenomenological Research* 89(1): 573–583.

Carey, S. 2010. *The origin of concepts*. Oxford: Oxford University Press.

Carroll, J.M., and J.A. Russell. 1996. Do facial expression signal specific emotions? Judging emotions from the face in context. *Journal of Personality and Social Psychology* 70(2): 205–218.

Carruthers, P. 1996. *Language, thought, and consciousness*. Cambridge: Cambridge University Press.

Carruthers, P. 2006. *The architecture of mind: Massive modularity and the flexibility of thought*. Oxford: Clarendon.

Churchland, P. 1988. Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science* 55(2): 167–187.

Churchland, P. 2006. Vector completion, relevant abduction, and the capacity for 'globally sensitive' inference. In *Cognitive penetrability and the problem of induction*, ed. A. Raftopoulos. New York: Nova Science Publishers, Inc.

Clark, A. 2013. Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(03): 181–204.

Dasgupta, N. 2004. Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research* 17(2): 143–168.

Deroy, O. 2013. Object-sensitivity vs. cognitive penetrability of perception. *Philosophical Studies* 162: 87–107.

Eberhardt, J. 2004. Seeing Black: Race, crime and visual processing. *Journal of Personality and Social Psychology* 87(6): 876–893.

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

Firestone, C. & Scholl, B. (2015). Can you *experience* 'top-down' effects on perception?: The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*.

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

Hohwy, J. 2013. *The predictive mind*. Oxford: Oxford University Press.

Levin, D.T., and M.R. Banaji. 2006. Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General* 4: 501–512.

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.

McDowell, J. 1979. Virtue and reason. *The Monist* 62(3): 331–350.

Moss, J. 2012. *Aristotle on the apparent good: Perception, phantasia, thought, and desire*. Oxford: Oxford University Press.

Murdoch, I. 1970. *The sovereignty of good*. London: Routledge.

Nosek, B., M. Banaki, and A. Greenwald. 2002. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice* 6(1): 101–115.

Pinker, S. 1997. *How the mind works*. New York: W.W. Norton & Company.

Pollock, J., and J. Cruz. 1999. *Contemporary theories of knowledge*, 2nd ed. Lanham: Rowman & Littlefield.

Prinz, J. 2006. Is the mind really modular? In *Contemporary debates in cognitive science*, vol. 14, ed. R.J. Stainton, 22–36. Malden: Blackwell Publishing.

Pylyshyn, Z. 1999. Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22: 341–423.

Quine, W.V.O. 1951. Two dogmas of empiricism. *Philosophical Review* 60: 20–43.

Richeson, J., S. Trawalter, and J.N. Shelton. 2005. African Americans' implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition* 23(4): 336–352.

Siegel, S. 2010. *The contents of visual experience*. Oxford: Oxford University Press.

Siegel, S. 2013. The epistemic impact of the etiology of experience. *Philosophical Studies* 162(3): 697–722.

Smithies, D., and D. Stoljar. 2012. *Introspection and consciousness*. Oxford: Oxford University Press.

Sturgeon, N. 2002. Ethical intuitionism and ethical naturalism. In *Ethical intuitionism: Re-evaluations*, ed. P. Stratton-Lake. Oxford: Oxford University Press.

Vance, J. 2014. Emotion and the new epistemic challenge from cognitive penetrability. *Philosophical Studies* 169(2): 257–283.

Witzel, C., H. Valkova, T. Hansen, and K.R. Gegenfurtner. 2011. Object knowledge modulates colour appearance. *iPerception* 2: 13–49.

Zeimbekis, J., and A. Raftopoulos (eds.). 2015. *Cognitive penetrability.* Oxford: Oxford University Press.