Cognitive Penetration and the Epistemology of Perception

For Blackwell Compass
Nicholas Silins
Cornell University / Yale-NUS College

Introduction

Detective Lennie is interviewing Lex about who committed the crime. Detective Lennie believes that someone in a red sweater did it, and little does he know, Detective Lennie’s own belief causes him to ask highly leading questions. Lex ends up asserting that someone in a red sweater did it because Detective Lennie believes so. Does Lex’s testimony caused by Detective Lennie’s belief now support Detective Lennie’s belief? You might think not, given the way the testimony came about.

According to one foundationalist tradition, your visual experience functions as a mirror, and is nothing like Lex’s testimony. Your experience reflects what is before you and does not reflect your own mind. Given that your experience is not influenced in any way by your theories or expectations, it is thereby in an optimal position to confirm or disconfirm hypotheses about the world.

According to many philosophers and scientists, that particular foundationalist picture of experience is wrong. Just as Detective Lennie unwittingly causes Lex to testify a certain way, your own mind sometimes causes you to experience the world to be the way you antecedently believed or expected it to be. For example, a rose might look red to you because you expected it to look red, and indeed would have looked red to you whether or not it really were so. If your experience is influenced in this way, it’s not so clear whether your experience is in a good position to support your belief that the rose is red.

The debate here is broadly about how your prior non-perceptual states of mind might influence your experience, and about the ramifications in epistemology of any such influence. Given that the stimuli we receive are compatible with ever so many distal causes, our experience cannot be entirely determined by the stimuli we receive---the perceptual system needs some way to narrow the field of candidates for being what’s there. Our focus is on how your own beliefs, expectations, desires, hopes and so on might intervene in the perceptual process so that it comes up with verdicts about what is there. Since the potential interventions we are interested in would be by cognitive states from outside the perceptual system, where those cognitive states are not entirely determined by the stimuli received by the subject, we may speak of “cognitive penetration” or of “top-down effects”. While one might use the terms to pick out related but different phenomena, we will use those terms interchangeably. We will also leave aside how elements internal rather than external to the perceptual system might narrow down the field of candidates for being in the scene. (For some discussion of how to delineate what’s on the side of perception and what’s on the side of cognition, see Stokes 2013).
To survey the debate, we will first clarify what top-down effects would be, and briefly examine whether they happen. We will make our way to recent literature starting from classic philosophy of science discussions, where the debate is framed in terms of the “theory-ladenness of observation”. Here we will see that some common ways of denying that there are top-down effects have epistemic consequences of their own. We will then turn to examine what epistemological consequences top-down effects might themselves have. Here we will see that top-down effects might have positive as well as negative consequences for rational belief and knowledge.

Our focus will be on the case of vision, although we will discuss one case of flavor. Similar questions do arise for other senses, as well as for the case of memory (Michaelian 2013).

1. Definitions

We speak of cognitive penetration or top-down effects, but top-down effects on what? At least two types of relevant states might be said to depend on your theory or cognitive states: perceptual experiences, or beliefs formed on the basis of perceptual experiences (perceptual beliefs). In the classic Müller-Lyer illusion, two lines perceptually look to be of different lengths, even though their length is the same.


Learning that the lines have the same length does not change the way the lines look. On the standard picture here, the informed perceiver experiences the lines as differing in length, and yet does not believe on the basis of experience that the lines differ in length. On the standard picture then, seeing is not believing.¹ Contrast the following example:

---

¹ For some challenges to the standard view of how seeing can come apart from believing, see Pitcher (1971), Glüer (2009), or Byrne (forthcoming).
Fig 2: no illusion

Here you experience the lines as different in length, and also believe on the basis of your experience that they are different in length.

On the picture used in this entry, your perceptual experiences of figures 1 and 2 have “content” in the sense that they are accurate only if the lines are different in length. Your experience of figure 1 is accurate with respect to length; your experience of figure 2 is not. In the case of both figures, your experience has the same content as your belief that the lines are different in length, or at any rate a content closely related to that one. Still, your perceptual experience differs from perceptual belief in terms of “phenomenal character” or in terms of what it is like for you to have the experience. When you close your eyes and lose your experience, but retain your perceptual belief, there is a dramatic change in your conscious perspective. When we speak of the “character” of your experience in what follows, we are trying to home in on your visual conscious perspective, that aspect of your visual conscious perspective that is absent when you close your eyes.

Having distinguished perceptual experience and perceptual belief, we can now look at how the term “observation” is used in different ways. When “observation” denotes perceptual belief, then observation is theory-dependent when the usual process of forming beliefs on the basis of experience is disrupted or otherwise influenced by a prior theory. For instance, in the Müller-Lyer case, it is disrupted by the perceiver’s prior theory that the lines are the same length, while the perceptual state itself remains untouched. There is little debate that “observation” in this sense depends on theory.

Let’s now consider cases where “observation” denotes perceptual experience. Here perceptual experience would be theory-dependent roughly when a prior theory influences what the perceptual experience you have is of. So if things look a certain way to you because you have a certain theory, then your experience is theory-dependent. For example, suppose that the Müller-Lyer lines changed their appearance to you when you learned they are the same length, and came to look to have the same length because you believed they have the same length. That change would result in a case of theory-dependence of “observation” were it to occur.

There is much more debate about whether “observation” in the sense of “experience” ever really does depend on theory. Since the New Look movement of the mid-20th century and the publication in 1983 of Jerry Fodor’s influential book *The Modularity of Mind*, psychologists have debated the extent to which perceptual processes are insulated from the rest of the cognitive system, in the way that the perception of the Müller-Lyer lines seems insulated from the perceiver’s knowledge that the lines are the same length. Perceptual processes are said to be *modular* when they are free from any such influence.²

---

² For further discussion of modularity, see Stokes (2013: 653-654).
What exactly would it be for experience to be subject to theory-dependence, or to be cognitively penetrable? As a first pass, one might say that your experience is cognitively penetrable just in case it can be altered by your beliefs, desires, or other cognitive states. But suppose you hear something go bump in the night and move your eyes towards the door. Your experience will then change, but there’s nothing controversial about the view that your desire to see what caused the sound can alter your experience. Relatedly, there’s nothing controversial about the view that your experience can be altered by your desire to secretly shift your visual attention to someone else at the cafe, without even moving your eyes towards them (Fodor 1984, Macpherson 2013, although see Mole 2014 for critical discussion). Theorists have had something else in mind when they spoke of theory-dependence or cognitive penetration.

To make progress here, it should help to hold certain factors fixed, and to see whether people can still differ with respect to their experience due to cognitive factors. In particular, one might say that experience is cognitively penetrable just in case the following scenario is possible:

two people are the same with respect to their sensory inputs, the state of their sensory organ, and the orientation of their spatial attention, and they are still different with respect to what their experience is like, because of their beliefs, desires, or other cognitive states (see Macpherson 2012).

There remains a difficulty, given the distinction between attending to a location and attending to a feature.

Suppose that two people looking at qualitatively identical leaves attend to different features due to their different interests---the one in the color of the leaf and the other in the form of the leaf. They arguably can do so even if they are the same with respect to spatial attention. Here the character of their experience might be different, even though they are the same with respect to proximal stimuli, the condition of their sensory organ, and spatial attention. The case still isn’t the sort of case most are trying to capture with the term “cognitive penetration” or “top-down effect”. It certainly doesn’t seem to challenge claims about perception being modular, or to raise any questions about the ability of the respective experiences to justify beliefs.  

In response to the problem, one might define cognitive penetrability by holding all attentional phenomena fixed and not just spatial attention fixed. The new approach would say that experience is cognitively penetrable just in case the following scenario is possible:

---

4 There’s also a general difficulty for the approach of holding certain factors fixed, and checking how experiences could still vary because of cognitive factors. Suppose that there’s some content or character that all experiences have, say the self-referential content that this is an experience. Perhaps such content could be the result of cognitive penetration. But since all experiences have it, there’s no way for our current approach to count it as the result of cognitive penetration. I learned of this sort of point from Uriah Kriegel.
two people are the same with respect to their sensory inputs, the state of their sensory
organ, and the orientation of their attention, and they are still different with respect to
what their experience is like, because of their beliefs, desires, or other cognitive states.

The new proposal threatens to embroil us in debates about the relation between
experience and attention. According to some controversial views of the relation between
experience and attention, if two subjects are the same with respect to what they attend to
in the visual scene, then they will be the same with respect to their experience as well
(e.g. Prinz 2012). On the proposed gloss of “cognitive penetrability”, it would then turn
out that no experiences are cognitively penetrable regardless of how much they are
influenced by our cognitive states. But a useful definition of “cognitive penetrability”
arguably should allow us to debate cognitive penetrability while being neutral with
respect to controversy about the relation between experience and attention (for more on
cognitive penetrability and attention, see Mole 2014).5

In sum, it’s not so easy to define “cognitive penetration” in a theoretically
satisfying way (for further discussion of how do so, see Machery 2014, Stokes 2014, or
Shea 2014). Insofar as possible, a useful definition shouldn’t presuppose answers to
controversial questions about the nature of experience (e.g. its relation to attention). And
cognitive penetration will involve the influence of perceptual experience by cognitive
states, in a way that is not trivial, but how exactly to pin down “cognitive penetration”
more sharply remains open (as well as whether there will be a unique theoretically useful
definition of “cognitive penetration”). In what follows, we will proceed primarily by
orienting our discussion around paradigmatic examples.

2. Potential Examples of Cognitive Penetration and the Epistemic Consequences of
Denying Cognitive Penetration

Let’s now survey some potential cases of top-down effects, and begin to survey
potential epistemic implications of the phenomenon, starting with cases from the older
philosophy of science literature. Our movement here will be from less convincing cases
to more convincing cases. The detour is worth taking since it reveals that denying
cognitive penetration often has epistemic consequences of its own.

Theorists such as Hanson, Kuhn and Churchland made much of work by so-called
“New Look” psychologists such as Bruner and Postman. Consider for instance Bruner
and Postman’s classic 1949 experiment involving anomalous playing cards, in which they
briefly showed their subjects the following sort of card:

5 A further related worry about attention. Suppose I experience a feature you don’t
because my experience has been cognitively penetrated. Either I attend to this feature or
I don’t. If I do attend to the relevant feature, then we will differ with respect to what we
attend to, and our attempted gloss of “cognitive penetrability” won’t apply. If I don’t
attend to the relevant feature, then I experience it only if those theories are wrong that
demand that you attend to an entity in order to experience it. So applying the proposed
definition would require taking a stand on the controversial question of whether you
experience something only if you attend to it.
Even though the card is a red six of spades, many subjects reported the card to be a six of hearts.

According to Bruner and Postman, such cases are ones in which “a red card [is] seen as a heart or diamond regardless of its true suit (1949: 215),” with “the dominance of one principle of organization which prevents the appearance of incongruity (1949: 222).” They seem to interpret the impact of our expectations as being one on experience. On this interpretation, the way the card looked to the subject is the same as the way a genuine red six of hearts looks to the subject in good viewing conditions. However, further evidence is needed to conclude that the effect is on perceptual experience rather than merely on perceptual belief (Fodor 1983, Pylyshyn 1999). Perhaps the card looked like a red six of spades, or otherwise failed to look like a red six of spades, and the subject merely formed the belief that it is a red six of hearts.

Hanson is aware of this type of challenge. He writes in response:

To say that Tycho and Kepler, Simplicius and Galileo, Hooke and Newton, Priestley and Lavoisier, Soddy and Einstein, De Broglie and Born, Heisenberg and Bohm all make the same observations but use them differently is too easy. It does not explain controversy in research science. Were there no sense in which they were different observations they could not be used differently (1965, ch. 1).

The force of Hanson’s response is unclear. The disagreeing scientists trivially have different “observations” in the sense of having different perceptual beliefs, but there’s no clear reason why their perceptual experiences must be different for them to have different perceptual beliefs. Consider how you might no longer take an experience at face value when you are informed that you are in a case of illusion, as would happen to you when first informed of the Müller-Lyer illusion. Here there is a change in how you use your “observation” without a change in your experience.

Let’s now look at more recent experiments in psychology used in the current debate about cognitive penetration uses.

For a contemporary case in which the belief interpretation seems to be positively supported over the experience interpretation, consider work by Payne 2001 on the impact
of racial stereotypes. White American participants were primed either with faces of Black men or faces of White men, and then had to indicate by keystroke, under time pressure, whether they were shown an image of a tool or a gun. Participants given the Black prime miscategorized tools as guns more often than participants given the White prime.

![Primes](image1)

![Targets](image2)

Fig 4: Primes, Tool, and Gun
Source: personal communication from Keith Payne

In a follow-up study with similar results (Stokes and Payne 2011), participants were required to say after each trial how they reached their verdict, by choosing between three options: SEE (if they saw all or part of the object), GUESS (if they felt they were just guessing), or KNOW (if they didn’t see the object very well but felt they ‘just knew’ whether they were shown a tool or a gun, even without having a clear visual image or memory of it). When participants said they saw the stimulus (the SEE option), they hardly ever made mistakes, whereas when subjects with the Black prime miscategorized tools for guns, they almost always said they ‘just knew’ without seeing. Call these subjects GUN KNOWERS. Given that GUN KNOWERS described themselves in terms of knowing what was present, and did not take the option of saying they guessed, we have evidence that they did believe a gun was present. Given that GUN KNOWERS denied having a clear image of a gun, we also have evidence that they didn’t have a visual experience of a gun. Instead, they either had a visual experience of a tool, or a degraded visual experience that was undecided between whether there is a gun or tool present. Either way, GUN KNOWERS seem to have lacked an experience of a gun and just jumped to the conclusion that a gun was present.

Even if the Payne cases are not ones of cognitive penetration of experience, they might still be of great epistemic importance (not to mention the life-or-death practical importance of misclassifying a tool as a gun). To see why, bear in mind that suspension

---

6 You might agree that the subjects who chose KNOW did not merely guess about what was present, but you might still think that, in this forced choice condition, they still did not believe that a gun was present. Instead, perhaps they merely had a higher than 50% level of confidence that a gun was present. Either way, the points in what follows go through if understood in terms of higher than 50% confidence rather than belief.

7 For potentially conflicting results, see study 3 in Correll et al 2015, but also their discussion on 231-2.
of judgment is sometimes the justified attitude to take up in response to weak evidence. For example, an experience whose content is too impoverished to justify either a gun-belief or a tool-belief arguably does have justificatory force, in that it supports considered suspension of judgment. Alternatively, if a GUN KNOWER’s experience does delineate a tool clearly, while still being side-stepped in the belief-forming process, their experience arguably justifies the negation of the proposition they come to believe. Either way, they end up forming a belief that isn’t justified because of the mismatch between their perceptual belief and their perceptual experience. (See Siegel 2013 for further defense of this sort of claim).

In the Payne case, GUN KNOWERS don’t use an epistemic resource they have. They arguably also are unable to use the epistemic resource they have, given that they are making a good faith effort to form opinions about the surroundings in light of their experiences, and yet still fail to do so. The cases thus have implications for wider debates in epistemology.

Consider the principle that, if you have justification to take up an attitude on the basis of x, then you are able to take up the attitude on the basis of x (Turri 2010). Such a principle is favored by those who would like to analyze “propositional justification”---roughly a matter of having reasons---in terms of “doxastic justification”---roughly a matter of using reasons. To get the distinction, compare how Holmes and Watson might have the same reasons to believe that the butler did it, where only Holmes has properly formed his belief in response to those reasons. Watson on the other hand might not have followed through on those reasons to form a belief that the butler did it.

On the approach to understanding propositional justification favored by Turri 2010, if your experience gives you reason to take up a certain attitude to the proposition that p, it follows that you are able to take up that attitude to the proposition that p on the basis of your experience. However, the Payne subjects either have reason to suspend judgment about what they just saw, or have reason to disbelieve, apparently without being able to form those attitudes on the basis of their experiences. They arguably illustrate that you can have reasons without being able to use them.

Suppose reasons and justification are understood in terms of epistemic obligations. On this picture, if your experience gives you reason to take up an attitude A, then you ought to take up attitude A. Now consider the question of whether “ought implies can”. Here the Payne cases might also illustrate that an epistemic “ought” does not imply “can”. The Payne subject arguably ought to suspend judgment about whether a gun is present, or ought to believe that a gun is not present, and yet still cannot do otherwise.

One might respond to the cases by adding qualifications to the principles they challenge. One option would be to add a qualification about time. Here the claim would be that, if you ought to take up attitude A in response to your experience, and you have

---

8 The Bruner/Postman case might make the same point. Suppose that when you are showing the red six of spades, you have the sort of experience you normally have when you see a red six of hearts. Your experience might then give you reason to believe that a red six of hearts is present, even though you are unable to take advantage of the experience you have.
proper time to do so, then you can do so. However, one might worry that the qualification is ad hoc.

Another move would be towards idealization. Here the claim would be that, if your experience gives you reason to take up an attitude towards the proposition that p, then a suitably idealized counterpart of you does take up the attitude towards the proposition that p on the basis of the experience. For further discussion of this sort of move, see Turri 2010, Smithies 2012, or Siegel and Silins forthcoming.

Rather than qualify principles about how “ought” implies “can”, one might instead argue that GUN KNOWERS do have the ability to take up the right attitude in response to their experience, but simply fail to exercise their ability. We might have good evidence that they don’t take up the right attitude, but perhaps we need further evidence that they can’t take up the right attitude.

Having looked at unsuccessful cases of cognitive penetration, and having seen that they still have potential implications for epistemology, let us now turn to more promising candidates to be cases of cognitive penetration.

Consider the banana (that of Hansen et al 2006). When subjects were asked to adjust an image of a characteristically yellow object such a banana until it was achromatic, they overcompensated by adjusting until the image was slightly bluish. Subjects did not overcompensate with objects that were not characteristically yellow. The pattern suggests that the image of the characteristically yellow object looked slightly yellow when the image was in fact achromatic, and that the image looked achromatic when it was in fact slightly blue. According to recent discussions by Macpherson 2012 and Stokes 2013, this sort of case is harder to explain away than those in the classic philosophy of science literature. First, the case does not seem to be one of a mere attentional shift. Second, the case doesn’t seem to be one of a mere effect on perceptual belief. Third, cognitive states of the subject are plausible candidates to explain why the achromatic banana still looked slightly yellow to the viewer.

For examples that are similar to the banana case, see Delk and Fillenbaum 1965, Olkonnen et al 2008 or Witzel et al 2011. To get a feel for how wide the range of cases might be, consider the medley on display in the following figure from Witzel et al 2011: 33:
Witzel et al’s experiments were on German subjects. So Germans’ experiences might be cognitively penetrated whenever they see fire extinguishers, mailboxes, ping-pong tables, Cokes, road signs, or Smurfs!

Following Hansen et al, let’s label the current cases as putative ones of “memory color”. One way to try to explain away memory color is by trying to show that the effect here is after all on perceptual belief rather than perceptual experience---call this the “belief response”. The belief response is of special interest for its ramifications in epistemology. Here the implications concern our access to our own minds rather than the external world.\(^9\)

First, consider whether we are infallible about what our experiences are like. If the belief response to the memory color cases is correct, we presumably make mistakes about what our experiences are like. If we consider a conceptually sophisticated subject in the experiment, when she falsely believes that she has adjusted the banana image to be achromatic (it is actually slightly blue), she will end up believing that she is experiencing the image to be achromatic. If the belief response is right, she will instead be experiencing it as slightly blue. The belief response predicts that we make mistakes about what our experiences are like, across a wide range of putative cases of memory color.

According to Macpherson 2012, the consequence of the belief response is implausible:

> explaining away the alleged cognitive penetration of experience by means of this strategy involves, in this case, the postulation of a gross, brute and inexplicable error on the part of the subject – an error on the personal level that we think

\(^9\) A different way to explain away the memory color cases is to grant that the effect is on perceptual experience, but to insist that the cause is states in the perceptual system rather than cognitive states of the subject. For more on this response see Deroy 2013.
subjects are responsible for. No doubt many people think errors of this kind are possible, but this view predicts that these errors occur in a systematic way – for example they will always occur in the conditions that the experiment specifies. And that thought is unpalatable, for what then ensures any of our judgments reflect our experience? (Macpherson 2012: 42)

In response, one might say that the errors predicted by the belief response can’t be inexplicable if they are systematic. For all the belief response says, we may make mistakes about our experiences only in the conditions specified by the relevant experiments, and indeed because of the conditions specified by the relevant experiments. Alternatively, one might simply accept that we can make systematic introspective mistakes about our own experience, as do Dennett (1988) and Schwitzgebel (2011).

The worry about introspective mistakes might also go too far. Consider the Bruner/Postman card case, or the Stokes/Payne gun case. If the belief response applies there, an introspective subject in the experiment will end up falsely believing that her experience was of a red six of hearts, or of a gun. Indeed, introspective subjects will end up being systematically mistaken about what their experiences were like. So if you forbid the belief response and the ensuing introspective error in the Hansen banana case, then you might also have to forbid the belief response and the ensuing introspective error in the Bruner/Postman card case and the Stokes/Payne gun case. But the belief response might have seemed quite good with some of those cases, as argued by Payne 2011.

However, perhaps there is a way to treat the cases differently. For example, perhaps we should expect introspective mistakes when judgments are made about fleeting experiences, but not when they are made under the better time conditions of the Hansen case.

The belief response to memory color might have a further negative consequence about our access to our experiences. Someone might grant that we sometime have false beliefs about what our experiences are like when our introspective performance misfires, but insist that we are always in a position to know what our experiences are like if we properly reflect on the matter. If the belief response is true, the introspective subject in the Hansen experiment experiences the banana image to be slightly blue, and yet seems unable to know that she does, given that she is doing her best to reflect on what her experience is like and is stuck with the belief that she experiences the banana image to be achromatic. The belief response thus predicts not just false belief about what your experience is like but also an inability to know about what your experience is like. Even if you are willing to grant that we can make many mistakes about what our experiences are like, you might still think that we are always in a position to know what our experiences are like even when our performance misfires. So there’s a further reason to suspect that the belief response to memory color has consequences that are false.

One might again respond that the belief response predicts no more introspective skepticism than we should anyways expect. For arguments that we aren’t always in a position to know what our experiences are like, see Williamson (2000) or Greenough (2013).

3. The Epistemic Consequences of Affirming Cognitive Penetration
So far we have focused on the following questions:

- Is there cognitive penetration?
- What if there isn’t?

Let’s now ask:

- What if there is cognitive penetration?

We will assume that there are some genuine cases of cognitive penetration, and consider what epistemic consequences cognitive penetration might have. Here we will start by considering positive epistemic consequences, and then will turn to negative epistemic consequences, considering both rational belief and knowledge. Assuming that there are some negative cases of cognitive penetration, we will then consider what theories in epistemology are affected by them. (While our discussion is framed in terms of actual cases, it might well be that non-actual but possible cases would do well enough to challenge various claims in epistemology, as Siegel 2012 argues).

3.1. Are all top-down effects epistemically negative?

According to an extreme view, no experience that is cognitively penetrated justifies any (external world) belief. This view is almost certainly too demanding. Suppose a debit card looks closer because you desire it, as argued by Balcetis and Dunning 2010 (see Stokes 2012 for more on cognitive penetration by desire). Perhaps your experience is no longer capable of justifying your belief that the card is at such and such distance from you. But your experience is presumably still capable of giving you reason to believe that there is a debit card in the scene. Top-down effects are not epistemically lethal.

A more moderate proposal is that, if your experience has the content that p due to cognitive penetration, then your experience does not give you reason to believe that p. So if your experience is of the debit card as being at such and such location, and your experience is cognitively penetrated with respect to that locational content, it does not give you reason to believe the specific locational content, but may well still give you reason to believe that a debit card is in the scene.

The more moderate proposal might still be too demanding. Consider the possibility of cognitive penetration by expertise, where expertise might actually improve the status of our experiences as guides to the world (Churchland 1988).

To get a particular example in focus, consider what you experience when you look at the photograph below:
The shrewd kids in the background on the right arguably have experiences that are not just of colors and shapes, but in particular of Josh Kalis catching a 360 flip (the blank-looking kid in the middle? Well, maybe he’s just getting colors and shapes). The expert’s experience here would make a commitment about who is present in the scene, and about what trick he is doing, and about what stage he is in with the trick. Part of this is a matter of attending to Kalis’ skateboard that isn’t yet straight and his feet that aren’t yet fully on the bolts, but the non-expert might attend to those objects without yet having the expert’s experience. The case is arguably one of top-down influence on experience, by background knowledge that the expert viewer has.

Here the experts are presumably at an epistemic advantage thanks to top-down effects, where their experience justifies them with respect to those contents that their experience has thanks to their cognitive background. Cognitive penetration need not be pernicious even with respect to those contents that it generates.

For another potential case of benign cognitive penetration, we can go beyond the case of vision, and consider the possibility that your expectations affect your taste experiences. Consider the experiment of Lee et al 2006 involving beer flavored with balsamic vinegar and normal beer. If subjects are informed before tasting they are about to have some beer with balsamic vinegar in it, as well as some normal beer, they generally say they prefer the beer without vinegar upon tasting. However, if they first taste the respective beers, and are only informed afterwards that one contained vinegar,
more subjects say they prefer the beer with vinegar. This pattern suggests that the information that a beer contains vinegar changes your taste experience rather than merely changing your evaluation in judgment of the taste experience. If the information merely changed your evaluation in judgment of the taste experience, presumably it shouldn’t matter whether you get the information before or after tasting. Even if you get the information after tasting, you presumably would still downgrade your evaluation in judgment of the taste. While other hypotheses are possible, let’s assume that there is a change in taste experience due to the information that the beer contains vinegar, and let’s assume in particular that the change is a proper case of cognitive penetration.

It’s not clear whether any such dependence cripples your experience epistemically. If you judge that a beer containing vinegar tastes worse than one without it, and the beer indeed does taste worse because of the information you possess, you still arguably are epistemically justified in believing that the beer containing vinegar tastes worse than the one without it.

In short, some top-down effects might be epistemically beneficial or neutral rather than harmful (see Vance forthcoming for more on how some top-down effects might be epistemically beneficial).

3.2. Are some top-down effects epistemically negative?

3.2.1. The case of justification

To see why some cases of top-down effects might still be epistemically problematic, let’s now turn to cases of perception and anti-expertise. We saw how expertise might be epistemically beneficial thanks to putting more information into your experience. We can think of anti-expertise as the subtraction of information from experience due to a top-down effect. Consider for instance the notorious difficulty of telling between different members of the same racial outgroup. According to the “asymmetric feature selection” hypothesis reviewed in Gendler 2011 section 2, when White subjects encounter White subjects, they do not visually encode information about race (see also Levin 1996 or Levin 2000). The comedian Stephen Colbert would then be at least partly right when he says, “I don’t see race…people tell me I’m white, and I believe them” (The Colbert Report, Episode 2138, November 2, 2006). On the other hand, when White subjects see subjects who aren’t White, they do visually encode information about race. Since cognitive resources are scarce, this classification comes at the expense of more specific information about the faces of racial outgroup members, making it harder to tell between racial outgroup members. Now suppose such cases are ones of cognitive penetration. Here cognitive penetration might put you at an epistemic disadvantage by taking information away.

It’s uncontroversial that your experience will be a weaker epistemic resource when it carries less information. Are there cases where your experience does present the world to you as being a certain way due to a top-down effect, and still fails to give you reason to believe that the world is that way? Here it will help to move to the hypothetical case of "Angry Looking Jack" (Siegel 2012). Here we will finally consider a case along the lines of Detective Lennie from the introduction, where Lennie’s own belief that p caused Lex to testify that p even though Lennie had no idea that process took place.
Suppose that Jill antecedently has an unjustified belief that Jack is angry at her, where, as a result of a top-down effect from her belief, Jack does look angry at her when she next sees him. In response to her experience, Jill reaffirms her belief. Is she now justified by her experience in believing that Jack is angry at her? (Markie 2005 presents a related example of a gold prospector whose wishful thinking makes a nugget visually seem to be gold, and asks whether the prospector’s experience justifies him in believing that the nugget is gold).

You might appeal to an intuition that the answer to the question is “no” (Siegel 2012). But others may or may not share the intuition. For example, Pryor 2000: 540-1 suggests that the processes leading to your experience seem irrelevant to whether it justifies subsequent beliefs. So he at least would seem to lack an intuition that Jill is not justified by her experience.

Rather than appealing to intuitions about the example, you might bring out a further argument. For arguments that appeal to an analogy with the unjustified formation of a belief on an unjustified belief, see McGrath 2013 and Siegel 2013. The broad strategy of the approach is to rely on the principle, “garbage-in, garbage-out.” When a chain of inference starts from an unjustified belief, it tends to lead only to other unjustified beliefs. When an unjustified belief has a top-down effect on your experience, and you form a belief on the basis of your experience, your formation of the downstream belief might be too analogous to such a chain of inference—your experience is arguably unable to somehow compost the garbage of the initial unjustified belief.

Unjustified belief $\Rightarrow$ unjustified belief $\Rightarrow$ unjustified belief
Unjustified belief $\Rightarrow$ experience $\Rightarrow$ unjustified belief?

As an objection, one might insist on disanalogies between beliefs and experiences. When you believe something, there needn’t be anything it’s like for you have the belief. For example, one minute ago you believed that your name is ___, but there needn’t have been anything it was like for you to believe that your name is ___. On the other hand, when you have an experience, there is something it is like for you to have the experience—consider how things vividly change when you open and close your eyes or cover and uncover your ears. Given the dramatic difference between beliefs and experiences, perhaps an experience is able to lead to a justified belief when the experience results from an unjustified belief, even if a belief resulting from an unjustified belief tends to lead only to another unjustified belief.

A related argument by Vance (2014) aims to avoid the challenge, by using an analogy with the case of emotion rather than belief. First, Vance asks us to consider someone with an unjustified belief that foreigners are dangerous, who feels fear when he meets a foreigner as a result of his unjustified belief. Here Vance claims that the subject’s emotion is in no position to give any confirmation to the hypothesis that foreigners are dangerous, because of its etiology, especially given that the fear is itself assessable as irrational. Vance goes on to carry the verdict over to perception by highlighting analogies in how emotions and perceptual states justify, and especially with respect to what it is like to be in them.
As an objection, one might now insist on disanalogies between emotions and experiences. Perhaps a relevant difference between them is that emotions are rationally assessable although experiences are not. Here one might respond that the similar epistemic roles of experience and emotion outweigh the differences between them. More radically, one might respond by arguing that experiences are rationally assessable after all (Siegel ms).

3.2.2. The case of knowledge

Let’s now consider how top-down effects might result in the absence of knowledge, whether or not they lead to an absence of justified belief. Assuming that knowledge requires true belief, top-down effects will inevitably rule out knowledge when they lead to false beliefs. But in cases where top-down effects lead to true beliefs rather than false beliefs, other requirements for knowledge still might not be met. For example, Nozick (1983) and others think you must have a sensitive belief that \( p \) in order to know that \( p \). As a first pass, the requirement is the following one:

If you know that \( p \), then if \( p \) were false, you wouldn’t believe that \( p \).

To get a feel for the requirement, suppose you hold a losing ticket and you believe on statistical grounds that you hold a losing ticket. Even if your ticket were to be a winner, you would still believe that it is a loser. Arguably you don’t know that your ticket is a loser when it is a loser, and arguably your insensitivity to the truth explains why. Now let’s go back to the case of top-down effects. Suppose that Jack is angry and looks angry to Jill because of her unjustified belief, and still would have looked angry to her even if he hadn’t been angry. If he hadn’t been angry, she still would have reaffirmed her belief on the basis of her experience. So her belief that he is angry fails the sensitivity requirement for knowledge even though it is true. Arguably many cases of top-down effects will result in beliefs that are insensitive to the truth even if they happen to be true. Perhaps then all those beliefs will fail to be cases of knowledge.

A complication here is that sensitivity requirements for knowledge are controversial. Consider the following example adapted from Goldman (1979). You see a shaggy Great Pyrenean Mountain Dog guarding the sheep in the field, and you form a belief that there is a dog in front of you. Here you plausibly know that there is a dog in front of you since the Pyrenees is clearly a dog. However, if the Pyrenees hadn’t been there, a wolf that looks just like a husky would have been there, and would have caused you to falsely believe that there is a dog in front of you. If knowledge requires sensitive belief, then you don’t even know there’s a dog in front of you when you’re seeing the Pyrenees. (For discussion of attempts to refine the sensitivity approach, see Williamson 2000, Roush 2007, or DeRose 2010).

An alternative, less controversial requirement for knowledge is in terms of safety (Sosa 1999, Williamson 2000). Here the rough idea is that, if you know that \( p \), then you couldn’t easily have been mistaken about whether it is the case that \( p \). One might support the safety approach by saying that it explains the sorts of cases sensitivity was meant to explain, while avoiding the pitfalls of sensitivity. Your belief that your lottery ticket will lose isn’t safe from error, since the ticket could easily have been a winner while you still
believed it’s a loser. However, when you see the Pyrenees dog, your belief that there is a
dog in front of you is safe from error (assuming that the misleading wolf couldn’t easily
have been in the field).

Safety approaches provide more promising candidates to be requirements for
knowledge, and also seem to predict true beliefs that are failures of knowledge in some
cases of top-down effects. Consider the Hansen subject when she aims to adjust the
image of the banana until it is achromatic. When the image is nearly achromatic, she is
in danger of being mistaken about whether it is yellowish, since it will soon be
achromatic, but will still look yellowish to her where she remains at least fairly confident
that it is yellowish. Assuming that a safety requirement holds for knowledge, and
assuming the subject believes that the image is yellowish when it is nearly achromatic,
she will have a true belief that the image is yellowish that fails to be a case of knowledge.
Assuming that a safety requirement does not hold for belief, she might even have a
justified true belief that fails to be a case of knowledge.\footnote{For further discussion of similar examples, see Williamson 2000 or Berker 2008.}

Top-down effects can threaten knowledge even when they don’t threaten true
belief. It remains an open question how many cases of top-down effects will threaten
knowledge.

3.3. What theories are affected by epistemically negative cases of top-down effects?

We’ve now seen how some cases of top-down effects might be epistemically
problematic. A central question in the literature focuses on those cases that are
epistemically problematic, and tries to identify the specific theories in epistemology for
which the cases are a threat. If top-down effects are sometimes an epistemic problem,
who exactly are they a problem for?\footnote{Earlier we saw how some cases of top-down effects might be epistemically
unproblematic. A further question in the literature is about how to sift cases into one or
another of these categories (Siegel 2013).}

Given that the existence of top-down effects is largely to be settled by reflection
on science rather than purely from the armchair, you might think that problematic top-
down effects challenge “internalist” approaches in epistemology. When you are in a
problematic case of cognitive penetration after all, you might have no way of telling that
you are in it. For example, when Jill sees Jack, he looks angry to her, and she has no way
of telling that he looks that way because of her own belief. Internalist approaches in
epistemology might be thought to falsely predict that she is nevertheless justified in
believing that he is angry.

Much depends here on exactly how internalism is understood. Here we will
survey the two versions of internalism distinguished by Conee and Feldman 2001.

Conee and Feldman first identify a version of internalism understood in terms of
access--- “accessibilism”. The idea is roughly that, if something is invisible to you when
you introspect, then it doesn’t make a difference to what you have reason to believe. A
bit less roughly, the idea is that, if two people are the same with respect to what is
introspectively accessible to them, then they are the same with respect to what they have
reason to believe. So if two people differ in some way, but not in any way that is
available to introspection, that difference won’t affect what they have reason to believe.

Accessibilism predicts introspectively inaccessible top-down effects to be irrelevant to whether you are justified in believing any proposition. Compare the case of Angry Looking Jack to a benign case where Jack does look angry to Jill, but not as the result of any cognitive penetration. These two cases will be the same with respect to what is introspectively accessible, and yet---if Jill is not justified in the original case of Angry Looking Jack---different with respect to what Jill has reason to believe.

Regardless of how accessibilism fares, internalists might be able to take problematic cases of top-down effects on board. That’s because they might only endorse the second version of internalism discussed by Conee and Feldman (2001). This version is formulated in terms of metaphysics rather than epistemology. According to “mentalism”, the factors that determine whether a subject is justified in believing a proposition are mental. So if two people are mentally the same, then they will be the same with respect to what they have reason to believe (some versions of mentalism will privilege a narrower range of mental states). Now, if one person’s experience is the result of a top-down effect, and the other’s experience is not, they are not mentally the same. Different mental processes led to their experiences! Given the mental character of top-down effects, some versions of mentalism could allow an epistemic role for top-down effects.  

Having discussed the big picture approach of internalism, let’s now consider some more specific theories in the epistemology of perception.

According to “Liberalism”, experiences justify beliefs without the contribution of auxiliary beliefs (Pollock 1974, Pryor 2000, Huemer 2001). Liberal theories say that experiences are a source of non-inferential justification. Compare how your headache might justify you directly in believing you have a headache, and contrast how a comment in an online forum might justify you in believing its content only in conjunction with your (justified) background belief that the comment was made by a reliable source. There’s an open question as to why Liberal views must be challenged by any cases of top-down effects. As Pryor writes,

Why should the fact that your background beliefs causally affect what experiences you have show that the justification you get from those experiences relies on or derives from those background beliefs? Your sunglasses causally affect your experiences, but none of your perceptual beliefs are justified to any extent by your sunglasses. (2000: 540)

Even if background beliefs play a role in generating your experience, that need not imply that they play a role in justifying any of your beliefs justified by your experience (for further discussion see Siegel 2012 or Tucker 2014).

Some Liberal theorists such as Pryor 2000 or Huemer 2001 accept the following sort of Sufficiency thesis:

---

13 Versions of mentalism differ according to which mental entities they privilege. Versions that privilege only introspectively accessible or conscious mental entities might still be in trouble from some cases of top-down effects.
(Sufficiency): If you have an experience E with the content that p, and no evidence that anything is amiss with your experience, then you have justification to believe that p from E.

The Sufficiency thesis does face a threat from negative cases of cognitive penetration. Jill has an experience with the content that Jack is angry, and even though something is amiss with her experience, she has no evidence that anything is amiss with her experience. So the Sufficiency thesis predicts that she has justification from her experience to believe that Jack is angry. But some will insist against the Sufficiency thesis that she does not.

Liberal views do not stand or fall with the Sufficiency thesis. Liberal views say that your experiences sometimes give you non-inferential justification to believe that p, but they leave open what further factors need to be in place for that to happen. A Liberal theorist could say that, in order for an experience to give you justification to believe that p, the experience needs to be reliable. Now, someone could have an experience with the content that p, not have any evidence that something is amiss, when in fact their experiences fails to be reliable. Some Liberals can allow for counterexamples to Sufficiency, and some Liberals can take epistemically negative cases of top-down effects on board.

The denial of the Liberal theory is endorsed by the “Conservative” theory, on which the ability of experiences to justify beliefs is partly explained in terms of auxiliary beliefs (Wright 2002, Cohen 2002). On Conservative approaches, perceptual justification is in effect a special case of inferential justification. When you have an experience with the content that p, any justification you get from the experience to believe that p is routed through your background belief that, if you have an experience with the content that p, then p. Again compare the view on which, if you get justification from a comment in an online forum that p, that justification is routed through your background belief to the effect that, if the source of that comment says that p, then p. Just as Liberalism is compatible with tricky cases of top-down effects, Conservatism is too.

In the problematic cases of top-down effects we have seen, the subjects have no inkling that anything is amiss. That is how top-down effects generated a problem for accessibilist versions of internalism, and how they generated a problem for Sufficiency. Given that the subjects have no indication that anything is wrong, their justification for the background beliefs privileged by Conservatives is presumably perfectly in order. They should then have justification from their experience thanks to the availability of the following sort of inference:

I have an experience with the content that p.
If I have an experience with the content that p, then p.
So, p.

14 For discussion of top-down effects and reliabilist theories, on which the ability of experiences to justify beliefs is explained in terms of their reliability, see Goldman (2008), Lyons (2011), or Tucker (2014).
Their justification to believe the premises is perfectly in order. Presumably their justification to believe the conclusion is in order too, given that the conclusion manifestly follows from the premises, and there is no plausible candidate to prevent the justification for the premises to flow to the conclusion.

Does perceptual justification essentially depend on background beliefs? Do only mental factors play a role in perceptual justification? These questions seem relatively insensitive to whether there are epistemically negative cases of top-down effects. Do your perceptual experiences suffice for you to have perceptual justification? Do only introspectively accessible factors play a role in perceptual justification? These questions seem quite sensitive to whether are any epistemically negative cases of top-down effects.

**Conclusion**

Whether you think that there are top-down effects or not, you cannot escape the implications of debates about top-down effects for epistemology. If you deny the existence of top-down effects, you might take on commitments about our ability to know our own minds and our ability to form beliefs on the basis of our experiences. If you accept the existence of top-down effects, you might take on commitments about the failure of experiences to justify beliefs in some bad cases of top-down effects. These bad cases in turn have implications for a wide range of debates in the epistemology of perception.16

**References**


Byrne, Alex (forthcoming). The Epistemic Significance of Experience. *Philosophical Studies.*


16 For their help with this entry, thanks to Mike Blabac, Amber Carpenter, Qu Hsueh, Jay Garfield, Peter Kung, Fiona Macpherson, Matt McGrath, Neil Mehta, Jessie Munton, Keith Payne, Michael Pelzcar, Susanna Siegel, Weng Hong Tang, Lu Teng, Jona Vance, and David Wong.


Susanna Siegel (ms). The Rationality of Perception.