

# Cognitive Penetration

Lu Teng, New York University Shanghai, [lt77@nyu.edu](mailto:lt77@nyu.edu)

(This is a penultimate draft. Please cite the published version)

## 1. Introduction

In a study by Hansen et al. (2006), subjects had to adjust the color of a yellow banana until it looked achromatic, and the subjects tended to settle on a point slightly in the direction of blue—the opposite direction from yellow. This suggests that the subjects still experienced the banana as somewhat yellow at the achromatic point. One explanation is: their background cognition that bananas are typically yellow caused them to perceive an entirely grey banana as yellowish grey. Such a study is often cited to support the cognitive penetrability of perceptual experiences, broadly understood as the possibility of perceptual experiences being influenced by some non-perceptual psychological states, such as beliefs, expectations, emotions, desires, and so forth.

Epistemologists have also been interested in the phenomenon of cognitive penetration, not so much about whether there is good empirical evidence that it occurs, but rather about what epistemological implications it has for perceptual justification and knowledge *if* cognitive penetration ever occurs. Consider this hypothetical case from Siegel (2012):

**Anger:** Jill unjustifiably believes that Jack is angry with her. Unbeknownst to her, when she meets with Jack, this belief makes her experience Jack's face as expressing anger. However, Jack does not have an angry face at all.

One epistemological question to ask is: does Jill's experience give her justification to believe that Jack's face indeed expresses anger?

Notice that we might get more fine-grained versions of the anger case by distinguishing between different stages at which Jill's experience is cognitively penetrated. To bring out some of the possibilities, consider that perceptual experiences might have both low- and high-level contents, where the former are about colors, shapes, volumes, and so forth, and the latter are about kind properties, mental states, semantic properties, and so forth. When Jill's unjustified belief influences her experience of Jack's facial expression, one possibility is that the influence is directly on the low-level contents of her experience (e.g., the low-level features of Jack's facial muscles). Another possibility is that the influence is solely on the high- rather than low-level contents (i.e., the anger expression).

The epistemological question then becomes: in different versions of the case, does Jill's experience give her justification to believe the relevant cognitively penetrated contents? We survey three general approaches to this question in the sections below.

## **2. Reliabilism**

Many epistemologists reject that Jill's experience gives her justification to believe the relevant cognitively penetrated contents. If the influence is directly on the low-level contents, then Jill's experience fails to justify believing that Jack's facial muscles display such-and-such low-level features. And if the influence is solely on the high-level contents, then her experience fails to justify believing that Jack's face indeed expresses anger. What are some arguments? In this section, we discuss the reliabilist approach.

According to one version of reliabilism, an experience provides justification to believe that P only if there is a belief-forming process available from the experience to the belief that P and the process is reliable. Reliabilists can support their conclusions about the anger case with this principle of justification (Lyons, 2011 and 2016). The idea is that no matter whether the influence from Jill's unjustified belief is directly on the low-level contents of her experience, or solely on the high-level content, the cognitive penetration makes the relevant belief-forming process unreliable.

One merit of the reliabilist approach is its ability to readily accommodate epistemically beneficial/innocent cognitive penetration:

**Priming:** You wait for a friend at the train station. Your desire to see her primes your perceptual system and causes it to be more sensitive to her presence.

**Expertise:** When an expert looks at an eastern kingbird, her expertise causes her to experience it as a kingbird. A novice fails to have such an experience.

Reliabilists explain these cases by appealing to the fact that the cognitive influence makes the subjects' belief-forming process more reliable.

There are, however, concerns about the reliabilist approach. Consider a standard BIV, who is externally caused to experience that someone's face is angry. Compare it with a second BIV, who experiences that someone's face is angry due to being caused by its own unjustified belief that that person is angry. If the two BIVs both believe that there is an angry face in front of them based on their experience, then their beliefs are both unreliable. However, as highlighted by McGrath (2013a), the second BIV's belief appears even more epistemically compromised. This reveals that unreliability does not adequately capture the flaw in the second BIV's belief.

In light of this contrast, McGrath thinks that a different explanation is required for the original anger case, one that goes beyond mere unreliability.

Another concern about this approach is that it encounters a specific generality problem within the context of cognitive penetration. Tucker (2014) argues that if a perceptual process maintains very high reliability even when occasionally influenced by desires to produce wishful seeing, then reliabilism might need to acknowledge the justification of a belief formed through wishful seeing. Tucker's intuition, however, suggests a serious deficiency in such a belief. As a response, Lyons (2019) proposes a more fine-grained method for individuating belief-forming processes, which combines the taxonomy rooted in empirical cognitive psychology with extra variables.

### **3. Experiential Inferentialism**

An alternative significant approach to the epistemology of cognitive penetration draws a comparison between the etiologies of cognitively penetrated experiences and certain beliefs. In the anger case, Jill's experience has an etiology that structurally resembles an epistemically inappropriate belief inference; this explains why the experience fails to give Jill justification to believe the relevant cognitively penetrated contents (Siegel, 2013a and 2017; McGrath, 2013b). Such an approach is referred to as inferentialism. There are different ways to spell out the idea. We examine one proposal in this section and turn to another one in section 4.

McGrath (2013b) introduces the concept of quasi-inference: a transition between two experiences is quasi-inferential just in case upon replacing the experiences with corresponding beliefs, the outcome would be a genuine inference by the subject. McGrath posits that certain

cognitive penetration cases involve quasi-inferences. For example, if the influence from Jill's unjustified belief is solely on the high-level content of her experience, then it may involve the following quasi-inference:

(1) **Input experience with low-level contents:** Jack's facial muscles display such and such low-level features (that do not indicate anger).

(2) **Unjustified belief:** Jack is angry with Jill.

↓

(3) **Output experience with a high-level content:** Jack's face expresses anger.

McGrath does not require the input and output experiences to be distinct states. They could be sub-states within the same experience and occur together, but they nonetheless bear some inference-like relationship with each other.

McGrath thinks that a quasi-inference can be epistemically inappropriate in at least two different ways, mirroring the ways in which a belief inference might be epistemically bad. To begin with, if any of the inputs lacks justification, then the absence of justification extends to the output experience. Secondly, if the inputs inadequately support the output, then the quasi-inference hastily arrives at a conclusion. In the current example, one of the inputs, namely Jill's belief that Jack is angry with her, lacks justification. The lack of justification is then transferred to the resulting experience, elucidating why the latter fails to justify believing that Jack's face expresses anger.

A major objection to McGrath's inferentialism is that it fails to cover cases in which the cognitive penetration is directly to the low-level contents. Such an experience might very well lack justificatory power with respect to beliefs about the relevant low-level contents. However,

the lack of justificatory power could not be explained through the identification of any quasi-inference that moves from the low-level contents to high-level facets of the experience (Siegel, 2013b; Lyons, 2016).

One potential response is to contend that a quasi-inference could also occur between two distinct experiences (McGrath, 2013b, fn. 25). If the influence from Jill's unjustified belief is directly on the low-level contents of her experience, then there might be a quasi-inference from an initial experience that Jack's facial muscles display such and such low-level features (that do *not* indicate anger) to a subsequent experience that his facial muscles display such and such low-level features (that indicate anger), due to the influence from Jill's unjustified belief. The inadequacy of this quasi-inference accounts for why the second experience fails to justify believing the relevant low-level contents.

However, a challenge is that we lack empirical evidence that a transition between distinct experiences is implicated in cognitive penetration to the low-level contents (Long, 2018; Teng, 2021).

#### 4. Subpersonal Inferentialism

Siegel (2013a and 2017) presents another version of inferentialism. According to Siegel (2013a)'s theory, if the influence from Jill's unjustified belief is solely on the high-level content of her experience, then the etiology might be described as follows:

(4) **Unjustified belief:** Jack is angry with Jill.

↓

(5) **Output experience with a high-level content:** Jack's face expresses anger.

Siegel suggests that this etiology bears structural similarity to an inference from an unjustified belief that Jack is angry with Jill to another belief that Jack's face expresses anger. In the context of belief inference, when the input belief lacks justification, the output belief inherits the same lack of justification. In the context of experience, a structurally analogous etiology also makes the resulting experience lack justificatory power.

Moreover, there is a corresponding explanation if the influence is directly on the low-level contents:

(6) **Unjustified belief:** Jack is angry with Jill.

↓

(7) **Output experience with low-level contents:** Jack's facial muscles display such and such low-level features (that indicate anger).

The etiology might be likened to an inference from an unjustified belief that Jack is angry with Jill to another belief that Jack's facial muscles display such-and-such low-level features (that indicate anger). Similar to the former case, an unjustified input transfers its lack of justification to the output.

One objection to Siegel (2013a)'s proposal is that it faces a problem akin to the widely-discussed generality problem for reliabilism: the etiology of a cognitively penetrated experience can be type-individuated in multiple ways, of which some are more abstract than others (Lyons, 2011 and 2016). Consider that Jill's unjustified belief might cause her perceptual system to be more sensitive to an angry expression. The resulting experience could indeed have justificatory power, although the etiology can be compared to an epistemically bad inference, which moves from an unjustified belief that Jack is angry to another belief that Jack's face expresses anger.

Siegel (2013a)'s theory also encounters difficulty in explaining epistemically pernicious cognitive penetration from a justified belief (Lyons, 2016). Consider a hypothetical case based on Hansen et al. (2006):

**Banana:** A banana picture is completely grey. However, due to the influence from a justified background belief that bananas are typically yellow, the subjects perceive the picture as somewhat yellow.

The etiology of the subjects' experience can be compared to an epistemically good inference, which moves from a justified belief that bananas are typically yellow to another belief that this banana is somewhat yellow. Depending on the specific psychological mechanism underlying the cognitive penetration, however, the justificatory power of the resulting experience might be compromised.

In her recent book, Siegel (2017) modifies her proposal by allowing not only personal-level psychological states, such as beliefs and desires, but also unconscious, subpersonal states postulated by psychological explanations of perception to play a role in individuating etiologies. Moreover, Siegel (2017) now contends that perceptual experiences can themselves result from inferences. Her new proposal has the advantage of being able to elucidate the epistemological implications of various cognitive penetration cases. For example, the banana case above might involve this perceptual inference:

(8) **Unconscious, subpersonal state:** This banana is completely grey.

(9) **Justified belief:** Bananas are typically yellow.

↓

(10) **Output experience:** This banana is somewhat yellow.



The subjects' belief fails to bridge the gap between the unconscious, subpersonal state and the resulting experience. The inference hence jumps to the conclusion, and the experience fails to give the subjects justification to believe the relevant content.

However, Siegel (2017)'s proposal faces some new challenges, of which one is that the theory might end up implying that numerous ordinary experiences are caused by epistemically bad perceptual inferences, and might lead to extensive skepticism. Teng (2021) illustrates this problem by applying Siegel (2017)'s proposal to Bayesian theories of perception. In generating an experience, a Bayesian perceptual inference might move from an unconscious, subpersonal state that contains information about the posterior probability of a hypothesis,  $H$ , to a non-probabilistic experience that  $H$ . Such an inference seems to hastily arrive at a conclusion. Even if the posterior probability of  $H$  is extremely high, reducing the hypothesis to a simple truth in the resulting experience overlooks the probabilistic subtleties present in the input state.

## **5. Phenomenal Conservatism/Dogmatism**

So far, we have investigated two different approaches to the epistemology of cognitive penetration: reliabilism and inferentialism. Both acknowledge that some cognitively penetrated perceptual experiences fail to provide justification due to inappropriate etiologies. In this final section, we examine the third approach—phenomenal conservatism or dogmatism (hereafter “dogmatism”), which maintains that in the anger case, Jill's experience can nonetheless justify believing that Jack's face expresses anger.

According to dogmatism, absent defeaters, undergoing a perceptual experience that  $P$  can give the subject justification to believe that  $P$  merely in virtue of the unique phenomenal

character of the experience (Huemer, 2001; Pryor, 2000). Dogmatists might emphasize that if Jill is unaware of the influence from her unjustified belief on her experience, then there are no relevant defeaters, and Jill's experience should be sufficient to confer perceptual justification (Huemer, 2013).

To support such a conclusion, dogmatists might appeal to the lack of an epistemic status other than justification, such as knowledge or blamelessness, to account for the deficiency in Jill's belief formed based on her experience (Huemer, 2013; Tucker, 2010). Moreover, another argument provided by dogmatists points out that if the influence from Jill's unjustified belief had already made her experience devoid of justificatory power, then learning about the etiology would not worsen the justificatory status of her newly formed belief. However, learning about the etiology does seem to worsen the belief's justificatory status. This prompts us to cast doubt on the idea that the mere occurrence of the bad etiology "destroys" the justificatory power of Jill's experience (Silins, 2020).

Some dogmatists think that their theory is compatible with certain negative conclusions about cognitive penetration cases. Chudnoff (2019), for example, argues that in the anger case, if the influence from Jill's unjustified belief is solely on the high-level content of her experience, then it gives her defeasible justification to believe that Jack's face expresses anger. In addition, the part of Jill's experience that represents low-level properties gives her defeasible justification to believe that Jack's face displays such and such low-level features (that do *not* indicate anger). The latter justification constitutes a defeater to the former justification. Consequently, Jill lacks *all things considered* justification to believe that Jack's face expresses anger.

On the other hand, Chudnoff claims that if the influence from Jill's unjustified belief is directly on the low-level contents of her experience, then it gives her defeasible justification to

believe both that Jack's face displays such and such low-level features (which indicate anger), and that Jack's face expresses anger. In this case, the justification provided by Jill's experience is not defeated.

## References

Chudnoff, E. (2019) "Experience and Epistemic Structure: Can Cognitive Penetration Result in Epistemic Downgrade?" in T. Chan & A. Nes (eds.) *Inference and Consciousness*, New York: Routledge.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. (2006) "Memory Modulates Colour Appearance," *Nature Neuroscience* 9(11): 1367-1368.

Huemer, M. (2001) *Skepticism and the Veil of Perception*, New York: Rowman and Littlefield Publishers, Inc.

Huemer, M. (2013) "Epistemological Asymmetries between Belief and Experience," *Philosophical Studies* 162(3): 741-748.

Long, R. (2018) "How Wishful Seeing is Not Like Wishful Thinking," *Philosophical Studies* 175(6): 1401-1421.

Lyons, J. (2011) "Circularity, Reliability, and the Cognitive Penetrability of Perception," *Philosophical Issues* 21(1): 289-311.

Lyons, J. (2016) "Inferentialism and Cognitive Penetration of Perception," *Episteme* 13(1): 1-28.

Lyons, J. (2019) "Algorithm and Parameters: Solving the Generality Problem for Reliabilism," *Philosophical Review* 128(4): 463-509.

- McGrath, M. (2013a) "Siegel and the Impact for Epistemological Internalism," *Philosophical Studies* 162(3): 723-732.
- McGrath, M. (2013b) "Phenomenal Conservatism and Cognitive Penetration: the 'Bad Basis' Counterexample," in C. Tucker (ed.) *Seemings and Justification*, New York: Oxford University Press.
- Pryor, J. (2000) "The Skeptic and the Dogmatist," *Noûs* 34(4): 517-549.
- Siegel, S. (2012) "Cognitive Penetrability and Perceptual Justification," *Noûs* 46(2): 201-222.
- Siegel, S. (2013a) "The Epistemic Impact of the Etiology of Experience," *Philosophical Studies* 162(3): 697-722.
- Siegel, S. (2013b) "Reply to Fumerton, Huemer, and McGrath," *Philosophical Studies* 162(3): 749-757.
- Siegel, S. (2017) *The Rationality of Perception*, New York: Oxford University Press.
- Silins, N. (2020) "Reading the Bad News about Our Minds," *Philosophical Issues* 30(1): 293-310.
- Teng, L. (2021) "Cognitive Penetration: Inference or fabrication?" *Australasian Journal of Philosophy* 99(3): 547-563.
- Tucker, C. (2010) "Why Open-Minded People Should Endorse Dogmatism," *Philosophical Perspectives* 24(1): 529-545.
- Tucker, C. (2014) "If Dogmatists Have a Problem with Cognitive Penetration, You Do Too," *Dialectica* 68(1): 35-62.