

A defense of cognitive penetration and the face-race lightness illusion¹

Kate Finley

To cite this article: Kate Finley (2022): A defense of cognitive penetration and the face-race lightness illusion¹, *Philosophical Psychology*, DOI: [10.1080/09515089.2022.2083591](https://doi.org/10.1080/09515089.2022.2083591)

To link to this article: <https://doi.org/10.1080/09515089.2022.2083591>



Published online: 30 Jul 2022.



Submit your article to this journal [↗](#)



Article views: 15



View related articles [↗](#)



View Crossmark data [↗](#)



A defense of cognitive penetration and the face-race lightness illusion¹

Kate Finley 

Department of Philosophy, Hope College, Holland, Michigan, United States

ABSTRACT

Cognitive Penetration holds that cognitive states and processes, specifically propositional attitudes (e.g., beliefs), sometimes directly impact features of perceptual experiences (e.g., the coloring of an object). In contrast, more traditional views hold that propositional attitudes do not directly impact perceptual experiences, but rather are only involved in interpreting or judging these experiences. Understandably, Cognitive Penetration is controversial and has been criticized on both theoretical and empirical grounds. I focus on defending it from the latter kind of objection and in doing so, highlight important features of Cognitive Penetration mechanisms and effects. I first sketch promising criteria for Cognitive Penetration and then address widespread Replication and Demonstration worries about purported instances of it. Next, I present one of the most compelling Cognitive Penetration studies and address specific objections against it and others. I demonstrate that each of these objections misunderstands important features of either Cognitive Penetration mechanisms or the specific studies to which they are applied. Ultimately, I conclude that multiple key studies are not undermined by these objections and continue to provide support for Cognitive Penetration. Correcting these misunderstandings bolsters empirical support for Cognitive Penetration and contributes to a better understanding of the mechanisms involved in perceptual processing.

ARTICLE HISTORY

Received 15 July 2020
Accepted 17 May 2022

KEYWORDS

Cognitive penetration;
perception; race; color;
vision; illusion

1. Introduction

We often take it for granted that, except in cases of illusory perception, perception enables us to experience much of the world as it really is and thus provides us with at least *prima facie* evidence that the world is the way that it appears to us. According to traditional accounts of perception, it informs us about the external world in a way that is relatively unaffected by what we “bring to the table”, so to speak – our beliefs, expectations, desires, etc. These and similar cognitive states and processes only get involved *after* perception occurs (e.g., we may develop desires as a result of perceptions or use our beliefs to interpret perceptions).

Cognitive Penetration refers to a phenomenon that challenges the aforementioned understanding of the relationship between perception and cognition. Mainstream Cognitive Penetration theorists hold that a cognitive state like a belief or a desire may penetrate a perceptual experience, and thereby directly affect some of its phenomenal qualities.² For example, the following are purported instances of Cognitive Penetration: a more valuable dollar bill appears closer than a less valuable one because of one's desire for it, and a banana appears yellower than it is, because of one's belief that bananas are yellow.³ According to advocates of Cognitive Penetration, two individuals in equivalent viewing conditions – including the object of perception, the physical state of their sensory organs, and any other perceptually relevant contextual features (e.g., the lighting surrounding the object) – can have two qualitatively different perceptual experiences due to the impact of different penetrating cognitive states.⁴ Cognitive Penetration theorists include philosophers Dustin Stokes (2013, 2015, 2020), Wayne Wu (2013, 2017), Fiona Macpherson (2012, 2015), Susanna Siegel (2012), and Paul Churchland (1988); and psychologists Levin and Banaji (2006, p. 2016), Delk and Fillenbaum (1965), among others.⁵

In contrast, many object to Cognitive Penetration, instead claiming that cognitive states like beliefs and desires cannot directly impact our perceptual experiences but can at most be used to interpret and judge experiences. Some object on theoretical grounds, like Jerry Fodor (1984, 1988) and others, who are primarily concerned with issues of cognitive architecture and encapsulation. However, this focus on cognitive architecture in the Cognitive Penetration literature (both on the part of its defenders and detractors) has more recently given way to focus on the phenomenological effects of Cognitive Penetration and the epistemological implications of those effects.⁶ Important questions about epistemological implications include: does the existence or pervasiveness of Cognitive Penetration undermine the possibility of theory-neutral observation? What role can cognitively penetrated experiences play in justifying beliefs about the external world? What social implications might some of the purported effects of Cognitive Penetration have? Siegel's well-known illustration concisely highlights the epistemic circularity worries motivating some of these questions: Jill believes Jack is angry, Jill sees Jack and, as a result of Cognitive Penetration, Jack appears angry to Jill, Jill then takes her perceptual experience as evidence for her belief that Jack is angry.⁷ These worries have given rise to objections that target empirical support for Cognitive Penetration from those like Chaz Firestone and Scholl (2014, 2015, 2016), and Edouard Machery (2015).

In what follows I will sketch a set of three criteria for Cognitive Penetration. There are many different versions of each criterion, some of which I will address before providing reasons to favor the particular criteria that I focus on. After addressing the criteria for Cognitive Penetration, I will

address multiple key objections that have been posed against empirical support for it. First, I will address the *Replication and Demonstration worries* which challenge empirical support for Cognitive Penetration on a general level. Notably, Levin & Banaji's well-known study is not vulnerable to these worries, and it is on this study that I will focus throughout the rest of the paper. It is one of the most cited by supporters of Cognitive Penetration, and most criticized and scrutinized by its objectors. I will detail its four experiments, explain how it fulfills the criteria for Cognitive Penetration, and how it survives the remaining objections that I address: the *El Greco* objection, the *judgment Effects* objection, and the *Low-level Perceptual Effects* objection. I will argue that each of these objections misunderstand important features of either the Cognitive Penetration mechanisms or the specific experiments to which they are applied and ultimately do not undermine Levin & Banaji's study, nor multiple other key Cognitive Penetration studies.

2. Criteria for cognitive penetration

Although there is a lack of consensus over a definition of Cognitive Penetration, most often, criteria for the phenomenon address the following questions.⁸ In order for a particular instance to count as Cognitive Penetration: 1) At what stage in the perceptual process must the cognitive state influence the resulting perceptual state? 2) What must the relationship be between the relevant cognitive and perceptual states? 3) Is the influence of the cognitive state on the perceptual state undermined by any external factors? The definition I will sketch and defend is one according to which the three questions above are answered in the following criteria:

Cognitive Penetration occurs iff a cognitive state causes a change in phenomenal features of a subject's perceptual experience such that 1) the cognitive state penetrates processes in the *early stages* of perception; 2) the perceptual experience draws on the cognitive state as an *informational resource*; and 3) this change is not undercut by *explanatory defeaters*.⁹

I will now say a bit more about each of these three criteria.

2.1 Early stages

This criterion addresses at what stage in the perceptual process the cognitive state needs to impact the resulting perceptual state. Although there are disagreements over how to distinguish early and late vision, most agree on a few characteristics of these stages.¹⁰ *Early Vision* processes are those that take place during between the first 120–150 ms after the initial stimulus and are often defined functionally as those that receive bottom-up feedback, including information about the stimulus' color, shape, size, and motion.

Late Vision processes are those that take place between 120–150 ms and 200 ms after the initial stimulus and are involved in more “recognitional” capacities that draw on semantic or representational information which aids in the interpretation and/or categorization of stimuli. Importantly, the processes involved in both early and late vision are not consciously accessible to the subject, nor are they under the subject’s control – this kind of control, and indeed access to a perceptual experience only occurs at least 200 ms after the initial stimulus.

Pylyshyn and many others hold that the cognitive state must penetrate the *early stages* of visual processing, because only these processes are proprietary to the visual system and are involved in true visual awareness as opposed to visual understanding.¹¹ Although I do not find reasons given for restrictions to early vision particularly compelling, because both early and late vision processes contribute to phenomenal features of our perceptual experiences, in order to make the strongest case for Cognitive Penetration I will adopt this criterion going forward.

2.2 Informational resource

Next, we turn to the many criteria that have been proposed to address the nature of the connection between the penetrating cognitive state and the penetrated perceptual state. Some relatively uncontroversial criteria require that the connection between the cognitive and perceptual states be both internal and direct. However, more controversial criteria address the connection between the content of these states. Macpherson proposes that there must exist some kind of intelligible connection between the relevant cognitive and perceptual states: the connection between the content of these states must be readily understandable.¹² However, many find this too vague, including Wu who instead proposes the *Informational Resource* criterion, according to which, in order to be an instance of Cognitive Penetration, a perceptual state must draw on the relevant cognitive state as an informational resource.¹³

To see the difference between these two criteria, take the following example: a subject believes that she will perform poorly on an upcoming presentation and then experiences blurry vision. In this case the subject’s cognitive state clearly impacts the phenomena of her perceptual experience; however, this is not a case of Cognitive Penetration. The “Intelligible Connection” criterion fails to capture this: the connection between the contents of the two states *is* readily understandable or intelligible because this effect is well-known to occur in performance anxiety situations. However, the *Informational Resource* criterion does: although the content of the cognitive state “(I believe) I will perform poorly” helps cause the resulting perceptual state, it does not do so in virtue of the information it

provides but rather due to the fact that it causes anxiety, which results in increased adrenaline, which in turn increases stress on the eyes, resulting in blurred vision.

2.3 Explanatory defeaters

Lastly, the third set of criteria address the amount and kind of influence that external factors can have on perceptual states in instances of Cognitive Penetration. Some hold that the viewing conditions of a subject's perceptual experience – including the object of perception, the state of the sensory organ (whether or not it is properly functioning), contextual features (e.g., the surrounding lighting), and attention of the subject (what objects or aspects of objects the subject focuses on) – must be held fixed, in order for something to count as an instance of Cognitive Penetration. But this criterion problematically excludes instances in which Cognitive Penetration mechanisms *and* external factors may both be partially responsible for the relevant perceptual effects; as well as instances in which Cognitive Penetration mechanisms may be fully responsible for the relevant effects but co-occur with some change in the aforementioned viewing conditions.

In order to recognize situations like this as legitimate instances of Cognitive Penetration, Wu, Macpherson, and Stokes have all proposed some version of the *No Explanatory Defeaters* criterion. This criterion holds that while a change in viewing conditions may occur in a legitimate instance of Cognitive Penetration, this change cannot be *solely* responsible for the relevant effect on the subject's perceptual experience.¹⁴ Rather, some portion of the perceptual effect must be left unexplained such that they can be appropriately attributed to Cognitive Penetration. To see the difference between these criteria take an example in which a subject a banana as yellower than it is while in a room with yellow lighting. The “Fixed Viewing Conditions” criterion would immediately rule this out from counting as a legitimate case of Cognitive Penetration. However, while the yellow light may serve as an explanatory defeater, it need not. Cognitive Penetration mechanisms could in fact be partially responsible for the more intensely yellow color of the banana, in which case, *No Explanatory Defeaters* criterion would correctly categorize this as a legitimate instance of Cognitive Penetration.

3. Empirical support for cognitive penetration

Now that I have sketched the criteria for Cognitive Penetration, I turn to address empirical support for Cognitive Penetration and the objections leveled against it. In what follows, I will present and address some of the

more notable challenges to empirical support for Cognitive Penetration offered by Firestone, Scholl, and Machery. Some see these objections as delivering decisive blows against Cognitive Penetration – for example, after presenting their objections to many well-known Cognitive Penetration studies, Firestone & Scholl conclude that “*none* of these hundreds of studies – either individually or collectively – provide compelling evidence for true top-down effects on perception”, and that “there is in fact *no* evidence for such top-down effects of cognition on visual perception, in *every sense* these claims intend” (emphases mine).¹⁵ While each of the objections I address below undermine *some* of the empirical support for Cognitive Penetration, these claims are overstated. Many of the objections reveal misunderstandings of key aspects of Cognitive Penetration mechanisms and/or the studies they are applied to. Crucially, they fail to undermine the support provided by Levin & Banaji’s study as well as many of the specific studies that Firestone, Scholl, and Machery attempt to target with these objections.

3.1 Replication & demonstration worries

Firestone, Scholl, and Machery all introduce Replication and Demonstration worries about empirical support for Cognitive Penetration. According to these worries, because attempts to recreate the results of key Cognitive Penetration studies, in both experimental and non-experimental settings (replication and demonstration attempts, respectively), have failed, we ought to decrease our confidence in these results and the support they provide for Cognitive Penetration.

One experiment that Machery claims is undermined by Replication worries, is Hansen et al.’s (2006) study, *Memory Modulates Color Appearance*. In some trials, subjects are presented with realistic, “characteristically-colored” computer images of fruits (e.g., an image of a yellow banana) and instructed to adjust the coloring of the images until they are achromatic. Researchers found that subjects consistently over-adjusted the images past the point of achromaticity toward the color opposite the fruit’s typical or characteristic color (e.g., over-adjusted the image of the yellow banana to be bluish). Explanations according to which this effect is due to Cognitive Penetration maintain that a subject’s cognitive state (such as her belief that bananas are yellow) penetrates her perceptual experience of the banana image, such that she perceives it to be yellower than it is; consequently, when the photo is achromatic, it still appears slightly yellow to her, which causes her to adjust the image to be bluer in order to “balance out” the yellow she perceives and make it appear achromatic. Machery claims that because a later trial of the experiment – in which subjects were presented with line drawings of fruit – did not find as strong of an effect, this amounts

to a replication failure and thus challenges the rest of Hansen et al.'s results, as well as the results of similar key Cognitive Penetration experiments like Delk & Fillenbaum's seminal fruit color experiments.¹⁶ He claims that because, in these and other similar experiments, replication attempts failed, this ought to call into question the support that these experiments provide for Cognitive Penetration.¹⁷

At first glance, this objection seems quite worrisome; however, it is ultimately not as problematic as it appears, because we can explain away these and other purported replication failures by noting overlooked features of Cognitive Penetration mechanisms and of the targeted experiments. These features provide compelling reason for why this "failure" occurred, and thus why it ought not undermine the support that the original experiment provides for Cognitive Penetration. One crucial feature in the above experiment is the difference in the lifelikeness of the images used in Hansen et al.'s original experiment and the purported replication. If the results of the original experiment, in which subjects were presented with life-like images of bananas, were due to Cognitive Penetration, it is likely that the cognitive state that penetrated subjects' perceptual experiences – a belief that "bananas are yellow" – was informed by and based on the subjects' past experiences of real bananas. Thus it is plausible that the color-intensifying effects on the subjects' perceptual experiences are stronger when the object perceived (e.g., a picture of a real banana vs. a relatively unrealistic line drawing of a banana) more strongly resembles the objects of past experiences on which the penetrating cognitive state is based (e.g., real bananas). Thus, weaker effects in the purported replication are what we might expect if Cognitive Penetration mechanisms are responsible for the perceptual effects at issue.

A related worry concerns the apparent failure of stimuli used in Cognitive Penetration studies to serve as demonstrations – in other words, Cognitive Penetration effects are not readily apparent to someone observing the stimuli used in these studies in non-experimental, everyday settings.¹⁸ Firestone & Scholl highlight two apparent demonstration failures, writing,

"the possibility that valuable items look closer is testable not only in a laboratory (e.g., Balcetis & Dunning, 2010), but also from the comfort of home: right now you can place a \$20 bill next to a \$1 bill and see for yourself whether there is a perceptual difference . . . Similarly, knowledge of an object's typical color (e.g., that bananas are yellow) reportedly influences that object's perceived color, such that a grayscale image of a banana is judged to be over 20% yellow (Hansen et al., 2006; Olkkonen et al., 2008); however, if you look now at a grayscale image of a banana, we predict that you will not experience this effect for yourself – even though the reported effect magnitudes far exceed established discrimination thresholds . . ." ¹⁹

The worry here, similar to that above, is that these failures of demonstration cast doubt on the relevant studies and that the overall lack of demonstrations casts doubt on the phenomenon of Cognitive Penetration.²⁰ Again, I think

that the instances referenced above ought not be understood as “demonstration failures” and furthermore, that viewing them as such results from overlooking and/or misunderstanding important features of Cognitive Penetration and the studies themselves. Both of these “demonstration suggestions” from Firestone & Scholl differ in crucial ways from their laboratory counterparts, such that we ought not expect them to provide demonstrations of the Cognitive Penetration effects found in their laboratory counterparts. The important difference is that ambiguities that exist in the experimental versions of these studies are eliminated in the proposed demonstrations, which is important because Cognitive Penetration effects are significantly more likely to occur, stronger and/or more salient when aspects of the observed stimuli are degraded, ambiguous or otherwise uncertain.²¹

In Balcetis and Dunning’s (2010) *Wishful Seeing: motivational influences on visual perception of the physical environment*, subjects were asked to estimate the distance between themselves and a \$1 bill, and then to estimate the distance between themselves and a \$20 bill – in two separate instances, thus the difference between the subject’s distance from each of the bills is ambiguous. However, the proposed demonstration above would have subjects estimate the distance between themselves and both dollar bills, side-by-side, at the same time. This resolves the aforementioned ambiguity because the subject herself would be placing these bills such that they are the same distance away from her. A similar difference exists between Hansen et al.’s (2006) original experiments and the proposed demonstration. In the original experiment, when subjects are adjusting the color of the banana image, it remains ambiguous to subjects when it reaches achromaticity. However, in the proposed demonstration, this ambiguity would be eliminated because the subject would be aware of the banana image’s achromaticity (according to the description above, she would be explicitly looking at an image of an achromatic banana). Thus, this might plausibly result in the elimination or dampening of the relevant Cognitive Penetration effects. Furthermore, the information available to the subject in the demonstrations: that the bills were equidistant from her, and that the banana image was achromatic might also contribute to postperceptual or judgment effects that mask or counteract potential Cognitive Penetration effects (e.g., the achromatic image may look somewhat yellowish but the subject judges that it is achromatic). Thus, I maintain that both demonstration proposals are unsuitable as such and should not be thought of as “demonstration failures” which undermine the original experiments.

Any suitable demonstration proposals would end up having to look quite similar to their laboratory counterparts in order to preserve the ambiguities mentioned above. For example, a suitable demonstration of Balcetis & Dunning’s study would be one in which a subject placed different value dollars in front of herself one-by-one, estimated their distance, and then

compared the estimations. Similarly, for a demonstration of the effects found by Hansen et al. a subject would need to adjust the color of on-screen images of bananas while remaining unaware of when they reach achromaticity. Or perhaps the following might also be a suitable demonstration because it preserves the ambiguity about level of chromaticity: a subject looks at many different versions of the banana image – all adjusted to different levels of chromaticity (e.g., one at 30% chromaticity, 15%, 0%, –10%) and sees if she can pick out the truly achromatic one – or if she ends up picking out images that are more bluish.

Furthermore, although I do not have space to address them in detail here, there are similar features that can be used to explain why many other purported “demonstration proposals” (instances that we might expect to serve as a demonstration of Cognitive Penetration) connected to Cognitive Penetration studies fail to deliver. In other words, there are similar compelling reasons for why the Cognitive Penetration effect might occur in the experimental version but not in the demonstration version. However, it still seems like there are other “demonstration proposals”, disconnected from experimental Cognitive Penetration studies, that are not susceptible to the worries address above of “diminished ambiguity”, but which nevertheless do not seem to demonstrate. For example, take the following case.²² Say that you see an object in a fruit bowl across the room that you at first believe to be a lemon but as you get closer realize is actually a tennis ball. According to Cognitive Penetration, it seems like once you have this realization (and thus shift your belief from “that is a lemon” to “that is a tennis ball”) you should experience a corresponding shift in color phenomenology, since lemons and tennis balls are different colors. However, intuitively it doesn’t seem like you would experience such a shift. Thus, this case, which notably preserves the ambiguity unlike the demonstration proposals above – because you at first are not made aware of whether the object is a lemon or a tennis ball – appears to be a failed demonstration.

This is an interesting case that at first glance, does seem to constitute a “demonstration failure”. However, a couple of things are important to keep in mind when evaluating our intuitions about whether we would experience a perceptual shift in this case (and similar cases). First, if such a shift in color phenomenology did occur, it would be quite subtle. This is because of the necessarily small difference between the colors involved: the colors of a lemon and a tennis ball are relatively close together – thus the expected shift in color would only be from yellow to a greenish-yellow. Indeed, in order for this case to be plausible (confusing tennis ball for a lemon) the colors would need to be quite close together – a situation in which you confused a lime for a lemon would be much less plausible because of the larger distance between the colors involved. Second, it would actually be fairly difficult to notice whether you experienced the color shift in question without a direct comparison or standard against which to match your color perception. So, for

example, the alleged color shift might be much more noticeable if there were two objects in the fruit bowl, both of which you originally believed to be lemons, but then as you got closer, realized one was a lemon and one was a tennis ball. In this situation, the shift in color of the lemon mistakenly thought to be a tennis ball would actually be noticeable because you would have a reference point against which to mark the change.

However, I would argue that in a more convoluted situation like this, we ought to be even less confident in our intuitions about whether we would or would not experience a shift in color phenomenology. I have never experienced a situation remotely like this as far as I can tell and do not have strong intuitions about whether or not I would experience a shift in color phenomenology. All this to say, although I think this is an interesting case that it would be fascinating to study empirically, there are compelling reasons to doubt our intuitions that we would *not* experience the predicted shift in color phenomenology in this case. In this situation we may experience a shift in color phenomenology and yet this shift may, for good reason, not be salient to us. Like the explanations above, this does not *prove* that we would experience the proposed demonstration in the relevant case, but rather gives compelling reasons based on the purported mechanisms of Cognitive Penetration, that we *might*, appearances to the contrary. Thus, either empirical studies would be needed to test such a case or more would need to be said to defend initial intuitions about this case for it to be a more compelling demonstration failure.

Finally, one might still think that, while addressing individual proposed demonstration failures is well and good, there is still the broader worry about the relative dearth of compelling demonstrations of Cognitive Penetration compared to what we might expect if the phenomenon existed. Although I do not have space here to respond to this worry in full – it is worth briefly sketching a response that I make in more detail elsewhere in which I articulate ways in which Cognitive Penetration is a “self-obscuring” phenomenon and how this provides a plausible explanation of the aforementioned paucity of notable demonstrations. By “self-obscuring” I mean that, if Cognitive Penetration exists, the operation of its mechanisms minimize evidence of that operation. This is because purported Cognitive Penetration mechanisms often encode an individual’s expectations, and thus often result in perceptual effects that are minor and unsurprising, and thus largely inconspicuous. For example, the penetrating cognitive state mentioned above that “bananas are yellow” comes from a subject’s past experiences with bananas which have informed her expectations that future bananas she encounters will also be yellow. Although this cognitive state may penetrate the subject’s perceptual experience such that she sees the banana as yellower than it is, this accords with her expectations and is thus less noticeable (perhaps entirely unnoticeable in non-experimental situations). Furthermore, these expectations are often widely shared. For example, most people are likely to share the belief



Figure 1. SEQ Figure * ARABIC 1.



Figure 2. SEQ Figure * ARABIC 2.

that bananas are yellow and thus have their perceptions cognitively penetrated in similar ways. This then makes it less likely that there will be a noticeable difference between the Cognitively Penetrated perceptual experiences of different individuals, and this in turn further contributes to the inconspicuous nature of these effects. Thus, understanding Cognitive Penetration as a “self-obscuring” phenomenon helps make sense of how it might be true both that Cognitive Penetration exists, and is even quite pervasive, and that there is a dearth of salient demonstrations of it.

To sum up, although some empirical support of Cognitive Penetration may be challenged by the Replication and Demonstration worries discussed above, at least the specific studies targeted by the objectors are not. Furthermore, the “self-obscuring” nature of Cognitive Penetration mechanisms may help further explain a relative lack of salient demonstrations. Finally, it is notable that Levin & Levin and Banaji’s (2006) study does not fall prey to replication concerns and provides very clear demonstrations (see, [Figures 1 and 2](#) below). Because of this, the study is one of the most commonly cited as providing compelling support for Cognitive Penetration by its defenders, and one of the most critiqued by its objectors.²³ Thus, it is to this study that I now turn: I will present it in detail as well as how it fulfills the criteria for Cognitive Penetration outlined above, and then will defend it against the remaining objections to Cognitive Penetration that I address below.

3.2 Distortions in the perceived lightness of faces: The role of race categories

In this study, Levin & Banaji address “how people perceive the shading of faces of different races”.²⁴ Before each of the four experiments in the study, subjects were informed of the purpose of the study and asked to do a variety of “matching tasks” involving “Black”, “White”, and “Ambiguous”

prototype faces. The Black and White faces were each created by “blending a set of 16 faces from the race the prototype was to represent” while the Ambiguous face was a composite of features from the Black and White faces. At the end of each experiment, subjects completed a brief questionnaire that measured their explicit attitudes toward different races.

Experiment 1: Testing the lightness distortion effect

In Experiment 1, the subject viewed two faces (a ‘reference face’ and an ‘adjustable face’) for each trial. They were instructed to adjust the adjustable face until it matched the reference face in terms of lightness. Note that ‘lightness’ refers to the subjective perception of shading, while ‘luminance’ refers to the objective level of shading. On half of the trials, the faces were of the same race (‘same-race’ trials), on the other half, they were of different races (‘mixed-race’ trials).²⁵

Experiment 2: Removing stimulus artifacts

In Experiment 2, an additional “Ambiguous” face was added and labeled either “Black” or “White”, depending on the trial. In each trial, the subject was presented with one face and a uniformly gray rectangle and was instructed to adjust the rectangle to match the face in terms of lightness.²⁶

Experiment 3: Eliminating attentional the attentional focus alternative hypothesis

In Experiment 3, the experiment was adjusted to reduce potential attention-mediated effects. Levin & Banaji note the possibility that, in previous experiments, subjects’ lightness perception may have been affected by their focus on different parts of the faces.²⁷ To eliminate this possibility they used faces with consistent luminance throughout and alternated between faces with either brighter or darker outlines in different trials in order to minimize the impact of further attentional effects (see [Figure 3](#)). Subjects were again instructed to match a uniformly gray rectangle to the face in terms of lightness.²⁸

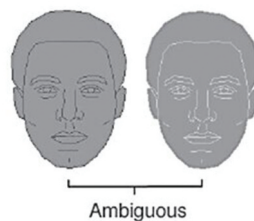


Figure 3. SEQ Figure * ARABIC 3.

Experiment 4

Experiment 4 was designed to address worries about postperceptual editing and/or demand characteristics potentially contributing to the effects seen in the previous three experiments. A subject may have perceived different faces to be the same level of lightness, but then postperceptually “adjusted” some to be darker or lighter based on a belief like “White faces tend to be lighter and Black faces tend to be darker”. In such instances, these beliefs may influence a subject’s postperceptual adjustments, while not affecting her perception itself: in such instances, subjects adjust the images “to be consistent with their *understanding* of the faces rather than what they currently *see*” (emphasis mine).²⁹ To address this concern, Levin & Banaji presented subjects with pairs of faces that were either of the same race (pairs of the identical Black or White faces), or pairs of faces from different races – and they varied the luminance in both of these kinds of pairs. They instructed the subjects to respond as quickly as possible whether the faces were the “same” or “different” races, and not to use mismatches in lightness as a basis for “difference” judgments.³⁰

Discussion

In the first three experiments in this study, Levin & Banaji, found that “White faces were consistently judged to be relatively lighter than Black faces, even for racially ambiguous faces that were disambiguated by labels.”³¹ And in the fourth experiment, they found that subjects’ reaction speed of correctly selecting “same” or “different”, varied with the lightness (*perceived* lightness or darkness) rather than the luminance (*actual* lightness or darkness).³² Both Cognitive Penetration and its objectors can agree to this articulation of their findings. However, in order to make the more contentious claim that Cognitive Penetration mechanisms were responsible for these effects one has to further make the case that: 1) the faces were *perceived* (rather than judged) to be relatively lighter or darker based on effects on *early vision* (fulfilling the first criterion for Cognitive Penetration); 2) these effects were the result of the relevant cognitive state serving as an *informational resource* for the perceptual state (fulfilling the second criterion); and 3) that the explanation according to which Cognitive Penetration mechanisms were responsible for these effects is not undercut by *explanatory defeaters* like context factors, postperceptual effects, or demand characteristics (fulfilling the third criterion).³³ Now let’s take a first pass at evaluating whether this study fulfills these criteria – this will also be further evaluated in light of some of the objections below.

First criterion. The clear demonstration provided by the stimuli used by Levin & Banaji (the fact that one can experience the lightness effects for oneself by looking at them) help make a compelling case that the effects they found were indeed *perceptual*. Furthermore, Levin & Banaji specifically address worries about potential postperceptual effects in the construction of their study as a whole and explicitly in their fourth experiment. Because they found evidence of the lightness effect in the fourth experiment, in which subjects were required to make judgments quickly (the timeframe for judgments was much shorter than in previous experiment iterations) and relatively unreflectively (subjects were told to ignore lightness effects in their judgments but their results indicate that lightness effects did impact their judgments), this provides support for the claim that at the very least, the effects found by Levin & Banaji in this fourth experiment are indeed *perceptual* effects. And furthermore, because perceptions of lightness and lightness distortions take place during the *early stages* of visual perception, this makes a good case that Levin & Banaji's results from at least one of their experiments fulfill the first criterion for Cognitive Penetration.³⁴

Second criterion. The effects found by Levin & Banaji also appear to clearly fulfill the second criterion. The likely contents of the penetrating cognitive state ("White faces are lighter and Black faces are darker") quite plausibly serve as *informational resources* for the resulting perceptual experiences in which subjects experience lightness effects on White and Black faces. Furthermore, Levin & Banaji also rule out the most plausible potentially mediating influences in some of their third and fourth experiments (attention mechanisms and postperceptual effects, respectively), thus further strengthening claims that these effects occurred directly and internally. Thus, Levin & Banaji's results, again, at the very least from their third and fourth experiments, fulfill the second criterion for Cognitive Penetration.

Third criterion. Finally, a Cognitive Penetration account of Levin & Banaji's results does not appear to be undercut by any *explanatory defeaters*. Many of the viewing conditions (the state of the sensory organ and the object of perception) are held fixed during all of the experiments, and further viewing conditions (contextual features and attention effects) are held fixed in the latter two experiments. Thus, changes in viewing conditions are unable to contribute to *explanatory defeaters* and the effects of at least Levin & Banaji's latter two experiments appear to fulfill the third criterion for Cognitive Penetration.

Although initially, Levin & Banaji's results appear to easily fulfill the three criteria for Cognitive Penetration, application of the objections below to this study (and other key studies) attempt to call into question. Specifically, the El Greco objection further calls into question fulfillment of the second and third criteria; the judgment Effects objection, the first criterion; and the

Low-level Perceptual Effects objection, the second and third criteria. Now I turn to the objections and will argue that none of them undermine Levin & Banaji's study, nor do they undermine other key Cognitive Penetration studies to which they've been applied.

3.3 *El greco objection*

This objection, originally from Firestone & Scholl and also taken up by Machery, claims that alleged Cognitive Penetration effects fall prey to the “El Greco fallacy”, which then indicates that the effects were not caused by Cognitive Penetration mechanisms.³⁵ According to this objection, when a subject's perception of a “target property” (which is the property being manipulated in an experiment) is *seemingly* Cognitively Penetrated, while her perception of a relevantly similar “measurement property” (the property against which effects on the target property are measured) is not, this provides reason to think that Cognitive Penetration did not cause the effects on the subject's perception of the target property. Firestone, Scholl and Machery claim that this objection undermines many purported instances of Cognitive Penetration, including Delk and Fillenbaum's (1965) seminal experiment: *Differences in Perceived Color as a Function of Characteristic Color*.

In Delk and Fillenbaum's (1965) study, subjects are presented with shapes cut out of orange paper, laid over a background of the same orange paper, and asked to adjust the color of the background paper so that it matches the color of the foreground shape. Delk & Fillenbaum found that subjects over-adjusted the color of the background paper to be redder when the cutout shapes in the foreground were of characteristically red-colored objects (e.g., heart, apple, etc.) but not when the cutout shapes were of characteristically non-red-colored objects (e.g., four-leaf clover) or of objects without a characteristic color (i.e., oval). Delk & Fillenbaum, and many others hold that these effects were caused by Cognitive Penetration: subjects saw the characteristically red-colored shapes as redder than they were as a result of the penetration of beliefs like “hearts are red” and thus needed to adjust the background to be redder in order to match the perceived color of the foreground shapes.

However, Machery invokes the El Greco fallacy, claiming that *if* Cognitive Penetration caused subjects to see the orange paper of the characteristically red-colored-object cutouts (the target property) as redder, it should have also caused them to see the orange paper of the background (the measurement property) as redder, and thus not caused subjects to adjust the color of the background (because the redder color of the foreground cutout shape and the redder color of the background would match). However, since this is not what occurred according to Delk & Fillenbaum's results – subjects *did* adjust the

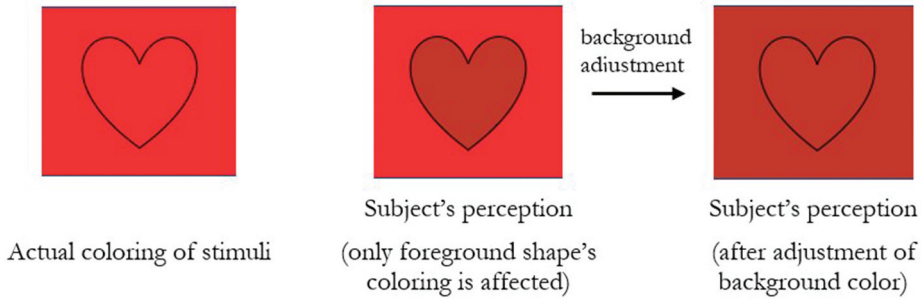


Figure 4. What happened in Delk & Fillenbaum's study.³⁶

color of the background – these effects, Machery concludes, were not due to Cognitive Penetration. In [Figures 4](#) and [5](#) below, we see the difference in the subject's perceptual experience according to the El Greco Objection.

In this application of the El Greco objection, Machery overlooks important ways in which Cognitive Penetration mechanisms operate. He seems to assume that Cognitive Penetration ought to affect larger portions of a subject's visual field, irrespective of content of the penetrating cognitive state. However, at least in many cases, it is quite plausible that Cognitive Penetration effects are localized to particular objects of perception, especially the objects of perception that are likely to “trigger” the influence of the particular penetrating cognitive state. In this study, presumably it is not *the color of the paper* (which is shared by both the foreground cutout shape and the background paper) that triggers the influence of Cognitive Penetration mechanisms, but rather *the shape of the paper* (which is not shared with the background paper).

So, in this particular instance, if the penetrating cognitive state is something like the subject's belief that “hearts are red”, the subject's perception of the color of a heart shape (the foreground cutout) is much more likely to be

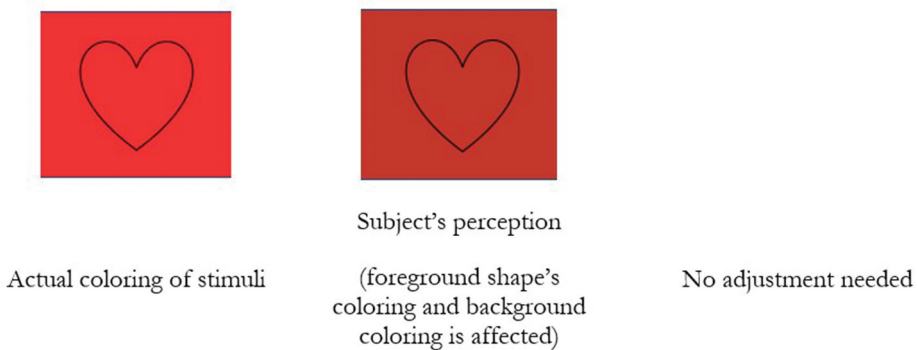


Figure 5. What Machery thinks *should* have happened, if results were due to Cognitive Penetration.

affected by Cognitive Penetration mechanisms than her perception of the color of the rectangular background paper. Thus, contra Machery, if the effects found in the experiment are due to Cognitive Penetration mechanisms, it makes sense that subjects behaved the way that they did (i.e., adjusted the background to be redder because they perceived the foreground heart shape to be redder but not the background), and the Delk & Fillenbaum's results are not undermined by this objection. For this reason, a version of this experiment whose results might more plausibly be due to the El Greco effect would be one in which *both* the foreground and the background paper were characteristically red-colored shapes (e.g., a small cutout apple superimposed over a big cutout heart).

Next, we turn to Levin & Banaji's study. Firestone & Scholl remark that Levin & Banaji's study, or at least certain experiments within it, might be susceptible to this objection. They note that in Levin & Banaji's first experiment (of their 2006 study), they "found lightness-distortion effects even where they "shouldn't" have – not only when observers judged the lightness of the faces by matching a grayscale patch to a reference face, but also when matching *copies of the faces themselves* to a reference face. This pattern of results implies an "El Greco fallacy"³⁷. Here, Firestone & Scholl are referencing Levin & Banaji's first experiment in which subjects were presented with "same-race" and "different-race" pairs of faces in which one face was adjustable. They note that subjects even adjusted Black faces to be darker when asked to match it with the same Black face and adjusted White faces to be lighter when asked to match it with the same White face. Firestone & Scholl note that, if the effects found in experiment 1 were due to Cognitive Penetration, this adjustment shouldn't occur: subject's perceptions of both White face 1 and White face 2 (or Black face 1 and Black face 2) should be penetrated by the belief that "White faces are light" (or "Black faces are dark") and the subjects should see both White faces as being equally lighter (or both Black faces as equally darker) and thus not need to adjust one to match the other.

This application of the El Greco objection is much more compelling than the application addressed above, because in this instance, the target property and the measurement property are identical (either identical White faces or identical Black faces). In their original study, Levin & Banaji note this worry and offer a brief potential explanation of the results of these same-race trials: they note that because subjects were not able to adjust the lightness of the "static" face, it may have seemed to be a more integral property of the face, and thus appeared more salient – thus prompting subjects to adjust the adjustable face to match this level of salience.³⁸ While an interesting proposal, it is not particularly compelling without further explanation.

However, a more compelling response can be offered on Levin & Banaji's behalf by looking at the difference in strength of effect found between the same-race and mixed-race trials: on the mixed-race trials "Black faces were judged to be 4.35 levels darker" while on the same-race trials they were only "judged to be 1.55 levels darker".³⁹ Thus, while this objection might highlight that a portion of the effect found on mixed-race trials might be due to causes other than Cognitive Penetration, it does not undermine the entire effect size, or even most of it. The difference in strength of effect between these trials, 2.8 levels, is still two thirds of the original effect. Recalling the *No Explanatory Defeaters* criterion, as long as external factors do not account for the entire effect, they do not act as explanatory defeaters for the original Cognitive Penetration account of these effects. Thus, it may be that Firestone & Scholl are right that *some* portion of the effects found by Levin & Banaji were not due to Cognitive Penetration mechanisms, but that doesn't address the much larger portion that still may be. Furthermore, note that this objection also doesn't call into question Levin & Banaji's other three experiments, two of which (experiments 3 and 4) as I've mentioned above, provide the strongest empirical support for Cognitive Penetration.

In closing, I agree with Firestone, Scholl and Machery that the El Greco objection is problematic for some purported instances of Cognitive Penetration.⁴⁰ However, it does not undermine the key Cognitive Penetration experiments addressed above, including Levin & Banaji's. Furthermore, similar defenses can be offered on behalf other studies similarly challenged by this objection.⁴¹

3.4 Judgment effects objection

Another objection, posed by Firestone, Scholl, and Machery, is that many experimental effects originally attributed to Cognitive Penetration may in fact be due to subjects' *postperceptual judgments* of their perceptions, rather than effects on the perceptions themselves, thus disqualifying them from counting as instances of Cognitive Penetration. A more specific version of this objection is that these effects may be due to *demand characteristics* which cause an experimental subject to interpret or judge her perception in light of what she believes the purpose of the experiment to be.⁴²

Machery again targets Delk and Fillenbaum's (1965) study with this objection. Recall that subjects were asked to adjust the color of an orange paper background to match the color of orange paper cutouts placed on top of them, and subjects tended to make the background redder when the cutout was the shape of a characteristically red object. Machery proposes that the effects found in this experiment may be due to subjects' *judgments* that the foreground objects were redder, rather than their *perception* that they were. Machery's claim is that subjects may be (accurately) *perceiving*

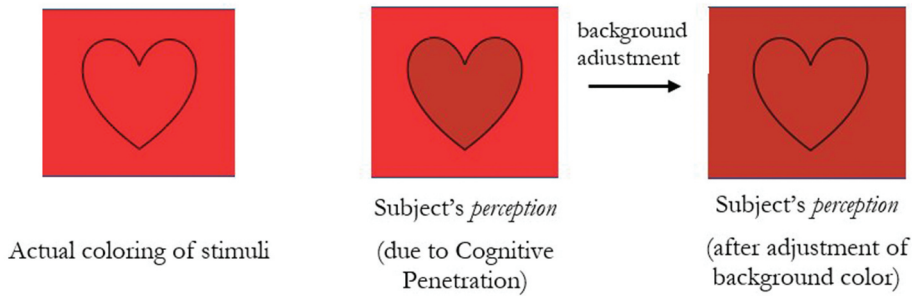


Figure 6. Cognitive Penetration interpretation of Delk & Fillenbaum's study.

both the foreground paper cutout and background paper as orange, and then *judging* the foreground paper object to be redder, and thus adjusting the background to be redder in order to match the foreground cutout on this basis (see [Figures 6 and 7](#) below).

However, Machery's interpretation is unconvincing because it fails to provide a satisfactory account of the "adjustment behavior" of subjects in Delk & Fillenbaum's study. According to Machery's interpretation, the subject initially accurately perceives the color of the foreground cutout and background (as both orange) and yet (according to Delk & Fillenbaum's results) is still motivated to adjust the coloring of the background – it is unclear why a *judgment* would motivate the subject to do this in spite of the subject perceiving the color of the foreground cutout and background as already matching. Furthermore, according to Machery's interpretation, even if the subject was initially motivated to adjust the background, after doing so, she would then accurately perceive the foreground cutout as orange and the background color as redder and yet (according to Delk & Fillenbaum's results) not readjust the background color back toward orange to try and make it match with the foreground cutout – in this case, it is unclear why the subject's *perceptions* would not motivate her to do this since there would be a mismatch in the color of the foreground

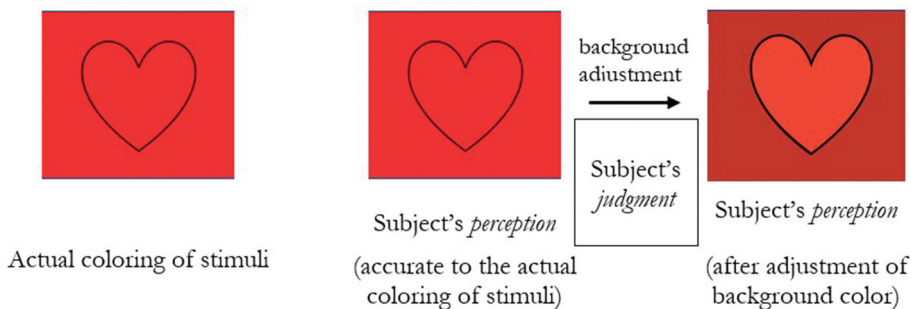


Figure 7. Machery's alternative interpretation of Delk & Fillenbaum's study.

cutout and background. In short, Machery tries to “substitute” the subject’s *judgment* of the coloring of the foreground cutout for her *perception* of the coloring of the cutout as what that drives her to adjust the background, however, this interpretation fails to explain the subject’s behavior as found by Delk & Fillenbaum.

In contrast, the Cognitive Penetration interpretation can provide a satisfactory explanation of this “adjustment behavior”. Because the subject *perceives* the heart to be redder, she then adjusts the background to be redder and then experiences a “perceptual match” in which both the background and heart cutout appear redder. The “thoroughly perceptual” setup of Delk & Fillenbaum’s study enables it to withstand this objection: it is designed to elicit and measure immediate, perceptual adjustments made in response to experiences of perceptual match or mismatch. Note that Machery’s objection would be much more compelling if applied to a version of this study (or another study) which measured subject’s perceptual effects through relying on subject’s self-reports of these effects (e.g., subjects verbally describe or otherwise “rate” or categorize the color of the foreground cutouts after looking at them). In this kind of study, it would be much more plausible for judgment effects to play the kind of role that Machery proposes that they do.

Now, turning to Levin & Banaji’s study, Firestone & Scholl pose the judgment Effects objection against them as well. Specifically they focus on their latter two experiments noting that these results may be due a specific kind of judgment effect: task demands.⁴³ They reason that since the stimuli used in these latter two experiments don’t serve as clear demonstrations – when you look at them there don’t appear to be lightness effects, like in their previous experiments – and subjects were informed of the purpose of the study, it is likely that subject’s lightness ratings were due to *postperceptual judgments* based on their knowledge of the purpose of the study rather than effects on their *perceptions* of the stimuli.

However, we can offer a similar defense of Levin & Banaji’s study because the tasks performed by subjects similarly involved making perceptual adjustments to stimuli on the basis of perceived matches or mismatches. Thus, if *postperceptual judgments* were what caused subjects to adjust the gray square to be lighter when paired with a White face (from experiment 3), they would then experience a perceptual mismatch after adjusting, which is not consistent with Levin & Banaji’s results. However, according to a Cognitive Penetration interpretation, subjects *perceive* the White face as lighter, and thus when they adjust the gray square to be lighter, they experience a perceptual match. Again, a version of this study in which subjects reported on their perceptions through other non-perceptual means, such as rating the lightness of faces numerically, would be much more vulnerable to this kind of

objection. In summary, again, while this objection may undermine some purported Cognitive Penetration studies, it does not undermine some of the key experiments that objectors target, including Levin & Banaji's.⁴⁴

III.5 Low-level perceptual effects objection

Now, I turn to a final challenge which is specifically leveled by Firestone and Scholl (2014, 2015) against Levin & Banaji's study. They argue that the effects found by Levin & Banaji may have been due to low-level perceptual effects, rather than Cognitive Penetration mechanisms. To do so, they carried out an iteration of Levin & Banaji's original experiment in which they presented participants with blurred versions of Levin & Banaji's original face images and asked subjects to select which of the images was darker; they also tested whether subjects seemed to be aware of any race-identifying differences between the faces. Firestone & Scholl found that of those who judged that the races of the two faces looked the same (and fulfilled various other criteria), the majority (72%) "reported that the blurry Black image was darker or that the blurry White image was lighter".⁴⁵ They then reason that, because many of their subjects reported a difference in lightness while apparently unaware of racial difference between the faces, the lightness effect shown by Levin & Banaji could have very well been due to low-level perceptual differences between the images, rather than the penetration of subjects' race-related beliefs through Cognitive Penetration mechanisms. They conclude by noting that

"... although the original effect (with unblurred faces) could of course still be explained entirely by race (rather than by the lower-level differences now shown to affect perceived lightness), it is clear that further experiments would be needed to show this - and so we conclude that the initial demonstration of Levin and Banaji (2006) provides *no evidence* for a top-down effect on perception."⁴⁶ (emphasis mine)

Although Firestone & Scholl bring up compelling points, there are a couple of crucial weaknesses in their study that undermine the force of this objection. First off, before addressing these weaknesses, it's important to note that results from Levin & Banaji's third experiment are not susceptible to this objection because they used line drawings (rather than life-like images) of faces, explicitly to address worries about low-level perceptual effects, and they still found lightness effects. Now, turning to weaknesses in their study – there are problems with the sensitivity of their "race-identifying" measures as well as their "lightness-effect" measures. First, on their "race-identifying" measures, Baker and Levin (2014) address this by attempting to replicate some of Firestone & Scholl's findings.⁴⁷ First, they replicate Firestone & Scholl's experiment using their blurred stimuli and "race-identifying" measures. However, Baker & Levin then perform a second experiment in which they use a more fine-grained measure to determine whether subjects were

able to detect racial information. They then found that, even when using the blurry stimuli of Firestone & Scholl, most participants (80.2%) were able to detect racial information according to this more sensitive measure. On this more sensitive measure – Firestone & Scholl used “same-different” measures – meaning that subjects only needed to report that two faces were of the same or different races; whereas Baker & Levin used “forced-choice” measures – meaning that subjects had to select which face out of a pair they thought was White and which was Black. Furthermore, Baker & Levin demonstrated that the participants who were not able to detect racial information (according to their measures) did not experience (a significant) lightness effect while those that were able to detect racial information did.⁴⁸ Thus, contra Firestone & Scholl, their results suggest that 1) the majority of subjects were able to detect racial information and 2) detection of that racial information was correlated with experiencing significant lightness effects.

In addition to Baker & Levin’s compelling critique of their “race-identifying” measures, I also argue that Firestone & Scholl’s “lightness-effect” measures were insufficiently sensitive. Although Firestone & Scholl use a forced-choice measure to measure lightness differences – meaning that subjects had to select which face was lighter and which was darker – these measures did not address the *strength* of these lightness effects. This is important because Levin & Banaji, in their original study, did: they measured not just whether subjects found faces lighter or darker but how much lighter or darker, and found that on average “subjects chose a Black reference that was 2.9 gray levels darker than the corresponding White face”.⁴⁹ This undermines Firestone & Scholl’s comparative claims about the relevance of their findings in relation to that of Levin & Banaji. Without the ability to compare the strength of the lightness effects that they found, this leaves open the possibility that the lightness effects found by Firestone & Scholl were significantly smaller than those found by Levin & Banaji. And in fact, because Firestone & Scholl used much less lifelike stimuli (very blurry images) in comparison to that used by Levin & Banaji (realistic images of faces), we would expect based on results from Levin & Banaji’s original study as well as those from studies like Hansen et al.’s that the lightness effects Firestone & Scholl found *would* be significantly smaller. This then means that it is possible, even quite plausible that the effects found by Levin & Banaji were due to both low-level perceptual effects (those highlighted by Firestone & Scholl’s results) *and* Cognitive Penetration mechanisms.⁵⁰ And because Firestone & Scholl did not measure the strength of the lightness effect, their data cannot be used to respond to this proposal.

A more compelling version of Firestone & Scholl’s study would be one in which they 1) used forced-choice (or other sufficiently sensitive) “race-identifying” measures, 2) measured the strength of lightness effects, and 3) used stimuli that were somewhat more life-like/less blurry (to potentially

elicit stronger effects) while still remaining blurry enough that obvious racial features were not apparent. However, taking these criticisms together, it is clear that Firestone & Scholl's Low-level Perceptual Effects Objection fails to undermine Levin & Banaji's study. Furthermore, Baker & Levin's replication attempts help provide further support to their claims that race-identifying features of (both clear and blurry versions of) these stimuli seem to drive lightness effects, presumably through the operation of Cognitive Penetration mechanisms.

4. Conclusion

I have argued that Levin & Banaji's experiment, as well as a number of the other experiments addressed above continue to provide compelling support for Cognitive Penetration, despite some of the objections leveled against them by Firestone, Scholl, and Machery. This is in stark contrast to claims by Firestone & Scholl that "none of these hundreds of studies" provide evidence for Cognitive Penetration and that "there is in fact *no* evidence" for Cognitive Penetration.⁵¹ In addition to their ability to withstand the critiques addressed above, both Levin & Banaji's original study and Baker & Levin's follow-up study discussed above are particularly notable and valuable examples of Cognitive Penetration for a couple of reasons. First, the cognitive states involved clearly result from higher-order conceptual processing – in order for the effects to occur, subjects must first categorize the faces according to race, based on generalizations about features of faces typically belonging to these categories. Second, many of the stimuli used serve as compelling demonstrations for everyday viewers in non-experimental settings. And finally, they serve as important examples of Cognitive Penetration effects that may be associated with, and in some cases contribute to, more substantive, socially-relevant value judgments.

Levin & Banaji note this possibility and try to briefly address it by measuring explicit attitudes toward Black and White people in their first three experiments in order "to test whether there was any association between favorability toward the groups and basic color perception".⁵² They note one hypothesis, that perhaps "those who are more negatively predisposed toward African Americans and willing to express it explicitly will show a stronger dark bias than those who are not" but ultimately report finding "no hint of a correlation between attitudes toward the races and the degree of lightness distortion."⁵³ While this might be a desirable result, it is again important here to pay attention to the sensitivity of the measures used. Levin & Banaji only asked subjects about *general favorability* – specifically, they asked subjects to complete a "feeling thermometer" in which they rated their attitudes toward the races from 0 (very coolly) to 100 (very warmly). So, we should be very hesitant to draw conclusions from these results about

a lack of potential associations between these Cognitive Penetration effects and value judgments related to race. Judgments of general favorability, especially highly explicit ones, are importantly distinct from unconscious attitudes as well as specific judgments about particular traits and features. There is a robust research program which indicates that perceptions of the lightness of skin color are associated with judgments about (among other things): psychological qualities (e.g., aggression, intelligence, enthusiasm), perceived level of income, educational attainment, and parenting skills.⁵⁴ As we can see from just this brief sketch, there are many possible connections between the kinds of perceptual effects found by Levin & Banaji certain kinds of social perception.

In addressing the objections against empirical support for Cognitive Penetration above, I have also clarified important aspects of its mechanisms and have highlighted the misunderstandings and mischaracterizations that seem to motivate some of the particular applications of these objections. It is important to keep these nuances in mind when evaluating particular studies of Cognitive Penetration and the current state of empirical support for Cognitive Penetration, as well as when designing novel studies and replications.

Notes

1. Thanks to Michael Rauschenbach, Robert Rupert, Ted Warfield, and two anonymous reviewers for helpful feedback on versions of this paper. Also, thanks to participants in sessions at the American Philosophical Association, Pacific and Eastern Division conferences at which I presented previous versions of this paper.
2. Some also hold that our perceptions may be penetrated by a wider set of cognitive states. For example, see, Cowan (2015).
3. Balcetis and Dunning (2010), Hansen et al. (2006).
4. Note that the subject need not be, and often is not, aware of these cognitive states or their effects at the time of the perceptual experience.
5. This relatively recent interest in Cognitive Penetration builds upon the previous New Look movement in psychology. New Look psychologists held that one's cognitive states often (according to some, always) impacted one's perceptions because cognition was continuous with perception. Cognitive Penetration theorists often draw on theoretical and empirical resources of this movement. See, Bruner (1957). Thanks to an anonymous reviewer for drawing my attention to the importance of this historical lineage.
6. See, Clarke (2020)
7. Siegel (2012).
8. Note that some, like Shevlin & Friesen, take instances which fulfill the criteria mentioned below to be instances of specifically *Radical* Cognitive Penetration rather than merely Cognitive Penetration more generally. Shevlin and Friesen (2020).
9. See, Macpherson (2012), Wu (2013), and Pylyshyn (1999).
10. For more on these disagreements, see, Lyons (2020), Raftopoulos (2019).

11. Pylyshyn (1999). Also see, Pylyshyn (2007), Marr (1982), and Raftopoulos (2009). Also see, Macpherson (2012).
12. Macpherson (2015).
13. Wu (2013), p. 657. Also see, Pylyshyn (1999) and Gross (2017) on the “semantic coherence” criterion according to which the content of the cognitive state and the perceptual state must be “semantically coherent”: the content of the former must serve as an epistemic basis for the latter.
14. Wu (2013), Macpherson (2012), and Stokes (2013).
15. Firestone and Scholl (2016).
16. I address this experiment in more detail below.
17. Machery (2015).
18. Firestone and Scholl (2015).
19. Firestone and Scholl (2016), p. 31.
20. Although Firestone & Scholl do not explicitly cite this as an *objection* against Cognitive Penetration but rather a worrisome fact about it, it has strong intuitive appeal and is often mentioned in discussions of Cognitive Penetration, so is worth addressing. Firestone and Scholl (2016), p. 31.
21. See, Hansen et al. (2006), Levin and Banaji (2006), Delk and Fillenbaum (1965), and Olkkonen et al. (2008).
22. Many thanks to an anonymous reviewer for proposing this case as a potential demonstration failure.
23. See, Macpherson (2012), Collins and Olson (2014), and Vetter and Newen (2014). Even those who object to Cognitive Penetration such as Firestone & Scholl acknowledge that, if true, it presents the most compelling empirical support.
24. Levin and Banaji (2006).
25. Ibid., p. 503.
26. Ibid, p. 505.
27. Ibid, p. 506.
28. Many mainstream accounts rule out effects mediated by attention mechanisms as candidate instances of Cognitive Penetration, however, there is disagreement on this point. Macpherson (2012), Siegel (2012). I, and some others, maintain that certain kinds of attention, namely exogenous or automatic attention mechanisms are legitimate mediators of Cognitive Penetration effects.
29. Levin and Banaji (2006), p. 507–508.
30. Ibid, p. 509.
31. Ibid, p. 501.
32. They explain that “subjects are slower to classify face pairs as *different* when they match in apparent brightness versus when they match in actual luminance. Ibid, p. 510.
33. Note that if the effects were due to attention mechanisms or postperceptual effects they would also fail to fulfill some of the other criteria.
34. Ibid, p. 502.
35. Firestone (2013).
36. The actual color differences are exaggerated in these images for clarity.
37. Firestone and Scholl (2014).
38. Levin and Banaji (2006) p, 505.
39. Ibid., p. 504.
40. For example, I agree with them that it undermines Goldstone (1994).
41. The impact of demand characteristics is well-documented, and some argue that this might be behind many of the purported instances of Cognitive Penetration. For example, Banerjee et al. (2012). which was also targeted by Firestone & Scholl.

42. See, Nichols and Maner (2008).and Machery (2015).
43. Firestone and Scholl (2016), fn. 3.
44. For example, I think it does undermine Caruso et al. (2009).
45. Firestone and Scholl (2015).
46. Ibid.
47. See, Baker and Levin (2014), Levin et al. (2016).
48. Baker and Levin (2014).
49. Levin and Banaji (2006).
50. However, note that Baker & Levin make a stronger claim arguing not just that some portion of the lightness effects may likely be due to perceived race information but that “*any* effects that might be observed can reasonably be explained by subtle signs of race that survived the blurring process.”
51. Firestone and Scholl (2016), p. 2, 6.
52. Levin and Banaji (2006), p. 203.
53. Ibid, p. 503–504.
54. On aggression see, Dasgupta et al. (1999); on enthusiasm, and parenting skills see, Wade and Bielitz (2005); On perceived level of income see, Wade (1996); and on educational attainment see, Keith and Herring (1991).

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Kate Finley  <http://orcid.org/0000-0002-8293-2391>

References

- Baker, L., & Levin, D. (2014). The face-race lightness illusion is not driven by low-level stimulus properties: An empirical reply to Firestone & Scholl, *Psychonomic Bulletin & Review*, 23(6), 1989–1995. <https://doi.org/10.3758/s13423-016-1048-z>
- Balcetis, E., & Dunning, D. (2010). Wishful seeing: Motivational influences on visual perception of the physical environment. Balcetis, E, Lassiter, G.D. *The Social Psychology of Visual Perception*, New York: Psychology Press, 77–101. <https://doi.org/10.1177/0963721412463693>
- Banerjee, P., Chatterjee, P., & Sinha, J. (2012). Is it light or dark? Recalling moral behavior changes perception of brightness. *Psychological Science*, 23(4), 407–409. <https://doi.org/10.1177/0956797611432497>
- Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64(2), 123. <https://doi.org/10.1037/h0043805>
- Caruso, E., Mead, N., & Balcetis, E. (2009) Political partisanship influences perception of biracial candidates’ skin tone. *Proceedings of the National Academy of Sciences*, 106.48, 20168–20173.
- Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science*, 55(2), 167–187. <https://doi.org/10.1086/289425>

- Clarke, S. (2020). Cognitive penetration and informational encapsulation: Have we been failing the module? *Philosophical Studies* 178, 1–22. <https://doi.org/10.1007/s11098-020-01565-1>
- Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, 21(4), 843–860. <https://doi.org/10.3758/s13423-013-0564-3>
- Cowan, R. (2015). Cognitive penetrability and ethical perception. *Review of Philosophy and Psychology*, 6(4), 665–682. <https://doi.org/10.1007/s13164-014-0185-4>
- Dasgupta, N., Banaji, M., & Abelson, R. (1999). Group entitativity and group perception: Associations between physical features and psychological judgment. *Journal of Personality and Social Psychology*, 77(5), 991. <https://doi.org/10.1037/0022-3514.77.5.991>
- Delk, J., & Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *The American Journal of Psychology*, 78(2), 290–293. <https://doi.org/10.2307/1420503>
- Firestone, C. (2013). On the origin and status of the “El Greco fallacy”. *Perception*, 42(6), 672–674. <https://doi.org/10.1068/p7488>
- Firestone, C., & Scholl, B. (2014). ‘Top-down’ effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, 25(1), 38–46. <https://doi.org/10.1177/0956797613485092>
- Firestone, C., & Scholl, B. (2015). Can you experience ‘top-down’ effects on perception?: The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*, 22(3), 694–700. <https://doi.org/10.3758/s13423-014-0711-5>
- Firestone, C., & Scholl, B. (2016). Cognition does not affect perception: Evaluating the evidence for ‘top-down’ effects. *Behavioral and Brain Sciences*, 39, 1–72. <https://doi.org/10.1017/S0140525X15000965>
- Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, 51(1), 23–43. <https://doi.org/10.1086/289162>
- Fodor, J. A. (1988). A reply to Churchland’s “Perceptual plasticity and theoretical neutrality”. *Philosophy of Science*, 55(2), 188–198. <https://doi.org/10.1086/289426>
- Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2), 125–157. [https://doi.org/10.1016/0010-0277\(94\)90065-5](https://doi.org/10.1016/0010-0277(94)90065-5)
- Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology*, 8, 221. <https://doi.org/10.3389/fpsyg.2017.00221>
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. (2006). Memory modulates color appearance. *Nature Neuroscience*, 9, 1367–1368. <https://doi.org/10.1038/nn1794>
- Keith, V., & Herring, C. (1991). Skin tone and stratification in the Black community. *American Journal of Sociology*, 97(3), 760–778. <https://doi.org/10.1086/229819>
- Levin, D., & Banaji, M. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology*, 135(4), 501. <https://doi.org/10.1037/0096-3445.135.4.501>
- Levin, D., Baker, L., & Banaji, M. (2016). Cognition can affect perception: Restating the evidence of a top-down effect. *Behavioral and Brain Sciences*, 39, 39–40. <https://doi.org/10.1017/S0140525X15002642>
- Lyons, J. (2020). *The cognitive impenetrability of early vision: What’s the claim?* Rivista internazionale di. *Filosofia E Psicologia*, 11(3), 372–384. <https://doi.org/10.4453/rifp.2020.0026>
- Machery, E. (2015). A no progress report on cognitive penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception: New philosophical perspectives* (pp. 71). Oxford University Press.

- 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. <https://doi.org/10.1111/j.1933-1592.2010.00481.x>
- Macpherson, F. (2015). Cognitive penetration and nonconceptual content. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception: New philosophical perspectives*. Oxford University Press, 441.
- Marr, D. (1982). *Vision: A computational investigation into human representation and processing of visual information*. Freeman.
- Nichols, A., & Maner, J. The good-subject effect: Investigating participant demand characteristics. (2008). *The Journal of General Psychology*, 135(2), 151–166. Machery (2015). <https://doi.org/10.3200/GENP.135.2.151-166>
- Olkkonen, M., Hansen, T., & Gegenfurtner, K. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, 8(5), 13. <https://doi.org/10.1167/8.5.13>
- Pylyshyn, Z. (1999). Is Vision Continuous with Cognition?. *Behavioral and Brain Sciences*, 22 3 , 341–365. <https://doi.org/10.1017/S0140525X99002022>
- Pylyshyn, Z. (2007). *Things and Places: how the mind connects with the world*. MIT Press.
- Raftopoulos, A. (2009). *Cognition and Perception*. MIT Press.
- Raftopoulos, A. (2019). Early and late vision: Their processes and epistemic status. In *Cognitive Penetrability and the Epistemic Role of Perception* (pp. 251–338). Palgrave Macmillan.
- Shevlin, H., & Friesen, P. (2020). *Pain, placebo, and cognitive penetration*. Mind & Language.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Nous*, 46(2), 201–222. <https://doi.org/10.1111/j.1468-0068.2010.00786.x>
- Stokes, D. (2013). Cognitive penetrability of perception. *Philosophy Compass*, 8(7), 646–663. <https://doi.org/10.1111/phc3.12043>
- Stokes, D. (2015). Towards a consequentialist understanding of cognitive penetration. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives*. Oxford University Press, 75–100.
- Stokes, D. (2020). Cognitive penetration and the perception of color. In D. Brown & F. Macpherson (Eds.), *Routledge Handbook of the Philosophy of Colour*, Routledge, 475.
- Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, 27, 62–75. <https://doi.org/10.1016/j.concog.2014.04.007>
- Wade, T. (1996). The relationships between skin color and self-perceived global, physical, and sexual attractiveness, and self-esteem for African Americans. *Journal of Black Psychology*, 22(3), 358–373. <https://doi.org/10.1177/00957984960223006>
- Wade, T., & Bielitz, S. (2005). The differential effect of skin color on attractiveness, personality evaluations, and perceived life success of African Americans. *Journal of Black Psychology*, 31(3), 215–236. <https://doi.org/10.1177/0095798405278341>
- Wu, W. (2013). Visual spatial constancy and modularity: Does intention penetrate vision?. *Philosophical Studies*, 165(2), 647–669. <https://doi.org/10.1007/s11098-012-9971-y>
- Wu, W. (2017). Shaking up the mind's ground floor: The cognitive penetration of visual attention. *The Journal of Philosophy*, 114(1), 5–32. <https://doi.org/10.5840/jphil201711411>