Synthese
https://doi.org/10.1007/s11229-020-02978-5

Is color experience linguistically penetrable?

Raquel Krempel[1]

Received: 14 March 2020 / Accepted: 1 December 2020

Abstract

I address the question of whether differences in color terminology cause differences in color experience in speakers of different languages. If linguistic representations directly affect color experience, then this is a case of what I call the linguistic penetrability of perception, which is a particular case of cognitive penetrability. I start with some general considerations about cognitive penetration and its alleged occurrence in the memory color effect. I then apply similar considerations to the interpretation of empirical studies of color perception in speakers of different languages. I argue that findings such as differences in categorical perception in speakers of different languages do not show that language affects color experience. They therefore do not support the claim that color experience is linguistically penetrable. But even if we grant that color experience is different in speakers of different languages, I argue that this might still not be a case of linguistic penetration. Finally, I consider some epistemological consequences of the assumption that speakers of different languages have different color experiences.

Keywords Color perception · Language · Cognitive penetrability · Color concepts · Categorical perception

It has long been known that languages don't all divide the color spectrum in the same way. Languages can have as few as two basic color terms, referring to light and dark colors (Berlin and Kay 1969). Several languages do not distinguish between blue and green (e.g. Himba and Berinmo, spoken by indigenous people in Namibia and in Papua New Guinea, respectively). Others, such as Russian and Greek, unlike

I would like to thank Evan Keeling, Plínio Smith, the members of the groups GEMF and CLEA (especially André Abath, Marco Aurélio Alves, Eduarda Calado, Felipe Carvalho, Nara Figueiredo, Caroline Marim, Marcos Silva and Beatriz Sorrentino), audiences at an SBFA Conference and at UNICAMP, and two anonymous reviewers for helpful comments and suggestions. This research was supported by Grant #2018/12683-9, São Paulo Research Foundation (FAPESP).

✉ Raquel Krempel
raquelak@gmail.com

1 Department of Philosophy, Federal University of São Paulo, São Paulo, Brazil

Published online: 07 January 2021

English, have two different words for light and dark blue. Two natural questions come to mind: what causes these differences in color terminology? And are any differences in color perception caused by differences in color terminology?

It is now believed that, despite differences in color vocabulary, speakers of different languages make equivalent judgements about whether two colors are identical or different (Roberson et al. 2009); they can make the same color discriminations (according to Pointer and Attridge 1998, humans can discern about 2 million colors). So it doesn't seem to be the case that languages lack different words for blue and green, say, because their speakers cannot see the difference between these colors. The explanation of what causes a language to have the color terms that it has probably involves at least cultural (what colors are important) and environmental (what colors are available) factors, which are constrained by facts about color vision.

The question that will occupy me is not what causes differences in color terms, but instead whether color terms can cause differences in color perception. Language induces discrimination habits, and speakers of languages with noncoextensive color terms[1] (terms which differ in the range of colors in their reference) will learn and practice different color discriminations. Does this make any difference in the color experiences speakers of different languages end up having? More specifically, will two colors look more different if they receive two different names than if they don't?

I propose to consider this issue in light of the debate about the cognitive penetrability of perception. This will allow us to consider different ways language and perception might interact, while singling out the kind of influence of language on color perception that would be the most interesting. If language directly affects color experience, that is, the phenomenology of color, or the way colors appear to us, then this is a case of what I'll call the linguistic penetrability of perception, which I take to be a particular case of cognitive penetrability of perception. Some recent studies seem to show that the language we speak directly affects color experience (Roberson et al. 2000, 2005; Zhong et al. 2018), resulting in speakers of different languages having different color experiences. If that is the case, then color experience is linguistically penetrable.

Before discussing the evidence for this, I believe it's helpful to consider some reasons for taking color experience to be cognitively (or linguistically) impenetrable (Sect. 1). It is also helpful to illustrate how some have denied this with the so-called memory color effect (Sect. 2). This will serve to illuminate the discussion about the linguistic penetrability of color experience. Both phenomena, if real, seem to imply that color experience, contrary to what most vision scientists have assumed, is affected by higher order cognitive states. In Sect. 3 I'll review some empirical studies involving speakers of different languages, and discuss whether we should interpret their findings as evidence for the linguistic penetrability of color perception. I'll argue that there is so far no strong evidence that speakers of different languages have different color experiences. But even if we grant that they do, I'll argue in Sect. 4 that this might still not be a case of linguistic penetration. In Sect. 5, I consider some epistemological consequences of the assumption that speakers of different languages

[1] Henceforth "speakers of different languages".

have different color experiences. I'll argue that if this were the case, it would have negative epistemological consequences, but only small ones.

1 Cognitive penetrability and colors

To a first approximation, we can say that perception is cognitively penetrated if perceptual experiences have their phenomenal character at least in part determined by cognitive states, such as beliefs, desires, expectations or (important to our case) the language one speaks.[2] In further characterizing cognitive penetration, we can consider some useful distinctions made by Fodor (1985) and Pylyshyn (1999). For them, cognitive penetration of visual perception would occur if information from *outside* the visual system affected some computation that the system performs, and therefore its output.[3,4] As Gross et al. put it, "not all top-down effects are cognitive penetration: it matters *where* the information is stored" (2014, p. 6). Contextual and top-down information processing that occurred within the visual system because of visual principles would not count as cognitive penetration (Pylyshyn 1999). In addition, the top-down cognitive effects must be *perceptual* (which I take to mean that, in the case of color, they would have to affect color experience, or phenomenology), and not just post-perceptual (affecting e.g. judgments about color appearance) (Fodor 1985, p. 204).[5] Also, it is usually accepted that perceptual differences should not be due to *attention* (Fodor 1988; Pylyshyn 1999; Siegel 2012; Firestone and Scholl 2016). This means that if two subjects have different perceptual experiences only because they are attending to different things, that shouldn't count as cognitive

[2] Here I follow Macpherson (2012) and Siegel (2012) in taking that alteration in perceptual experience, and not only in perceptual processing, is necessary for the cognitive penetrability of perception.

[3] This is roughly what Burnston (2017) calls "the computation condition" for the cognitive penetrability of perception, and he goes on to argue that it cannot be met, due to differences in format between perceptual and cognitive representations. That is one way to argue against one form of cognitive penetrability of perception. I here remain neutral about the format perceptual and conceptual representations take.

[4] To be more precise, Fodor's view seems to be a little more strict than Pylyshyn's, for, as I read him, cognitive penetration would occur if any kind of information from outside a given module were to affect its output. For Pylyshyn, on the other hand, the penetrating information would have to come from outside the visual system. Presumably, then, modules in the visual system could interact.

[5] Both Fodor and Pylyshyn tend to formulate the view that visual input systems, or early visual processes (such as the processing of color), are cognitively impenetrable in terms of their computations and outputs being unaffected by cognitive states. Unlike Siegel (2012) and Macpherson (2012), they tend not to be explicit about whether this should also mean that visual perceptual *experience* of those basic features is impenetrable (though I take they would accept that). Psychologists testing possible effects of memory, or language, on color perception are generally interested in revealing changes in conscious perceptual experience (and they tend to rely on that in their experiments), and not just in sub-personal color processing. In addition, it is cognitive penetrability in this sense that has interesting epistemological implications, which will be considered in the final part of the paper. So differences in color processing that do not generate differences in color experiences should not count as cognitive (or linguistic) penetrability for the purposes of this paper.

penetrability. That is because attention is assumed to change the input to perception, but not perceptual processing properly.[6,7]

As Fodor notes, "no one in his right mind doubts that perception interacts with cognition *somewhere*. What's at issue (…) is the *locus* of this interaction." (1985, p. 204). Though the influence of cognitive states on perceptual judgments, or on where and what we attend to, can be interesting and non-obvious, there is nothing really revolutionary about it; nothing that would force us to reconceive how the mind is organized (Firestone and Scholl 2016). Cognition directly affecting perceptual experience, on the other hand, might force us to reconceive the separation of these two psychological categories. "Cognitive penetration" is therefore reserved for the kind of interaction between cognition and perception that would have particularly revolutionary consequences for our understanding of the mind, as well as for certain epistemological views. The empirical work that will be considered here, as we will see, is often interpreted as providing evidence for this strong kind of interaction, where cognitive or linguistic representations directly affect color experience.

Philosophers care about the relation between perception and cognition for several reasons. Interest in cognitive architecture is one, which involves questions about how the mind is structured, how different faculties relate, whether the mind consists at least in part of modules, and so on. In addition, one might care about the epistemological consequences of different views about the relation between perception and cognition. Those who, with Fodor, like objectivity and "hate relativism" (1985), might wish to say that there are at least some aspects of perceptual experience that are not affected by anything we believe. For if what we believe affects all that we perceive, then perception can't be the common ground we use to gain knowledge and dissipate disagreement. If what people perceive will vary as much as what they believe, then perhaps no two people will ever share the same perceptual experiences. Besides, if perceptual experience is cognitively penetrated, that is, if phenomenal character is partly determined by what we believe, expect, etc., this could create problems for the natural view that perceptual experience provides justification for perceptual beliefs (Siegel 2012). In order to preserve these ideas, we might try to find at least some impenetrable aspects of experience.

Here is where the interest in color experience comes from. If we are looking for cognitively impenetrable features of experience, unaffected by cognitive states, basic visual properties, such as color, are a natural choice. How colors appear to us

6 The topic of how attention relates to cognitive penetration is controversial. In Sect. 4 I'll suggest why differences in attention can't explain away an assumed difference in color experience in speakers of different languages. But I leave it open whether attentional effects on perception could amount to cognitive penetration in other situations. See Raftopoulos (2009) for a development of Pylyshyn's view, with a discussion of different forms of attention and their role in early and late vision, and for why attention renders the latter cognitively penetrable. See Mole (2015), for a criticism of the view that attention-mediated influences of cognition on perception do not amount to cognitive penetration. For discussion, see Gross (2017).

7 More generally, Firestone and Scholl (2016) take it that cognitive penetration requires that higher order states affect perceptual processing, and not just the input to perception. So when our desire to experience darkness causes us to close our eyes, which leads us to experience darkeness, cognition is affecting the input to perception, but not perceptual processing. This is therefore not a case of cognitive penetration.

doesn't seem to depend on any cognitive states.[8] There seems to be nothing I need to know to see bananas as yellow. I just need to look at a ripe banana under the right conditions of illumination and have my visual system working properly: that is enough to explain the yellowness I experience. But one might be inclined to argue that that doesn't hold for the perception of objects. What I experience when I look at an elm tree, or at a cell under a microscope, or at the night sky through a telescope is affected by my background knowledge about trees, organelles and heavenly bodies. I can't perceive Jupiter as Jupiter through the telescope if I don't have the concept JUPITER and if I don't know what it is supposed to look like. But I presumably see the same colors and shapes as an astronomer does.[9]

The expertise required for perceiving Jupiter as Jupiter does not seem required for perceiving yellow. When people disagree about colors, it's often assumed that the explanation for this will fall under the scope of biology. Some people have only two types of cones in the retina, instead of three, and their color experience will be affected in different ways depending on which type is missing. Differences in color perception can also be found among trichromats and are explained by a number of factors, including differences in cone ratios (Webster 2015). These differences are not uncommon, and they tend to manifest themselves in ordinary life. But we don't expect them to be explained by what someone knows and the other doesn't. No amount of teaching could make a colorblind person perceive the difference between green and red.

Color perception is then a good candidate for a cognitively impenetrable aspect of experience. In fact, proponents of the view that perceptual systems are modular, such as Fodor (1983) and Pylyshyn (1999), have taken color processing to be a paradigmatic example of a visual process that is "encapsulated" and therefore not subject to the influence of higher cognitive states.

Given the previous characterizations, if one were to try to preserve the claim that color experience is cognitively impenetrable, in spite of apparent evidence against it, one might typically try to adopt one of the following strategies (cf. Macpherson 2012, p. 25). One is to say that the alleged effect is post-perceptual: color experience, despite appearances, is not really affected by cognitive states; only judgements about experience are affected. Since cognitive penetration of color perception requires alteration in color experience, denying that such alteration occurs would be tantamount to claiming that there is no cognitive penetration. Another strategy for denying cognitive penetrability would be to concede that color experience is affected, but deny that the causes are cognitive states. Differences in color experience might be

[8] Color is a paradigmatic example, but the same seems to be true of other basic visual properties, such as orientation, motion, shape and contrast.

[9] I'm not endorsing the view that one can perceive higher order properties such as Jupiter, and not merely judge a visual object to be Jupiter. I am just presenting it as a possible view, one which grants that perception of objects, but possibly not of their features, is cognitively penetrable (for discussion, see Siegel (2005), Mandelbaum (2018)).

caused by differences in attention, or by perceptual processes or constraints (which use information stored in the visual system).[10]

2 Memory color effect

Despite its intuitive appeal, the claim that color experience is cognitively impenetrable has been challenged. Recently there has been a renewed interest in experiments such as Delk and Fillenbaum's (1965), which seem to show that color experience can be affected by our memory or knowledge of the typical colors of objects. Because of this, many have chosen color experience as the standard example in discussions of cognitive penetrability. In their experiment, Delk and Fillenbaum placed different figures, all the same orange-red color, against a colored background. Subjects had to adjust the color of the background (by giving instructions to an experimenter) until it matched the color of the figure. Participants made the background redder when the figure was of a typical red object (such as a heart or an apple), than when it was of objects with no typical color (such as a circle). Similar results have been found more recently, with better experimental designs (Hansen et al. 2006; Olkkonen et al. 2008; Witzel 2016). In one kind of experiment (Hansen et al. 2006; Olkkonen et al. 2008), subjects had to adjust the color of fruit images until they looked grey. They found that instead of adjusting the color of, e.g., a yellow banana to an achromatic grey, subjects actually chose a grey that was slightly bluish (yellow's opponent color). The same pattern of color adjusting did not occur with a disc (an object with no typical color).

Results like these are usually interpreted as showing that one's knowledge or memory of the typical color of objects affects color experience, what has come to be known as the memory color effect. The idea is that an achromatic grey banana would still look yellow and that for the banana to really look grey, subjects had to adjust its color further away from yellow, making it slightly blue. In the case of Delk and Fillenbaum's experiment, the idea is that subjects make the background more red in the case of a heart because it is perceived to be more red that the circle. Macpherson (2012), for instance, argues that the strategies mentioned above to preserve cognitive impenetrability would not work here (see also Schirillo 1999). She claims, about Delk and Fillenbaum's findings, that if we assumed subjects accurately experienced the color of the heart and of the redder background to be different, but merely judged them to be the same, we would be attributing an inexplicable, "gross form of misjudgment to the subject" (2012, p. 41). She also argues against the idea that attention can explain the effect. Hearts look redder than circles of the same color, and achromatic grey bananas look yellow, because of what we know about their typical colors. Color experience is therefore said to be cognitively penetrable.

[10] For these and different ways to counter the claim that perception is cognitively penetrated, see Firestone and Scholl (2016), who consider a number of recent studies claiming to have found cognitive effects on perception.

But these findings have alternative explanations. Some flaws in Delk and Fillenbaum's experimental design (e.g. the use of a sheet of waxed paper between subjects and the figures) led Zeimbekis (2013) to suggest that, because this was a perceptually difficult task, the belief that hearts are red might have led subjects to initially shift the color of the background towards red, as a strategy to try to get the right matching in non-ideal conditions.[11] Because no typical colors are associated with circles, no such strategy was adopted in that case. Due to the conditions of the experiment, subjects might have reached a point of indiscriminability when trying to match the color of the figure and the background. Hearts, unlike circles, ended up being matched to a redder background, but not because they really looked redder.[12] The newer experiments avoid many of the design flaws affecting Delk and Fillenbaum's, but they also allow for alternative explanations. It is possible that subjects adjusted the color of the banana towards its opponent color not because the achromatic grey banana looked yellow, but as a strategy in order to make sure that there was nothing of the canonical color left in the image (Zeimbekis 2013; Valenti and Firestone, 2019). Because subjects know bananas are yellow, they moved a safe distance away from yellow, making the banana a grey that, while slightly blue, could still count as grey, and so allowed them to comply with the instructions given.

Witzel (2016)'s experiment seems to avoid this objection, because it doesn't rely on color adjustment. The image of a grey banana and of a bluish-grey banana were shown side-by-side against a grey background, and subjects were asked which looked grey. The bluish-grey banana was chosen more often than the grey one (the same did not occur with grey and bluish-grey discs). This too is taken to suggest that the grey banana is still perceived as yellow. But one potential problem with the design is that the question "which looks grey?" doesn't have a unique answer, since the two bananas had similar color and could both be called "grey". Although the instructions said that one of the items was more purely grey than the other, there could be variation in what people take to be a purer grey. The fact that subjects did not show a clear preference for the grey disc over the bluish-grey suggests that. Since both could be considered grey, it is possible that subjects chose the bluish-grey banana more often because it was the most clearly not yellow of the two, and not because the grey one looked yellow. A better question here, it would seem, would have been "which banana is the same grey as the background?". This would require a more direct comparison, instead of relying on people's judgment about what counts as a purer grey (which might be context sensitive).

It seems then that it is not easy to demonstrate that memory color affects color experience, for the results can be reinterpreted in terms of influences of beliefs on judgments about color. One reason to favor this reinterpretation comes from Valenti and Firestone (2019). They developed a more straightforward experiment, in which

[11] For other concerns about Delk & Fillenbaum's experimental design, see Deroy (2013) and Machery (2015).

[12] In addition, Gross et al. (2014) failed in an attempt to conceptually replicate their findings. Valenti and Firestone (2019) did replicate the findings, but, as we will see below, they ran other conditions whose results cast doubt on the interpretation that the effects were perceptual.

people were presented with three images side-by-side (different combinations of discs and bananas, grey or bluish-grey) and were asked to select "the odd color out". The idea is that, if the memory color effect is real, grey bananas should look yellowish, bluish-grey bananas should look grey, but grey and bluish-grey discs should look the color they are (because discs don't have a canonical color). In a scenario, e.g., where a grey banana, a bluish-grey banana and a grey disc are presented, the memory color effect predicts that people should choose the grey banana as the odd color out (and not the bluish-grey). Subjects, however, did not respond according to the memory color view in any of the combinations, but were mostly accurate in their answers. This suggests that color memory or knowledge doesn't affect color appearance. Grey bananas don't really look yellow. Subjects' beliefs about the typical color of objects may affect performance (such as in the previous experiments), but not by altering color appearance. And if there is no difference in color appearance depending on the object, then there is no cognitive penetration.

But even if we granted that objects do affect color experience, we might still reject the conclusion that this is a genuine case of cognitive penetration of perception. One alternative explanation for the difference in color experience (assuming it occurs) would be that color processing has access to other visual information of a stimulus, such as shape and texture, which are assumed to be part of early vision.[13] There might be a color-shape–texture processing system, whose database belongs to the visual system and which stores past color-shape–texture associations, which is encapsulated from higher cognitive processes.[14]

In addition, even if objects do affect color experience, this is a fast and most likely automatic process, probably insensitive to what one believes. Pointing out to the subject that the heart and the circle are the same color would presumably not change their experience. The analogy here is with visual illusions such as Muller-Lyer's, where learning that the two horizontal lines are the same length doesn't make them look the same. Visual illusions are used to illustrate how perception can be encapsulated from what we believe, for we continue to perceive something that we know not to be the case. If subjects formed the belief that the heart and the circle are the same

[13] This is supported by Olkkonen et al. (2008). Subjects made significant adjustments of objects' colors to the opponent color especially when the stimuli were realistic photographs of fruits, but the effect was smaller with less realistic images (with no texture) and absent with pure outlines of shapes of fruits. Given that recognizing the object is not enough for the effect to occur (as in the outline shapes condition), Olkkonen et al. point out that memory color effect, when it occurs, must be due to information about visual properties stored in a visual representation, and not to the activation of a semantic representation. This means that these experiments don't really show cognitive penetration, even if they show that color experience is affected, because this is likely not caused by beliefs or concepts (see also Deroy, 2013).

[14] Other possibilities are suggested by Deroy (2013) and Brogaard and Gatzia (2017). Deroy argues that multi-modal representations containing all and only sensory information about objects can affect color experience. These representations are assumed not to be conceptual, but they are not visual either, for they integrate information from all sensory modalities. On some views (such Fodor's and Pylyshyn's) this would probably still count as cognitive penetration. Brogaard and Gatzia (2017) suggest that color constancy could explain the effect, and so it would not qualify as a case of cognitive penetration.

color, but still perceived the heart as being redder, this would actually speak in favor of the view that color experience is encapsulated from cognitive states.

We have seen how color experience, through the memory color effect, has been taken as a case study for the cognitive penetrability of perception. I suggested that the studies showing the memory color effect, including the newer ones, are not successful in establishing that knowledge of the typical color of objects affects color experience proper, and not only post-perceptual judgments. This is what Machery (2015) calls “the *locus* problem” which, as he notes, seems to plague many attempts to show cognitive penetration: the evidence that tends to be provided in its support is often insufficient to determine the real locus of the effect, whether perceptual or post-perceptual. The studies reviewed here do not provide evidence for the cognitive penetrability of color experience, for they fail to establish the locus of the effect to be truly perceptual. In addition, as we have seen, cognitive impenetrability could still be preserved even if we assumed the effect to be perceptual and accepted that objects affect color experience, for this could be explained by processes involving information stored in the visual system. I will now show that similar considerations apply to the case of the influence of language on color perception.

3 Language and color experience

It is doubtful that knowledge of the typical color of objects affects how colors are experienced. But could one’s language affect color experience? Languages vary in how many basic colors terms they have. And even when two languages have the same number of color terms, they might not be coextensive (e.g. Berinmo and Himba, cf. Roberson et al. 2000, 2005). Can language A, with 5 color terms, cause its speakers to see colors differently from speakers of language B, with 11? In particular, will two colors look more similar when they receive the same name than when they don’t?

When a trichromat and a dichromat disagree about whether a given color sample is green, it is natural to assume that the disagreement originates from differences in color appearance, which in turn result from differences in cone types. How about disagreements between speakers of different languages? Although perhaps not as common, we might imagine English and Himba speakers disagreeing about whether the color of the sky and the color of the leaves should be placed in the same color category. English speakers would say that they belong to different color categories (leaves are green and the sky is blue), whereas Himba speakers would disagree (they would say that both are *yas*). But here, the source of the disagreement is open to debate. Do the color of the leaves and the color of the sky end up *appearing* more similar to Himba speakers than they do to English speakers? Or is the disagreement purely linguistic: colors look exactly the same, but they are categorized differently because of the languages one speaks?

If linguistic penetration occurs in color perception, there must be differences in how colors appear to speakers of different languages (i.e. differences in color experience, or phenomenology), and not simply in their judgments about colors. Moreover, let’s assume that for linguistic penetrability to occur it is necessary that differences in color experience are caused by information stored outside the visual system, being unexplained by perceptual principles or by differences in attention. More specifically, color experience is *linguistically* penetrable if linguistic representations directly cause colors to look different to speakers of different languages. Linguistic representations are assumed to be part of higher order cognition, so if they do affect color experience, this would also be a case of cognitive penetrability.

Just like with the case of memory color effect, it is not easy to experimentally test whether language affects color experience, but some attempts have been made. Studies comparing speakers of different languages found a number of behavioral differences in tasks involving color similarity judgements, color memory and discrimination. Let's consider first similarity judgment tasks. In a famous experiment, Kay and Kempton (1984) compared speakers of English and Tarahumara, an indigenous language spoken in Mexico. Tarahumara has one word for both blue and green. The researchers wanted to see if English and Tarahumara speakers would differ in how similar or different they judged colors in the blue-green area of the spectrum to be. Subjects were presented with three color chips at a time and were asked which of the three was the most different from the other two. When one of the colors belonged to a different linguistic category from the other two, it was usually judged by English speakers to be the most different, even when it wasn't.[15] Tarahumara speakers, on the other hand, didn't exaggerate distances when two colors crossed the blue-green divide. Roberson et al. (2000, 2005) got similar results when comparing, among other things, the similarity judgements of English, Berinmo and Himba speakers.

A natural reaction here would be to say that what best explains the fact that English, unlike Tarahumara speakers, judge color chips A and B (both called "green") to be more similar than they judge color chip B and color chip C (called "blue") to be, even when B and C were in fact closer, is that they *perceived* their distances differently. As Macpherson might put it, assuming that subjects perceived distances accurately, but judged them to be different, would be to attribute to them a gross form of misjudgment. In fact, English speakers did report that the color chosen as the most different *looked* more different to them (Kay and Kempton, p. 72). Why else would subjects make these similarity judgements unless colors *looked* that way to them? That is how Roberson et al. (2000, 2005) interpret their own results, which they take to "indicate that cultural and linguistic training can affect low-level perception" and that "the structure of linguistic categories distorts perception by stretching perceptual distances at category boundaries" (Roberson et al. 2000, p. 394). In their view, colors that belong to different categories look more different than colors that belong to the same linguistic category (even when they are equally distant, according to a color metric). Given that color categories vary from language to language, how similar or different two colors appear will to some extent depend on the categories in one's language. If differences in similarity judgment between speakers of different languages are indeed caused by differences in color experience, and linguistic

[15] What the authors took to be the real distance between colors was the measure of discrimination distance, whose unit is a just noticeable difference between colors.

representations are directly responsible for these differences, then this would be a case of color experience being linguistically penetrated.

However, there is an alternative explanation for differences in similarity judgment. The fact is that we may not be as good as we think we are in assessing our experiences, especially when differences in stimuli are small, as they were in these experiments. Kay and Kempton are in fact cautious when it comes to interpreting their results. According to them, subjects might well be using what they call a “name strategy”. When asked to perform a difficult task, they might be unconsciously reasoning that color chip C has a different name from A and B, and therefore is probably the most different.[16] This is assumed to be unconscious because subjects reported not to be relying on color names. If subjects were using the name strategy, that would mean that color terms affected judgements about color distance, but not color experience proper.

In order to test whether English speakers could judge color distances more objectively, Kay and Kempton designed a second experiment intended to block the use of the name strategy, by changing the method of presentation (only two color chips were visible at a time) and the instructions given. This time they asked participants to tell “which is bigger: the difference in greenness between the two chips on the left or the difference in blueness between the two chips on the right” (p. 73). They reasoned this would prevent subjects from using the name strategy, because they would be prevented from classifying the intermediate color chip as belonging to either blue or green. They were successful: with these differences in experimental design, English speakers’ similarity judgements were closer to the real distance between colors and not influenced by linguistic categories. Were their initial similarity judgments caused by permanent distortions in color perception (and not by implicit use of language, or categorization), we would expect judgments to remain unchanged after little tweaks in the experimental design, but they didn’t.

Let’s now consider color memory experiments. In a typical experiment (cf. Roberson et al. 2000, 2005) subjects are shown a color chip for a few seconds, which is removed from sight, then two color chips are presented and subjects have to indicate which they had just seen. Subjects remember better which chip they saw when the two options belong to different linguistic categories, so performance is different in speakers of different languages. We might be tempted to take this as suggesting that colors that cross a linguistic category appear more different, which is what makes it easier for English speakers, say, to remember the color they saw was blue when the alternative is green than when it is another shade of blue. But there is another equally plausible explanation for the results. As Pinker points out, experiments like these show only “that subjects remembered the chips in two forms, a non-verbal visual image and a verbal label, presumably because two types of memory, each one fallible, are better than one” (1994, p. 65). If an English speaker is shown a blue chip, she might register both the visual appearance and the name “blue”. When forced to choose between a blue and a green chip, it will be easier to say that the

[16] See Silins (2016) for discussion of what he calls “the belief response” to claims of cognitive penetrability and its implications for our access to our own minds.

blue chip is the one she saw, because the two forms of memories give the same result. When the alternative also gets the name “blue”, the subject will only be able to rely on visual memory, and might then be more prone to error. Color names might then help color memory without altering color appearance. Different performances by speakers of different languages might be explained by color names being available to some subjects but not to others.

We shouldn’t then immediately conclude, as Roberson et al. (2000) do, that speakers of different languages make different judgments about color distances, and exhibit differences in color memory, because they *see* colors differently. In perceptually difficult tasks, where differences in stimuli are small, subjects might use color names as a strategy to resolve uncertainty, just as in memory color experiments subjects might have used knowledge of the typical color of objects to help them solve the tasks. And if there is no evidence that speakers of different languages have different color experiences, there is no evidence for linguistic penetration.

However, other experiments compare performance of speakers of different languages in tasks that aim to be purely perceptual (which don’t rely on verbal reports or memory), and hence less subject to the name strategy. Winawer et al. (2007), for instance, compared English and Russian speakers in a timed color discrimination task. The Russian language has two basic color terms for blue: “goluboy” (light blue) and “siniy” (dark blue). On a computer screen, subjects saw a triad of color squares, two at the bottom and one above. They had to indicate as quickly and accurately as possible which of the two bottom squares had the same color as the square above. Sometimes the alternative colors belonged to different Russian categories, sometimes they belonged to the same category (both were goluboy, or siniy). Subjects performed the task in three different conditions: with no interference, with verbal interference and with spatial interference.[17] The researchers reasoned that, if language affects performance, then Russian speakers should show different patterns of reaction time than English speakers, and they should behave differently in the verbal interference condition. That is what they found. In trials without interference and with spatial interference, Russian speakers exhibited what is usually called *categorical perception*: they were faster at discriminating colors when they belonged to different categories in their language than when they belonged to the same category, even when the colors were equally spaced according to the adopted color metric. Categorical perception didn’t occur in trials with verbal interference, presumably because linguistic representations were being recruited for something else. English speakers didn’t show categorical perception under any condition.

With a different experimental design, Roberson et al. (2008) compared English and Korean speakers on a visual search task. The Korean language marks a distinction in the green region of the spectrum not marked by English. Participants had to fixate on a cross that was then surrounded by ten color squares. Nine were identical in color (the “distractors”) and one, the “oddball”, differed in hue from the others. In

[17] In the verbal-interference condition, “subjects silently rehearsed digit strings while simultaneously completing the color discrimination trials” (7781). In the control, spatial interference condition, “subjects maintained a spatial pattern in memory while completing color discrimination trials.” (7781).

some trials the oddball belonged to the same Korean category as the others, in some it belonged to a different linguistic category. Participants had to indicate on which side of the display (left or right) the oddball was presented. Korean speakers showed categorical perception: they were faster at identifying the location of the oddball when it belonged to a different color category from the distractors than when it belonged to the same color category. English speakers, on the contrary, did not show categorical perception, being equally fast on both types of trials.

Many seem to believe that categorical perception findings should be interpreted as showing that color experience is different in speakers of different languages. For instance, Zhong et al. (2018) claim that "studies based on different languages have demonstrated that language affects people's perceptions of color, as color perception varies according to the location of category boundaries in different languages" (p. 361). And here again, this is a natural interpretation. Russian speakers were faster at discriminating colors that belonged to different linguistic categories because they *looked* more different than colors that were equally distant but that belonged to the same category. English speakers did not show the same pattern of reaction time because they did not perceive color distances in the same way. The same explanation could be given to the results of similar studies, such as the one with Korean and English speakers. If these results indeed show that color experience is different in speakers of different languages, then this provides some evidence for linguistic penetrability.

But just as with the previous findings, we should be cautious in concluding that differences in reaction time result from differences in perceptual experience. Winawer et al. in fact consider that linguistic representations affect performance, but possibly not by promoting long-term distortions in the perception of color distances. One possibility we can envision is that color discrimination tasks induce the quick use of different sources of information (such as concepts or linguistic representations) to assist performance. Just as it might be easier to remember which color one has seen when both names and visual memory distinguish a target from the alternative, so too it might be easier to visually discriminate colors when two sources of information say they are different: visual appearance and language (or conceptualization). When two colors belong to the same color category, on the other hand, subjects will only be able to rely on how they look to discriminate them, and not on how they are categorized. In fact, in this case, as Roberson and Hanley (2010) note, subjects will have conflicting information: linguistic representations (or concepts) will group the two colors together while visual appearance will indicate that they are different. This might make the decision harder, making room for hesitation and so taking more time. It seems possible, then, that linguistic representations (or concepts) might interfere at a post-perceptual stage, affecting how quickly one makes perceptual decisions without affecting color phenomenology.

To help illustrate the point that differences in reaction time don't necessarily indicate differences in color phenomenology, we can consider the Stroop effect.[18] It is well known that subjects are faster at naming the ink color of a word when the word

[18] I thank an anonymous reviewer for the suggestion.

is congruent with its ink color (such as when "red" is written in red), and slower when the word is incongruent with its ink color (such as when "green" is written in red). No one takes this difference in reaction time to indicate that subjects experience different colors depending on the word written. Rather, for some reason (which needn't concern us here) the automatic process of reading seems to interfere with the process of naming the ink color when the two are incongruent, without affecting color experience. Just as conflicting semantic and visual information can slow reaction time in the Stroop effect, with no difference in color phenomenology, so too it seems possible that conflicting categorical and visual information can produce slower reaction times in visual tasks without affecting color experience.[19] If that is the case, then differences in categorical perception in speakers of different languages don't show that they have different color experiences.

A stronger reason for not assuming that categorical perception indicates a difference in perceived color similarity and dissimilarity is that it was eliminated when Russian speakers performed the task with verbal interference. If categorical perception indicated an alteration in perceptual experience, this would mean that Russian speakers saw different colors in the no-interference condition (where they exhibited categorical perception) than they did in the verbal interference condition (where categorical perception disappeared), which does not seem very plausible.[20] As Winawer and Witthoft (2013) note,

> "If a category effect goes away when labels become unavailable or not useful, then it is unlikely that the effect is due to color terms affecting early perceptual processes. While such an account is logically possible, it would require color appearance to be altered only during those moments when one is accessing the labels. A more parsimonious explanation is that the *decision* process is affected by language." (p. 5)

[19] As Winawer and Witthoft note, "when one makes a judgment or decision about color appearance, the knowledge that a color belongs to a particular category might affect the speed of the response or the content of the response without affecting the appearance of the color." (Winawer; Witthoft, 2013, p. 08).

[20] A similar difficulty applies to the interpretation of other findings, such as the lateralization of categorical perception of color to the right visual field (RVF). This was first observed in an influential study by Gilbert et al. (2006), and they explain the lateralization by saying that information coming from the RVF is more subject to the influence of linguistic representations, which are assumed to be stored in the left cerebral hemisphere (the same responsible for processing visual information from the RVF). Linguistic representations are accessed and interfere in discrimination when a target is presented to the RVF, but not the left. Here again, if we were to assume that categorical perception is a sign of a distortion in color experience, we would have to say that the colors participants *saw* in the RVF were different from the colors in the left visual field, which doesn't seem plausible. It's worth noting, however, that some (Witzel and Gegenfurtner 2011, Suegami et al. 2014) have failed to replicate the finding of lateralization of categorical perception. Similarly, other studies suggest that categorical perception is highly malleable and can have different patterns of occurrence with the same color stimuli depending on how they are categorized by a group of speakers of the same language (Zhong et al. 2018). But this too suggests that categorical perception is more likely an effect of color categories on post-perceptual, rather than perceptual processes. The name "categorical perception", as some have pointed out (Clifford et al. 2012) is then misleading.

Here is the central problem with the common interpretation of categorical perception as indicating perceptual differences directly caused by language. While the elimination of categorical perception by verbal interference is what suggests that language is responsible for categorical perception in no interference trials, it is also what suggests that the role of language is most likely not that of distorting perception of color distances. This is because, had language altered perception, we should expect categorical perception to persist even with verbal interference. The fact that it doesn't suggests that language does not alter perception. If however categorical perception had not been eliminated by verbal interference, this would have been a strong indication of an alteration in perception. But it would also have indicated that linguistic representations are not the proximate cause of differences in color appearance. This is because the elimination of categorical perception by verbal interference is what suggests that linguistic representations are directly responsible for categorical perception in normal trials; if there were no such disruption, there would be no such evidence. So when it comes to interpreting categorical perception effects, it seems that the very fact that indicates the role of linguistic representations (the elimination of categorical perception by verbal interference) also suggests that the effect is not perceptual. And the evidence that would have shown the effect to be perceptual (the persistence of categorical perception in the verbal interference condition) would not be evidence for the role of linguistic representations. Therefore, categorical perception and its elimination cannot show what many researchers take them to show: that linguistic representations directly affect color experience.[21]

The experiments that seem at first to suggest more strongly that there might be variation in how similar or different two colors appear to speakers of different languages are the ones using EEG. Thierry et al. (2009), for instance, compared brain responses to unexpected color change (unrelated to the task subjects were performing). Color could change from light to dark blue and vice versa, and from light to dark green and vice versa. They found differences between Greek and English speakers in early stages of visual processing. The Greek language, like Russian, has two different basic color terms for light and dark blue. They measured visual mismatch negativity (vMMN), which is a brain response to preattentive change in stimulus, and found that "the vMMN effect was (…) significantly greater for blue than green deviants in Greek participants", but not in English participants. The authors take the results to suggest that language affects "preattentive and unconscious aspects of perception" (p. 4567).

But it is unclear whether these differences in preattentive and unconscious stages of color processing correspond to differences in conscious color appearance. Different early electrophysiological responses to the same stimuli in speakers of different languages might indicate that color categorization is automatic and happens very

[21] Another thing that suggests that Winawer et al. results do not really indicate perceptual differences in English and Russian speakers is that Russian speakers only showed categorical perception when colors were near (and discrimination was harder). This suggests that concepts or linguistic representations can be recruited in order to assist difficult color discrimination. If color names had permanently altered color distances, it would be natural to expect categorical perception to have occurred with far colors as well.

fast, rather than point to differences in color phenomenology. More importantly, while Fonteneau and Davidoff (2007) and Forder et al. (2017) have also found category effects in early event-related potentials (ERPs) components associated with perceptual processes, Clifford et al. (2012) and He et al. (2014) only found color category effects after 200 ms post stimulus presentation, in ERP components that are believed to reflect post-perceptual processes. The evidence available is inconclusive and more research is then needed.

4 Differences in color experience do not imply linguistic penetrability

As we've seen, the issue of whether color perception is linguistically penetrable only arises if color experience is affected by linguistic representations. And as of yet there is no strong evidence that speakers of different languages indeed *see* different colors. Many of the studies purportedly showing that language affects color perception, just like the studies on memory color effect reviewed in Sect. 2, suffer from the locus problem. They fail to establish that the influence of language is truly perceptual. It's possible to interpret behavioral differences as color words (or concepts) affecting post-perceptual processes. More specifically, decisions about color identity (such as "*this* color is the same as *that* one"), for example, might involve not just sensory information, but also linguistic representations, which "meddle in even surprisingly simple objective perceptual decisions." (Winawer et al. p. 7784). And if color experience is the same in speakers of different languages, then words or linguistic representations do not really penetrate perception.

But in this section I will work with the assumption that speakers of different languages have different color experiences. Would that mean that color experience is linguistically penetrable? As I mentioned earlier, one way of denying cognitive penetrability when different individuals (or the same individual at different times) have different perceptual experiences is to say that they are due to differences in *attention*. Were they to attend to the same things, they would see the same things. Could this explain the assumed difference in color experience in speakers of different languages? In the case of Thierry et al.'s results probably not, for they focused on *preattentive* change detection, where differences in attention don't seem possible. But in the case of categorical perception (assuming it indicates differences in color experience), maybe Russian and English speakers have different attention patterns when they see light and dark blue.

But if there is a difference in attention here, it must be different from the usual cases. It is a difference that presumably cannot be resolved by indicating a place or a feature to which attention should be directed. When botanists distinguish elms from beeches faster than me, that might be because they know where to direct their attention when comparing their leaves, say. Even if we have different perceptual experiences when looking at the trees, that might just be because we are not attending to the same things. It seems unlikely, though, that the same sort of expertise is at play in color discrimination. It would be strange to say that Russians (unlike English speakers) know where to direct their attention, when looking at light and dark blue squares, because the stimuli in the experiment were simple uniformly colored squares differing only in color. There was nothing there to focus attention on except colors themselves. So assuming there is a difference in color experience between Russian and English speakers, attention doesn't seem to work to dismiss this as a genuine case of linguistic penetrability.[22]

If speakers of different languages had different color experiences not due to attention, that would be evidence for top-down processes being involved in producing color experience, given that color experience could not be entirely explained by the wavelength received, together with the particular characteristics of our visual apparatus. Past experience would somehow affect how colors are experienced. Much like the memory color effect says our past experiences with hearts makes us see them as redder than they are, so our past experience using words such as "blue" and "green" might make two colors that cross this boundary look more different than equally spaced colors that receive the same name.

But although it is necessary for linguistic penetrability that speakers of different languages have different color experiences, this is not sufficient. For this to be a case of linguistic penetrability of color perception, the proximate cause of differences in color experience must be linguistic representations. Researchers tend to agree that linguistic representations are directly responsible for categorical effects in perception and memory, even when they differ in their interpretation about their exact role (whether it is perceptual or post-perceptual). We should, however, distinguish between *linguistic representations* and *color categories*. Although there is evidence that color categories drive categorical perception, there is also evidence that these categories might not be linguistic. Holmes and Wolff (2012), for instance, found that categorical perception can occur not only with named, but also with novel unnamed categories. This suggests that color names, or linguistic representations, are not necessary for the occurrence of categorical perception.[23] In addition, other studies show that color categorization does not depend on language, since pre-linguistic infants as young as 4 months old show signs that they group colors in categories (Skelton et al. 2017). Although it is possible that linguistic representations drive categorical perception in some cases but not others, it is simpler to assume that categorical perception has a common cause in all its manifestations: color categories, not linguistic representations. It's possible that color categories are shaped by past discriminations of colors, which in turn were induced by language use, without these categories being linguistic in nature (Krempel 2018).

[22] According to Slobin (1997), language affects how one attends and thinks about a scene in order to speak about it (what he calls thinking for speaking). Speakers of different languages might attend and describe the same scene in different ways, depending on how events are usually encoded in their language. This can result in the same event being "experienced differently by speakers of different languages – in the process of making a verbalized story out of them" (p. 88). This is one way in which language might affect perception via differences in attention. However, I take that this is not a case of linguistic penetrability, because if attention were held fixed, speakers of different languages would see the same things.

[23] Another indication that linguistic representations might not be responsible for categorical perception is that it has also been observed, lateralized (see note 20 above), with non-basic categories of warm and cool colors, which are not frequently distinguished verbally (Holmes and Regier 2017).

If linguistic representations are not the proximate cause of the assumed differences in color appearance in speakers of different languages, this would not be a case of linguistic penetrability of perception. It could however still be a case of cognitive penetrability, depending on how we conceive color categories. There are at least two possibilities: they might be perceptual categories that belong to the visual system, or they might be higher order, conceptual representations, which are the constituents of our propositional attitudes about colors and which can be linguistically expressed, but which could exist even in the absence of language. Only the latter would count as cognitive penetrability, having particularly interesting consequences for how we conceive the cognition/perception divide.

One reason to take color categories to be perceptual categories, rather than conceptual, is provided by some ERP studies. Thierry et al. (2009), for instance, found differences in brain responses in speakers of Greek and English as early as 100 ms after stimulus presentation, which, as they note, is “a time range associated with activity in the primary and secondary visual cortices” (p. 4569). If these results prove to be reliable (for as I said, other ERP studies failed to find differences in early perceptual processes), they might be taken to suggest that color categorization is a visual, not a cognitive process.[24]

If the background information affecting the output of the color module belonged to the visual system, then that would be a top-down, but not a cognitive effect on perception. So even assuming that our personal history of color discrimination (which is in many cases motivated by language) may shape color categories, and that color categories in turn somehow affect color appearance, this still doesn’t show that color perception is cognitively penetrable. Even if growing up speaking English (and not Russian) leads one to continuously group dark and light blue together, so that they belong to the same color category and end up looking more similar to them than to Russian speakers, that in itself would not show that color experience is cognitively penetrable. The color categories affecting experience would have to be conceptual, not visual, for this to be a case of cognitive penetration.

One thing that would speak in favor of color experience being affected by perceptual color categories while still being encapsulated from cognition is it being different in speakers of different languages, but unaffected by current beliefs about color distances. If English speakers perceived two colors called “blue” to be more similar than a color called “blue” and a color called “green” (when both pairs are equidistant), and their experience did not change if they came to believe both color pairs to be equidistant, that would suggest that the output of the color module is affected by perceptual color categories, but encapsulated from higher cognition. I take that this is a reasonable prediction of what would happen if color experience is indeed affected by perceptual categories, though empirical confirmation would be needed. In sum, if color categories belong to the visual system, and not to cognition, then their influence on color experience would not render color perception cognitively penetrable.

[24] Mandelbaum (2018), for instance, argues that perception outputs conceptualized representations, based on findings suggesting that categorization happens very fast.

Now one might ask: if speakers of different languages had different perceptual color categories, which in turn affected color experience, wouldn't that show that language affects perception? Shouldn't we then say that color experience is linguistically penetrable? If that happened, this would be a case of what we could call diachronic linguistic penetration (cf. Fodor 1984), in which our current color experience is in part explained by the color discriminations we performed in the past, which was partly motivated by language. The following counterfactual would seem to hold: If S had learned language X instead of Y (where X and Y differ in at least one color term), S would have perceived at least some color distances in a different way. So in the scenario we are imagining, we could say that language is a *distal* cause of differences in color experience in speakers of different languages, and concede that language *indirectly* affects perception by promoting discriminations which in turn affected color categories. But in this kind of diachronic influence, the role of language would not be as strong as it is usually assumed to be. The same kind of perceptual alteration (assuming it occurs) could have happened without language, provided that similar patterns of color discrimination had occurred. What we've been discussing here (and what is investigated in the experiments surveyed) is *synchronic* linguistic penetration, which takes linguistic representations to be the *proximate* cause of alteration in perceptual experience. Perceptual color categories affecting color experience would not amount to linguistic penetrability in that sense.

I've been working with the assumption that color experience is different in speakers of different languages due to color categories that are perceptual, not conceptual. One difficulty with this assumption is to explain why categorical perception disappears with verbal interference. Why should the functioning of perceptual categories be affected by language use? The elimination of the effect might be taken to suggest that conceptual (and not perceptual) representations are responsible for categorical perception and (by assumption) for differences in color experience. If conceptual representations explain categorical perception, and categorical perception indicates differences in color experience, then this would be an instance of cognitive penetrability of perception.[25] However, though elimination of categorical perception by verbal interference suggests color concepts are responsible for it, as we've seen, the fact that categorical perception can be eliminated also suggests that this is an effect of concepts on perceptual decisions, or on categorization, not on perceptual experience. So categorical perception provides no evidence that color concepts affect color experience. And again: no alteration in color experience, no cognitive penetration.

I argued that empirical studies don't show that speakers of different languages perceive colors differently, so they don't show color experience is linguistically penetrable. But however exactly all these experiments are interpreted, they do suggest that color terminology can affect color discrimination, even if not by altering color experience. Even if speakers of different languages experience the same colors, they

[25] More work needs to be done, both empirical and theoretical, in order to reconcile findings involving ERPs, which sometimes suggest that color categorization is a perceptual process, with findings of categorical perception in visual search and discrimination tasks, which are better explained, in my view, by conceptual representations (Krempel 2018).

still differ in similarity judgments, in how well they remember certain colors and in how fast they judge two colors to be the same. These are differences we would expect to observe if they really experienced colors differently. Assuming they share perceptual experiences, the findings seem to show that behavior is not just affected by what we experience, but also by how experiences are categorized. And language is most likely affecting that, even if indirectly. Although interesting, this doesn't imply the kind of interaction between language and perception that proponents of linguistic penetrability often claim to occur.

5 Epistemological consequences

I've been considering different possible influences language might have on color perception and discrimination, and whether or not they should count as linguistic or cognitive penetration. Linguistic penetration would occur if linguistic representations directly affected color experience, and cognitive penetration if color experience were affected by color concepts. If perceptual color categories affected color experience, that would be an instance of a top-down effect on perception, but not of linguistic or cognitive penetrability.

As I said, the evidence available is insufficient to determine whether speakers of different languages have different color experiences, for several experiments aiming to show that allow for plausible reinterpretations. The issue of whether color experience is affected by language is empirical, and more investigation is needed. But if we assume it is true that speakers of different languages have different color experiences, would it really matter how that comes about: whether perceptual experience is directly affected by linguistic representations, by color concepts or perceptual color categories? It seems that either way, perceptual experience would produce indistinguishable behavior, regardless of what caused the differences in perceptual experience in the first place. These distinctions are important, though, if what we care about is cognitive architecture. They indicate different ways in which perception, cognition and language could interact, and different possible etiologies of perceptual experience. But I believe they are less important if what we care about is epistemology.

In order to see this, we can contemplate what epistemological consequences there would be if it turns out that speakers of different languages do after all have different color experiences. If that happens, then color experience will not be completely shared, for whether one perceives colors A and B to be closer together than B and C will depend in part on what language one grew up speaking. But importantly, color experience not being shared would not be a consequence specific to *linguistic* penetrability. If color experience were different in speakers of different languages not because of linguistic, but because of conceptual or perceptual representations, experience of color distances would still not be widely shared. Color experiences would be relative to language and culture. This threatens the view that perceptual experience can provide us with a common, neutral source of knowledge. We would not have, through perceptual experience, the kind of access to the world we ordinarily think we have. The problem, then, doesn't result from linguistic penetrability per se, but instead from top-down effects on perception (even those within perception) which are subject to variation among individuals.[26] Here, then, all possible causes of differences in color experience lead to the same epistemological conclusion, that experiences are not shared and so the world may not be directly accessible to us. Now, one might say that color experience not being shared would not be a problem, given that we already know that people frequently differ in how they experience colors. But as I said in the beginning, the differences in color experience we are familiar with are generally due to biological factors; they are not due to differences in color categorization. And we have at least an intuitive way of deciding which color experiences are more authoritative, namely, those that reveal more colors (and not those of color-deficient individuals). If language could provoke differences in how one experiences color distances (even if it were only the distal and not the proximate cause of them), that would indicate that perceptual experiences can differ in more extreme and unexpected ways. And how could we decide whose color experience is closer to revealing true color distances? Perhaps the very methods used to establish color distances would have to be revised, for they would have to take into account the differences caused, directly or indirectly, by language.

Linguistic, like cognitive penetrability (Siegel 2012), also creates problems for the view that perceptual experience is sufficient to justify perceptual beliefs. For even when things appear to be going right (good illumination, good visual acuity, no hallucination, etc.), language would be affecting perceptual experience, at least sometimes making us see things not as they are. Cognitive penetrability can, in some cases, introduce a problem of circularity, as Siegel points out. If I experience bananas as yellow because of my belief that bananas are yellow, then it is questionable that my perceptual experience of bananas as yellow can serve to justify my belief that bananas are yellow. The experience that is supposed to justify my belief is itself caused by that very belief. Similarly, the way colors look might be taken to justify the belief that colors A and B are closer together than B and C, as well as the belief that colors that belong to the same linguistic category are more similar than colors belonging to different categories. But if colors look the way they do because of my belief that A and B, unlike C, belong to the same linguistic category, and because I believe that colors that belong to the same linguistic category are more similar than colors belonging to different categories, then it would be doubtful that color experience provides justification for the belief. Here too, justification would be circular.[27]

But it is important to highlight that if the experience of color distances were affected by perceptual color categories which were shaped by past experience and language use (and not by linguistic or cognitive representations), that too would

[26] Here I'm following Pylyshyn (1999) in accepting that not all cases of top-down effects on perception are cases of cognitive penetrability of perception, even if all cases of cognitive penetrability are cases of top-down effects. As Raftopoulos (2009) notes, "top-down flow of information is compatible with a cognitively impenetrable perception" (p. xx). Some, however, use these terms interchangeably (cf. Silins, 2016).

[27] Lyons (2011) argues that circularity is not what is epistemically problematic with cognitive penetration. Cognitive penetration is bad when it decreases the reliability of perception. I take it that circularity can be a problem in some cases of cognitive penetrability (as in the cases just described), but I agree with Lyons that cognitive penetration can be epistemically bad even without it, such as when desires or fears penetrate perception, where presumably no circularity is involved.

challenge the view that perceptual experience is sufficient to justify perceptual beliefs. This is because how similar or different we experience colors to be would depend on our past experiences of colors, and not only on how similar or different they really are. And there would seem to be no reason to prefer one kind of influence of past experience over another as the one that more accurately reveals color distances. It seems that in all these cases, and not just in the case of linguistic or cognitive penetrability, the top-down influence on perception would be provoking a kind of insensitivity to stimuli (Siegel 2012; Lyons 2011). Whatever exactly caused the assumed differences in color experience (whether perceptual categories, concepts or linguistic representations), it would be questionable that color experience justifies our beliefs about color distances. The difficulties pointed out by Siegel are therefore not restricted to cases of cognitive penetration, but result from top-down effects on perception, being therefore more pervasive than she and others anticipated.

There is, however, a difference between cases where perceptual categories affect color experience, and cases of cognitive or linguistic penetrability. In the first case, the circularity of justification would not be a problem. Since there would be no beliefs affecting how colors look, this would not be a case where perceptual experience is affected by the very belief it is supposed to justify. But one could insist that perceptual experience would not offer proper justification for the belief it is supposed to justify, for the experience would not be giving us the color distances as they are. An insensitivity to stimuli would be at play. So denying that the direct cause of differences in color experience is linguistic or cognitive, saying it is instead perceptual categories, would not be of much consolation either to those who wish to say that perceptual experiences are a common ground shared by different people, or to those who think they provide justification for perceptual beliefs.

One issue that could be raised here is that perhaps, as some have pointed out (Siegel 2012; Lyons 2011; Silins 2016), not all cases of cognitive penetrability are epistemically bad. Some might actually be beneficial. For instance, if expertise affects perceptual experience, it presumably affects it in a positive way, such that experts' experiences reveal more information about the world than naïve observer's experiences (Silins 2016). Assuming this is a genuine instance of cognitive penetration (not reinterpretable, e.g., by differences in attention), this would be a good case, in which cognitive penetration increases reliability, making one more likely to get things right (Lyons, 2011).[28] One might then wonder whether the same is true of the linguistic penetrability of color experience. Perhaps, if language affects color

[28] Another purportedly good (or at least neutral) case of cognitive penetration is Lyons' (2011) snake case, in which one is more likely to spot actual snakes when one believes there are snakes nearby. According to him, there is nothing epistemically bad here, even when the penetrating belief is unjustified. My own take is that this and cases of expert perception are not obvious cases of cognitive penetration. The belief that there are snakes nearby, for instance, could be affecting perceptual experience only indirectly, by directing one's attention to signs of snakes, therefore changing only the input to perception. If it really were affecting perceptual processing and its output, it could presumably do so whether it is true or not that there is a snake in front of me. It could then make me see snakes even where there were none, which would decrease the reliability of perception. This would therefore be a bad case of cognitive penetration, according to Lyon's own characterization. I find it hard therefore, to think of a genuine case of cognitive penetration that would be epistemically good or neutral.

experience (whether directly or indirectly), it does so in a positive way, such that it makes one more sensitive to real differences in the world.

But that does not seem to be true. There is no evidence that speakers of different languages are able to make more color distinctions when they have more color terms (Roberson et al. 2009). Unlike the expert perception case, language does not make one capable of perceiving new color differences. It does not make color experience reveal more information about colors. Assuming it does affect color experience, it does so in a biased way, making colors look more different just because they get different names, and not because they really are that way (given that categorical effects conflict with more objective measures of distances).

It is also not the case that speakers of a language that labels two colors differently will be faster at discriminating them than speakers of a language that doesn't distinguish them. English speakers were overall faster at performing visual discrimination tasks than Russian (Winawer et al. 2007) and Korean speakers (Roberson et al. 2008) despite lacking distinctions that are present in those languages. This suggests that it would be misleading to think of Russians, say, as experts on blue. Language didn't give them any special advantage. It is not that color discriminations would be a lot harder to make without linguistic distinctions, for if that were the case, English speakers would have been slower than Russians. Given that the linguistic penetration of color experience, if it occurs, does not seem to reveal more accurate information about colors, nor to make one better at discriminating colors than speakers of languages with fewer distinctions, it would not be a good case of cognitive penetration.

Now one could say that if linguistic penetration occurs, so much the worse for our epistemological ambitions. After all, epistemology should adapt to empirical results about perception, not the other way around. But even though this would not be a good case of cognitive penetrability, if language really affects color experience, its effects are so minimal as to not render any strong, noticeable disagreements about colors – certainly no stronger than disagreements between trichromats and dichromats. Most of us at least were not even aware of it until the empirical results came out. The differences found, though significant, were subtle, so if there really are differences in color experience across languages, they are probably small. This suggests that color experiences are similar enough, if not completely shared.

In addition, the input information is still relevant to the production of color experience. Language might make colors that are in fact close together look closer than they are (when they get the same name). Or it might make colors that are close together look farther apart (when they get different names). But even if language stretches color distances, it presumably does so not in a completely unrestricted way. It presumably can't make colors that are really far part look very close together, or colors that are very close together look really far apart.[29] And while languages do differ in how many color terms they have, and in their extension, it has been shown that differences are not completely arbitrary, but follow some universal restrictions (Berlin and Kay 1969; Regier et al. 2010). Language presumably cannot make us see similarities or differences that aren't really there. Even if there are top-down

[29] Winawer et al. (2007) report that language played a role only in more difficult, near-color discrimination, and not on colors that were farther apart.

processes involved in color perception, such that color experience is affected by stored information about color similarities and differences, we might still be justified in believing, on the basis of color experience, that two colors are similar or different. It is just that they might not be *as* similar or *as* different as they look.

Gaining awareness of how past experience affects our current experience, if it really does, could move us one step closer to objectivity. Even if our experiences are not under our direct control, and are not universally shared, we might be able to judge their accuracy based on more objective measures of color distance and, with these, get a better grip on reality.

References

Berlin, B., & Kay, P. (1969). *Basic color terms*. Berkeley: University of California Press.

Brogaard, B., & Gatzia, D. E. (2017). Is color experience cognitively penetrable? *Topics in Cognitive Science, 9,* 193–214.

Burnston, D. C. (2017). Cognitive penetration and the cognition–perception interface. *Synthese, 194,* 3645–3668.

Clifford, A., et al. (2012). Neural correlates of acquired color category effects. *Brain and Cognition, 80,* 126–143.

Delk, J., & Fillenbaum, S. (1965). Differences in perceived color as a function of characteristic color. *The American Journal of Psychology, 78*(2), 290.

Deroy, O. (2013). Object-sensitivity versus cognitive penetrability of perception. *Philosophical Studies, 162,* 87–107.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences, 39,* E229.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (1984). Observation reconsidered. In: *Theory of content*. Cambridge, MA: MIT Press.

Fodor, J. (1985). Précis of *The modularity of mind*. In *Theory of content*. Cambridge, MA: MIT Press.

Fodor, J. (1988). A reply to churchland's 'perceptual plasticity and theoretical neutrality. *Philosophy of Science, 55*(2), 188.

Fonteneau, E., & Davidoff, J. (2007). Neural correlates of colour categories. *NeuroReport, 18*(13), 1323.

Forder, L., et al. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. *PLoS ONE, 12*(5), e0178097.

Gilbert, A., et al. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, 103*(2), 489–494.

Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology, 8,* 221.

Gross, S., et al. (2014). Problems for the purported cognitive penetration of perceptual color experience and Macpherson's proposed mechanism. *Baltic International Yearbook of Cognition, Logic and Communication*. https://doi.org/10.4148/1944-3676.1085.

Hansen, T., Olkkonen, M., Walter, S., et al. (2006). Memory modulates color appearance. *Nature Neuroscience, 9,* 1367–1368.

He, X., et al. (2014). Color categories only affect post-perceptual processes when same- and different-category colors are equally discriminable. *Journal of the Optical Society of America A, 31,* A322–A331.

Holmes, K. J., & Regier, T. (2017). Categorical perception beyond the basic level: The case of warm and cool colors. *Cognitive Science, 41,* 1135–1147.

Holmes, K. J., & Wolff, P. (2012). Does categorical perception in the left hemisphere depend on language? *Journal of Experimental Psychology, 141*(3), 439.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist, 86,* 65.

Krempel, R. (2018). Color terms and perception. In: W. J. Silva Filho (Ed.) *Novos ensaios de filosofia analítica.* Pelotas: Dissertatio Filosofia.

Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues, 21,* 289.

Machery, E. (2015). Cognitive penetrability: A no-progress report. In J. Zeimbekis & A. Raftopoulos (Eds.), *The cognitive penetrability of perception: New philosophical perspectives*. Oxford: 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*, 24–62.

Mandelbaum, E. (2018). Seeing and conceptualizing. *Philosophy and Phenomenological Research, 97*(2), 267.

Mole, C. (2015). Attention and cognitive penetration. In A. Raftopoulos & J. Zeimbekis (Eds.), *The cognitive penetrability of perception*. Oxford: Oxford University Press.

Olkkonen, M., et al. (2008). Color appearance of familiar objects. *Journal of Vision, 8*(5), 13.

Pinker, S. (1994). *The language instinct*. London: Penguin Books.

Pointer, M., & Attridge, G. (1998). The number of discernible colours. *Color research and application., 23*(1), 52.

Pylyshyn, Z. (1999). Is vision continuous with cognition? *BBS, 22*(3), 341.

Raftopoulos, A. (2009). *Cognition and perception*. Cambridge, MA: MIT Press.

Regier, T. et al. (2010) Language and thought. In: Malt and Wolff (Eds) *Words and the mind*. Oxford: Oxford University Press.

Roberson, D.; Hanley, R. (2010). Relatively speaking. In: Malt and Wolff (Eds) *Words and the mind*. Oxford: Oxford University Press.

Roberson, D., et al. (2000). Color categories are not universal. *Journal of Experimental Psychology, 129*(3), 369.

Roberson, D., et al. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50,* 378.

Roberson, D., et al. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition, 107*(2), 752.

Roberson, D., et al. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition, 112,* 482.

Schirillo, J. (1999). Color memory penetrates early vision. *BBS, 22*(3), 393.

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs, 46*(2), 201–222.

Silins, N. (2016). Cognitive penetration and the epistemology of perception. *Philosophy Compass, 11*(1), 24–42.

Skelton, A. E., et al. (2017). Biological origins of color categorization. *PNAS, 114*(21), 5545.

Slobin, D. (1997). From 'thought and language' to 'thinking for speaking'. In: Gumperz and Levinson (Eds) *Rethinking Linguistic relativity*. Cambridge: Cambridge University Press.

Suegami, T., Aminihajibashi, S., & Laeng, B. (2014). Another look at category effects on colour perception and their left hemispheric lateralisation: no evidence from a colour identification task. *Cognitive Processing, 15,* 217–226.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567–4570.

Valenti, J., & Firestone, C. (2019). Finding the "odd one out": Memory color effects and the logic of appearance. *Cognition, 191,* 103934.

Webster, M. (2015). "Individual differences in color vision". In: Elliot and Fairchild (Eds) *Handbook of color psychology*. Cambridge: Cambridge University Press.

Winawer, J., & Witthoft, N. (2013). Effect of color terms on color perception. *Encyclopedia of Color Science and Technology*. https://doi.org/10.1007/978-3-642-27851-8_77-4.

Winawer, J., et al. (2007). Russian blues reveal effects of language on color discrimination. *PNAS, 104,* 7780.

Witzel, C. (2016). An easy way to show memory color effects. *i-Perception, 7*(5), 2041669516663751.

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision, 11*(12), 16.

Zeimbekis, J. (2013). Color and cognitive penetrability. *Philosophical Studies, 165*(1), 167.

Zhong, W., et al. (2018). Is the lateralized categorical perception of color a situational effect of language on color perception? *Cognitive Science, 42,* 350–364.