

When do ratings implicate perception versus judgment? The “overgeneralization test” for top-down effects

Chaz Firestone and Brian J. Scholl

Department of Psychology, Yale University, New Haven, CT, USA

ABSTRACT

When looking at an object—say, a banana—we can both directly *perceive* its visual qualities (e.g., its size) and also make higher-level *judgments* about its visual and non-visual properties (e.g., not only its size, but also its cost). Suppose you obtain a rating of a property such as size. Does that rating implicate seeing or merely higher-level judgment? The answer often matters a great deal – e.g., determining whether such ratings imply “top-down” effects of cognition on perception. Too often, however, this distinction is ignored in empirical investigations of such effects. Here we suggest a simple test for when such ratings can be used to implicate perception: whenever the very same experiment “overgeneralizes” to an unambiguously *non*-perceptual factor, the results cannot be used to draw implications about perception, *per se*. As a case study, we investigate an empirical report alleging that conservatives perceive Barack Obama as darker skinned than liberals do. Two simple experiments show that the very same effects, measured via the same “representativeness” ratings, obtain with unambiguously non-perceptual and even silly factors (involving bright red horns vs. halos, rather than brightness differences). We suggest that this renders such methods unable to implicate visual processing and we recommend that “overgeneralization tests” of this type always be conducted in such contexts.

ARTICLE HISTORY Received 20 October 2015; Accepted 16 February 2016

KEYWORDS Top-down effects; cognitive penetrability; lightness perception; racial categorization

A foundational distinction in psychology is that between perception and cognition – and in particular, between directly *seeing* something vs. merely *judging* that something is the case. For example, whereas we can directly see visual properties such as the color or size of a banana, we can only infer, conclude, or judge that the banana is expensive or is grown in South America. However, the same properties can often be directly seen *and* indirectly

CONTACT Chaz Firestone  chaz.firestone@yale.edu  Department of Psychology, Yale University, Box 208205, New Haven, CT 06520–8205, USA

© 2016 Informa UK Limited, trading as Taylor & Francis Group

judged: we can both see that an object is yellow and also conclude that an object is yellow even without directly seeing it that way – for example, if the lighting is poor but the object clearly has a banana’s distinctive silhouette.

Indeed, one can often experience the distinction between perceiving and judging for oneself, especially when the two processes conflict. For example, when viewing a visual illusion, one often *sees* the world in a way one *knows* it not to be – for example when looking at an impossible figure (e.g., a 3D Penrose triangle), or when scrutinizing two objects that appear to be different sizes even when a ruler gives the same measurement for each (e.g., the Ebbinghaus illusion). Similarly, in a more everyday context, drivers who learn that objects in convex passenger-side mirrors are “closer than they appear” may often *see* an object as being at one distance and yet *judge* the object to be at a different distance.

Perception vs. judgment in top-down effects

Intuitively, a great many provocative and potentially groundbreaking findings are much more groundbreaking if they reflect perception per se rather than post-perceptual judgment. For example, the possibility that perception is affected by “top-down” factors such as our beliefs, desires, intentions, emotions, and other higher-level states is potentially revolutionary (for a critical review see Firestone & Scholl, [in press](#))—but only if such states alter *what we see*, and not if they simply change the judgments we make on the basis of what we see (in which case they may not interact with foundational questions in perception research, though they may of course be interesting for other reasons). For example, it would be truly groundbreaking to discover that linguistic labels literally alter an object’s perceived color (as has been reported by Mitterer, Horschig, Müsseler, & Majid, 2009), but it would be far more routine if it turned out that people *see* the same colors but simply call them by different names in different linguistic contexts.

Although theoretical discussions of such top-down effects often acknowledge the importance of distinguishing perception from judgment (e.g., Balceitis, *in press*; Collins & Olson, 2014; Philbeck & Witt, 2015), a surprising number of empirical investigations seem to ignore this distinction altogether, both in discussion and in terms of the actual experiments run (for a notable exception, see Witt, Proffitt, & Epstein, 2010). This is particularly true for experiments that rely on subjective ratings and other explicit, deliberate, unsped responses, which seem especially easily contaminated by post-perceptual judgment.

A case study: partisanship and Barack Obama’s skin color

Consider a concrete case study of both theoretical and practical importance. It has been reported that partisan attitudes can alter a politician’s perceived skin

color, such that Republicans and conservatives perceive Barack Obama as having darker skin than do Democrats and liberals (Caruso, Mead, & Balcetis, 2009). The data used to fuel this conclusion were ratings of how well certain photographs “represented” the politicians they depicted: when subjects were shown multiple photographs of Obama, some having been lightened and some darkened, conservative subjects were more likely than liberal subjects to rate a darkened photo of Obama as more “representative” of him, ostensibly because they thought the photograph looked more like the real-life Obama.

Does such an effect reflect a literal perceptual distortion, such that Obama’s skin truly *looks* darker to conservatives? Or is it instead that conservative subjects simply make *judgments* consistent with darker skin, while actual percepts may be unchanged? The original paper sometimes interpreted these ratings as evidence of an effect on “skin tone representations” (Caruso et al., 2009, p. 20169) without implying an effect on visual perception *per se*; but at other times they suggested that the effect was on “lightness perception” (p. 20169) and framed their results in the context of unambiguous discussions of true visual perception (such as Balcetis & Dunning, 2010; Bruner & Goodman, 1947; Levin & Banaji, 2006). In any case, many other researchers have taken this result to unambiguously reflect changes in visual processing. For example, a detailed follow-up study interpreted the results as showing that “top-down processing influences color perception” (Kemmelmeyer & Chavez, 2014, p. 138), and other discussions describe the results as demonstrating that “political affiliation can affect color perception” (Branigan et al., 2013, p. 1661) and that partisanship affects not only “our social judgments of others” but also “our visual perception of their physical features” (Broudy, 2009, p. 5).

The overgeneralization test

Regardless of how this result has been cited and discussed, the experiments themselves simply do not distinguish between perceptual and judgmental interpretations. (This is not an indictment of the original Caruso et al., 2009, work, however; it is not clear that their project ever intended to explore this distinction, and in fact some of their conclusions may follow just as readily from a higher-level judgmental interpretation.) Because this distinction really *matters*, though—insofar as a perceptual interpretation has bold and even revolutionary consequences for the nature of visual processing—we are interested in whether “rating” data of this sort can fuel such exciting conclusions.

Unlike many phenomena in vision science, top-down effects of cognition on perception such as those discussed above are almost never possible to “see for yourself” (Firestone & Scholl, 2015). As a result, the only data to go on in the case of Obama’s perceived skin color are the ratings themselves,

since there is no associated “demo” to experience. Here, we ask whether such ratings implicate perception, and we propose a simple test to help determine the answer: if the same measures produce the same pattern of results with an unambiguously *non*-perceptual factor, then the method cannot be used to draw implications about perception.¹

For example, suppose that, instead of rating the representativeness of artificially lightened or darkened photographs, subjects instead saw photographs of Obama in which bright red horns or a yellow halo had been drawn on his head, and they rated *those* images for “representativeness”. If the effect *overgeneralizes*, such that the same pattern of results obtains even with this clearly non-perceptual case (with conservatives rating the Obama-with-horns image as more representative than the Obama-with-halo image), then the measure ‘fails’ the test – since, presumably, conservatives don’t literally see Obama as having bright red horns on his head. However, if the experiment does not produce the same pattern of results with unambiguously non-perceptual factors, then it may well be that the ratings reflect visual percepts *per se*.

The current experiments

Here we apply the overgeneralization test, exactly as described, to the case of partisan attitudes and perception of Barack Obama (Caruso et al., 2009). We altered images of Obama to have either bright red horns or a yellow halo on his head (Figure 1), and then we asked subjects to rate the “representativeness” of these images. We subsequently collected information about subjects’ political attitudes, including whether they identified as Republicans or Democrats, to see whether political attitudes predicted which image was rated as more representative.

Experiment 1: horns and halos

Method

Participants

A total of 221 subjects were recruited online through Amazon Mechanical-Turk. This sample size was chosen to match that of previous research

¹Interestingly, there is already reason to be skeptical of a perceptual interpretation of such findings – namely, that they seem to exhibit an “El Greco fallacy” (Firestone, 2013; Firestone & Scholl, 2014). If the reason conservatives rate a darker Obama image as most representative is that their political orientation makes them *see* images of him as darker, then they should also see *the images in the experiment* (i.e., the ones they are rating) as darker, and the effects should cancel out. This seems to suggest at the outset that the result is not literally perceptual. (At the same time, not every perceptual interpretation of the effect would commit this fallacy. For example, if the reason conservatives rate a darker Obama image as most representative is that the news media they consume tends to portray Obama using darker images – and so they select the image that looks most like the images they usually see – then there would be no El Greco fallacy here.)



Figure 1. The altered images used in the experiment. Each subject rated two of these four images (one from each row and one from each column) for how well they “represent” Obama (along with an unaltered image, used in Experiment 2).

(Caruso et al., 2009, Study 2). (For discussion of this subject pool’s reliability, see Crump, McDonnell, & Gureckis, 2013.)

Stimuli

Two altered versions of each of two photographs of Barack Obama were created: for each “base” photograph, one version was altered to have red horns attached to his head, and one was altered to have a yellow halo encircling his head (see Figure 1).

Procedure

Following Caruso et al. (2009), subjects read instructions explaining that photographs of politicians can differ in how well they “represent a politician”

and capture their true essence. A single image of Obama (with either horns or a halo, randomly assigned for each subject) then appeared on its own page, and (again following Caruso et al., 2009), subjects rated this image on a 7-point scale for how well it “represent[ed]” Obama (from 1 = *Not at all* to 7 = *A great deal*). The page then refreshed and displayed a new single image of Obama that differed from the first image both in terms of the underlying base photograph and in terms of how it was altered (horns or halo). (As in all such studies, viewing distances were uncontrolled, but both images were presented scaled to a width of 300 pixels.) Subjects rated this second image using the same representativeness scale. Finally, subjects were asked to what degree they identified with the political terms *Democrat*, *Liberal*, *Independent*, *Republican*, and *Conservative*, each on an 8-point scale (from 0 = *Not at all*, through 1 = *Weakly*, to 7 = *Strongly*).

Results

Conservative subjects were overwhelmingly more likely than liberal subjects to rate the “horns” image as more representative of Barack Obama. Considering those subjects who assigned different ratings to the two images (following Caruso et al., 2009), 77% of conservatives (defined here as subjects who rated themselves higher on *Republican/Conservative* than on *Democrat/Liberal*) rated the image with horns as more representative than the image with the halo, as compared to only 19% for liberals, and more generally subjects’ representativeness ratings of the images varied as a function of their personal political views, $\chi^2(2, N = 189) = 57.72, p < .001, \phi = .61$. Similarly, conservatives as a group rated the horns-image as 1.98 points more “representative” than the halo-image, while liberals rated the horns-image as 1.39 points *less* representative than the halo-image, $t(187) = 8.90, p < .001, d = 1.37$. These effects were as large or larger than the effects with skin tone ratings in previous research. For example, conservatives in our study were 4.13 times as likely as liberals to give the highest rating to the image of Obama with horns on his head, whereas previous research found that conservatives were 1.67 times as likely as liberals to give the highest representativeness rating to a darkened image of Obama ($\phi = .24$). Thus, the same methods used to conclude that conservatives see Obama as darker skinned than liberals also “support” the (clearly false) conclusion that conservatives see Obama as having bright red horns on his head.

Experiment 2: unaltered images

One difference between our Experiment 1 and Study 2 of Caruso et al. (2009) is that our experiment presented only two images, both of which were altered (i.e., to include horns or a halo), whereas Caruso et al.’s study included *three*

images, two of which were altered (i.e., to be darker or lighter) but one of which was unaltered. One possibility, then, is that by failing to include an unaltered image, our Experiment 1 forced subjects to interpret the “representativeness” question metaphorically in a way that Caruso et al. did not. (Our subjects might have truly wished to say that neither image was the “real Obama”—and give both images the lowest “representativeness” rating—but didn’t want to be noncompliant subjects by failing to vary their ratings.) For this reason, we replicated Experiment 1, except that this time a third, unaltered, image of Obama was included.

Method

This experiment was identical to Experiment 1 except that a third, unaltered, image of Obama was rated for “representativeness”, in addition to the two altered images. The unaltered image was always the same, and the three images appeared in a random order for each subject.

Results

Conservatives were again overwhelmingly more likely than liberals to rate the “horns” image as more representative of Barack Obama, 75% vs. 12%, and again subjects’ representativeness ratings of the images varied as a function of their personal political views $\chi^2(2, N = 181) = 51.04, p < .001, \phi = .59$. And again, conservatives as a group rated the horns-image as more “representative” than the halo-image (by 1.61 points) while liberals rated the horns-image *less* representative than the halo-image (by 1.88 points), $t(179) = 8.01, p < .001, d = 1.35$. Moreover, these patterns remained even considering only those subjects who (by random chance) saw the unaltered image first (and so therefore knew that there were both altered and unaltered images at the time they rated the altered images): of such subjects, 75% of conservatives (but only 14% of liberals) rated the horns image as more representative than the halo image, $\chi^2(2, N = 59) = 24.26, p < .001, \phi = .71$. Thus, the very same results hold even when, as in Caruso et al. (2009), the altered images appear in the context of an unaltered image.

Discussion

Conservatives are more likely than liberals to judge darkened photographs of Barack Obama as more “representative” (Caruso et al., 2009) or as better reflecting the “real Obama” (Kimmelmeier & Chavez, 2014). Do such findings suggest that conservatives literally see Obama as having darker skin? The results of the simple overgeneralization test suggest otherwise: conservative subjects, more so than liberal subjects, also rate an image of Obama with

horns on his head as more representative than an image with a halo, and yet clearly this shouldn't imply that conservatives see Obama as having bright red horns on his head. Similarly, then, the same judgments with lightened or darkened images cannot be used to implicate visual perception in producing such effects: to implicate perception, other methods and controls would be required beyond simple ratings. (For a broader discussion of how that can be done for top-down effects in general, see Firestone & Scholl, in press. In that paper, we identified the conflation of perception and judgment as one of six "pitfalls" that collectively compromise alleged top-down effects of cognition on perception. In our recommendations for the future, however, we only suggested that researchers pay careful attention to this distinction, without going into much detail about how that could be done. The present paper represents one example of how this distinction can be empirically investigated in practice.)

A general test?

We tried to illustrate the utility of the overgeneralization test here with a single concrete case study, but we suggest that it could be much more broadly applicable, as a valuable tool to help distinguish perception from judgment when exploring the possibility of top-down effects of cognition on perception. Indeed, we think many other reports of alleged top-down effects that rely on judgments of this sort would benefit immensely from a similar control experiment—or at least a discussion of the possibility that such ratings needn't implicate perception.

For example, it has also been reported that, when shown a continuum of yellow-orange discs overlaid on a traffic light's middle bulb, German subjects (who call that light's color *gelb* or yellow) rated more discs as yellow than did Dutch subjects (who call it *oranje* or orange; Mitterer et al., 2009). Though this was interpreted as an effect of language on color *perception*—as indicating that "the perception of color relies on actively naming these colors" (p. 1561)—it seems possible that subjects were merely following linguistic convention, using the socially appropriate color names for that context. The overgeneralization test could be used to find out: one could test for analogous cross-cultural effects in circumstances where a perceptual interpretation is not plausible. For example, since people in the Netherlands are, on average, taller than people in Germany, subjects from these countries might apply the words "tall" and "short" differently when asked to rate a continuum of differently sized people—with Dutch subjects rating as "short" someone whom German subjects might rate as "tall". Yet, it seems unlikely that Dutch subjects in this case literally *see* other people as shorter (i.e., as extending a lesser distance from the ground plane). Instead, this could be just another case where

certain words—“tall”, “short”, “orange”, “yellow”—have socially appropriate meanings that can differ contextually.

Conclusion

It may be seen as ironic that an exploration of unambiguously non-perceptual (and even a bit silly) factors (such as the horns-vs.-halos contrast) can play a key role in determining when perception itself is implicated by ratings. But keep in mind that the outcomes of an overgeneralization test could also lend support to the existence of true top-down effects of cognition on perception—by showing that the relevant effect does not in fact overgeneralize to non-perceptual cases. This test can thus be a simple but valuable tool for clarifying whether rating data in many such contexts properly implicate perception or higher-level judgment.

Acknowledgments

For helpful conversation, we thank Ben van Buren, and the members of the Yale Perception and Cognition laboratory.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Balcetis, E. (in press). Approach and avoidance as organizing structures for motivated distance perception. *Emotion Review*.
- Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as closer. *Psychological Science*, 21, 147–152.
- Branigan, A. R., Freese, J., Patir, A., McDade, T. W., Liu, K., & Kiefe, C. I. (2013). Skin color, sex, and educational attainment in the post-civil rights era. *Social Science Research*, 42, 1659–1674.
- Broudy, D. (2009). The propaganda of patriotism and color. *Synæsthesia*, 1, 1–19.
- Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, 42, 33–44.
- Caruso, E. M., Mead, N. L., & Balcetis, E. (2009). Political partisanship influences perception of biracial candidates' skin tone. *Proceedings of the National Academy of Sciences*, 106, 20168–20173.
- Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, 21, 843–860.
- Crump, M., McDonnell, J., & Gureckis, T. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8, e57410.
- Firestone, C. (2013). On the origin and status of the “El Greco fallacy”. *Perception*, 42, 672–674.

- Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, 25, 38–46.
- Firestone, C., & Scholl, B. J. (2015). Can you experience "top-down" effects on perception? The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*, 22, 694–700.
- Firestone, C., & Scholl, B. J. (target article in press). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*.
- Kemmelmeier, M., & Chavez, H. L. (2014). Biases in the perception of Barack Obama's skin tone. *Analyses of Social Issues and Public Policy*, 14, 137–161.
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, 135, 501–512.
- Mitterer, H., Horschig, J. M., Müsseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1557–1562.
- Philbeck, J. W., & Witt, J. K. (2015). Action-specific influences on perception and post-perceptual processes: Present controversies and future directions. *Psychological Bulletin*, 141, 1120–1144.
- Witt, J. K., Proffitt, D. R., & Epstein, W. (2010). When and how are spatial perceptions scaled? *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1153–1160.