

The Face-Race Lightness Illusion Is Not Driven by Low-level Stimulus Properties: An Empirical Reply to Firestone and Scholl (2014)

Lewis J. Baker¹ · Daniel T. Levin¹

© Psychonomic Society, Inc. 2016

Abstract Levin and Banaji (*Journal of Experimental Psychology: General*, 135, 501–512, 2006) reported a lightness illusion in which participants appeared to perceive Black faces to be darker than White faces, even though the faces were matched for overall brightness and contrast. Recently, this finding was challenged by Firestone and Scholl (*Psychonomic Bulletin and Review*, 2014), who argued that the nominal illusion remained even when the faces were blurred so as to make their race undetectable, and concluded that uncontrolled perceptual differences between the stimulus faces drove at least some observations of the original distortion effect. In this paper we report that measures of race perception used by Firestone and Scholl were insufficiently sensitive. We demonstrate that a forced choice race-identification task not only reveals that participants could detect the race of the blurred faces but also that participants' lightness judgments often aligned with their assignment of race.

Keywords Lightness perception · Race perception · Top-Down effects

The perception of surface reflectance is a classic case where psychological experience does not map directly on to the actual amount of light striking the retina. Instead, our visual system can infer equal reflectance of surfaces despite variations in the physical amount of light falling on each of those

surfaces. An extensive tradition of research has assessed how early and midlevel vision accounts for complex variations in the arrangement of visual surfaces to derive relatively accurate perceptions of surface reflectance (for review, see Adelson, 1993, 2000; Gilchrist & Annan, 2002).

Several reports have extended analysis of lightness perception from general properties of illumination and surfaces to a perceiver's specific knowledge of object categories. Using a rating scale anchored by labels ranging from *light* to *dark*, MacLin and Malpass (2003) found that participants rated Black faces as being darker than White faces. Levin and Banaji (2006) further demonstrated that participants appeared to perceive Black faces as being darker than luminance- and contrast-matched White faces. This effect was demonstrated in several experiments using psychophysical paradigms in which participants adjusted the luminance of one face to an opposite-race reference face (Experiment 1); adjusted a neutral gray patch to match both unambiguous Black/White faces and racially ambiguous faces that were labeled Black for some participants and White for others (Experiment 2); adjusted a neutral patch to match Black/White line-drawn different-race faces with equiluminant gray filler (Experiment 3); and classified faces in a same/different lightness task in which faces deviated by incremental degrees of luminance (Experiment 4). Levin and Banaji argued that the category of a face can affect the perceived lightness of the face and that this effect was not attributable to lower level stimulus confounds. Thus, the finding would appear to be a top-down effect on vision, or, perhaps more controversially, a demonstration that a perceptual module is penetrable. They argued that the effect was based on social-categorical influence, and that the effect might rest on between-feature correlations between specific face forms and specific levels of luminance.

Firestone and Scholl (2014) cited the Levin and Banaji stimuli as an important, *experienceable* demonstration of top-down effects on perception. In two experiments, they

✉ Lewis J. Baker
lewis.j.baker@vanderbilt.edu

✉ Daniel T. Levin
daniel.t.levin@vanderbilt.edu

¹ Department of Psychology and Human Development, Peabody College, Vanderbilt University, #512, 230 Appleton Place, Nashville, TN 37203-5701, USA

argued that at least some parts of the Levin and Banaji (2006) effect were attributable to low-level stimulus confounds because the brightness distortion appeared even when the faces were blurred to eliminate the perceptibility of race. Participants who were apparently unaware of the race of blurred faces still judged the Black face to be the darker of the two. It is important to note that Firestone and Scholl acknowledged that their critique did not extend to Levin and Banaji's experiment where participants reported that racially ambiguous face appeared darker when preceded by an unambiguous White face versus an unambiguous Black face. However, they argued that these other experiments could derive from other experimental confounds, such as instruction effects or bias formed after multiple trials.

Here, we present a replication and reply to Firestone and Scholl in three experiments demonstrating that a more sensitive forced-choice measure of race classification reveals the presence of detectable race-specifying information in the blurred stimuli. It is possible that subtle racially identifying information slipped past Firestone and Scholl's questioning and that participants avoided reporting their socially undesirable detection of race in the stimuli. We present data demonstrating that these questions may have misclassified many participants as unable to detect race when they were, in fact, able to do so. This appears to be consistent with Firestone and Scholl's own data. Although it was relegated to a footnote, when asked to select from four races, 72.6% of participants chose "White" for the White face while only 45.2% of participants chose "White" for the Black face.

In addition to testing for detectable race information, we took seriously the possibility that the face stimuli might have contained artifacts, and we noted that both the original Levin and Banaji (2006) stimuli and blurred versions created by Firestone and Scholl had small compression artifacts that appeared as differences in gray-scale values visible in blank areas and in the forehead regions that should have been identical (see Fig. 1). In Experiment 2, we therefore added a new set of faces that were developed using lossless image formats for a recent cross-cultural project, which did not show compression artifacts during image analysis. The new faces produced a brightness distortion of magnitude very similar to that observed in the original Levin and Banaji experiments. In a related research project, we used these faces with 24 Vanderbilt University students who adjusted either a rectangular gray patch or a patch of forehead identical to both foreheads in the faces, and significant effects were observed in all conditions, with effect sizes (Exp. 1: $d=1.32$, Exp. 2: $d=1.08$) comparable to the original Levin and Banaji findings (Exp. 1: $d=.75$, Exp. 2: $d=1.26$).

Experiment 1

Experiment 1 directly replicated the design of Firestone and Scholl (2014), using their stimuli. As did Firestone and Scholl, we used Amazon Mechanical Turk to recruit participants to complete a single trial experiment in which they viewed a pair of blurred faces and judged how light or dark they were, and then indicated whether they could identify the race of the faces.

Method

Participants

Volunteers ($N=101$) completed the study via Amazon Mechanical Turk. Mechanical Turk is a website for outsourcing small tasks to large groups of people and has been empirically validated for perceptual experiments (Germine et al., 2012; Paolacci, Chandler, & Ipeirotis 2010). Participants were paid \$0.35 and took an average of 4.25 minutes to complete the experiment. Recruitment was restricted to volunteers who had completed at least 100 assignments with >95% approval.

Stimuli

Images were received via e-mail from Firestone and Scholl. Firestone and Scholl converted the original Levin and Banaji (2006) faces to JPEG and then to PDF. They were blurred using the Glass filter in Adobe Illustrator and exported as PNG files. For ease of presentation, we placed pairs of images horizontally in PNG files (402×252 pixels), with the nearest features of the distorted faces separated by 100 pixels.

Procedure

Upon giving consent, participants viewed a pair of blurred faces, one White and the other Black. Images were counterbalanced so that half of participants viewed the Black face on the left and vice versa. Participants answered four randomly ordered two-alternative forced-choice questions, asking which image was lighter, larger, blurrier, and more aesthetically pleasing. To avoid response bias, participants were randomly asked either "Which image is lighter" or "Which image is darker." On the next page, participants were asked to describe the images they had seen in two sentences. On the page after, participants were explicitly told that the images were faces and were asked to provide a two-sentence description to help someone "pick them out of a large set of many other faces." Following this, participants viewed the faces again and rated them as the same or different on six categories in randomized order: race, age, gender, shape,

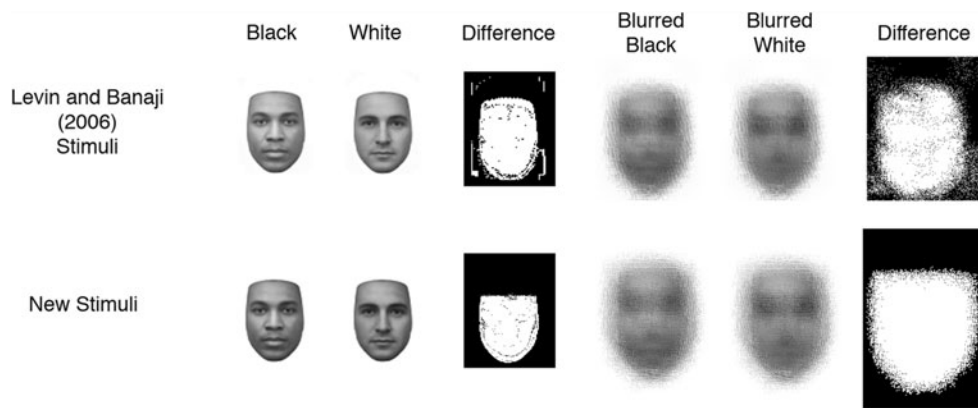


Fig. 1 The first row contains the original Levin and Banaji (2006) faces, and blurred versions created by Firestone and Scholl (2014) along with difference images contrasting the original and blurred faces. Difference images are thresholded such that pixels that were identical

between the originals are black and those that were not identical are white. The second row shows the new stimuli that recreated the originals using all lossless compression

facial expression, and facial hair. Finally, participants wrote their best guess as to the purpose of the study and responded whether they had ever participated in a similar study.

Results

We applied exclusion criteria similar to those employed by Firestone and Scholl to isolate a subset of participants who were nominally unable to detect the race of the stimuli. Only completed surveys from unique IP addresses were analyzed. Exclusion criteria included (1) identifying the faces as different races in open-ended descriptions ($n=0$) or when (2) giving details for a “line up” ($n=0$); (3) judging the faces to be of different races ($n=14$); (4) identifying race as the feature of interest for the study ($n=6$); or (5) claiming participation in a similar study design ($n=7$). Combined, 26 participants met at least one of the exclusion criteria, leaving a total of 75 participants for analysis.

A binomial test revealed that significantly more participants responded that the White face was lighter or the Black face was darker (64%; 50/70, $p=.005$). As described above, 14 participants were excluded solely for reporting that the faces were of different race. Repeating the analysis with the 13 participants who reported the faces to be different races but were not excluded by any other criteria also showed above-chance lightness judgments (68%; 60/88, $p<.001$).

Discussion

Experiment 1 replicated Firestone and Scholl’s finding. Despite their apparent inability to extract information about race categories, participants rated White faces as lighter or Black faces as darker. It is therefore possible that some basic

perceptual feature of these stimuli survived distortion, leading participants to judge the White face as lighter even when they could not detect the race of the faces, even though the original stimuli were, overall, isoluminant. However, we did observe that a substantial minority of participants (33%) judged the faces to be different races.

Experiment 2a

It is possible that participants were able to perceive subtle signs of race in the stimuli that in turn may have induced a subtle brightness distortion. Firestone and Scholl (2014) employed a forced-choice measure to assess lightness but a same–different measure to assess racial categorization. We added a forced-choice measure of racial categorization to Experiment 2 in case the same–different question underestimated the discrimination of race. In addition, the original Levin and Banaji (2006) faces contained image noise that may have been exacerbated by JPEG compression (see Fig. 1). Although we recognize that there are a number of potential uncontrollable variables inherent in crowdsourcing research, Experiment 2 also included stimuli using lossless compression. If image noise or other properties of the original Firestone and Scholl (“old”) faces induced a stimulus confound, then this should be absent in the new faces. If the availability of subtle race-specifying information was missed in Experiment 1, then it may appear in our forced-choice measure of race classification, and, in cases where above-chance race classification does appear, conditionalizing data on these responses may reveal a categorical brightness illusion.

Method

Participants

Participants ($N=242$) completed the experiment using Mechanical Turk. Participants were randomly assigned to see the old stimuli ($n=115$) or the new stimuli ($n=127$). Recruitment procedures were identical to those in Experiment 1.

Stimuli

As described in the introduction, we used a new set of stimuli similar to those in Levin and Banaji (2006) with lossless picture formats that eliminated most of the image noise in the original stimuli. The PNG files were distorted using the Glass filter in Photoshop before being exported as PNG files, thereby skipping conversion to JPEG and PDF. The resulting image subtraction revealed fewer differences in luminance between stimuli (see Fig. 1) and, most importantly, no between-image differences in backgrounds or in the forehead region, which were designed to be identical between the Black and White (see Levin & Banaji, 2006). These images were sized and paired in a single image file (588×400 pixels with 100 pixels between faces). A second condition used the same stimuli as Experiment 1 (the “old” faces) in the same size image file.

Procedure

The procedure was similar to Experiment 1 with two exceptions. Participants were randomly assigned to view either the old faces or the new faces. Also, participants viewed the faces for a third and final time at the end of the experiment and were asked, “Using your best guess, how would you differentiate these faces by race?” Participants chose from a randomly sorted drop-down menu, “Left face is Black—Right face is White” or “Left face is White—Right face is Black.”

Results

We applied two exclusion criteria to better test for an effect of racial judgments on brightness perception. The first criteria excluded participants who explicitly differentiated faces by race in open-ended description ($n=0$) or description for selection from a line-up ($n=2$), or claiming participation in a similar study design ($n=4$). The second criteria mirrored the exclusion principles of Firestone and Scholl, excluding participants who also judged the faces to be of different races ($n=43$). We did not exclude participants who identified race as the basis of the study, because the added final question likely suggested the

premise. Altogether, six participants were excluded for at least one of the first criteria, and 46 were excluded for any criteria.

This experiment did not replicate Firestone and Scholl’s basic finding. Binomial tests revealed that participants did not differ from chance in their assignment of lightness to faces in both the old stimuli, first exclusion criterion: 49.5% (55/111), $p>.99$; second exclusion criterion: 48.4% (44/91), $p=.83$, and the new stimuli, first criterion: 45.2% (56/125), $p=.33$; second criterion: 46.2% (48/104), $p=.34$. There were no differences in lightness judgment between conditions, first criterion: $\chi^2(1, n=237)=.440$, $p=.501$; second criterion: $\chi^2(1, n=195)=0.094$, $p=.759$.

As in previous experiments, most participants in both conditions rated that the faces were of the same race after applying the first exclusion criteria (old: 82% same race, new: 83% same race). This did not statistically differ by condition, $\chi^2(1, n=236)=1.424$, $p=.233$. However, when forced to assign race to the faces, 80.2% of participants viewing the old stimuli and 73.6% of participants viewing the new stimuli correctly differentiated race. This also did not differ between conditions, $\chi^2(1, n=236)=0.1875$, $p=.665$. This implies that some information about race survived the blurring procedure.

If participants can correctly identify race regardless of stimulus degradation, are their judgments reflected in their assessment of lightness? As visualized in Fig. 2, we found that participants averaged across both conditions were significantly more likely to rate the Black face as darker if they correctly differentiated race, correct differentiation: 59.1% (107/181) darker, incorrect: 32.7% (18/55) darker, $\chi^2(1, n=236)=11.791$, $p<.001$. This effect was significant for the new stimuli, correct: 65.2% (60/92), incorrect: 27.3% (9/33), $\chi^2(1, n=125)=14.141$, $p<.001$, but not for the old stimuli, correct: 52.8% (47/89), incorrect: 40.9% (9/22), $\chi^2(1, n=106)=0.999$, $p=.318$. Note that only 55 of the 236 participants across both conditions judged race incorrectly, thus

Experiment 2 Lightness by Forced Choice Race

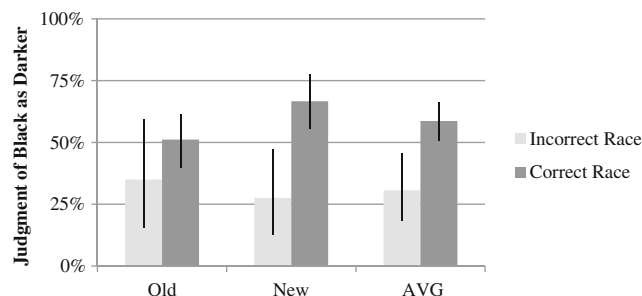


Fig. 2 Overall, participants who incorrectly categorized the faces by race tended to perceive the Black face as lighter or vice versa. Error bars denote Clopper-Pearson 95% binomial confidence intervals

limiting comparisons between groups. The proportion of individuals who rated the Black face as darker when correctly judging race was marginally different across conditions (old = 52.8%, new = 65.2%), $\chi^2(1, n = 181) = 2.882, p = .090$.

Experiment 1b

One other difference existed between the stimuli of Experiments 1 and 2: The images in Experiment 2 were 46% larger than the images in Experiment 1. Although this was done to better position stimuli on the online survey, it is possible that enlarging the stimuli somehow affected judgments of lightness. We therefore conducted a replication of Experiment 2a using identical stimulus sizes to Experiment 1 and Firestone and Scholl (402 × 252 pixels with 100 pixels between faces). The procedure and exclusion criteria otherwise matched Experiment 2a.

Participants did not judge the Black face to be darker than chance in either condition using Firestone and Scholl's (2014) exclusion criteria, binomial tests; old: 54.2% (38/45), $p = .51$; new: 44.7% (42/52), $p = .35$, nor did participants differ between conditions, $\chi^2(1, n = 177) = 0.0216, p = .8831$. If we included participants who judged faces to be of different races, significantly more participants in the old condition judged the White face to be lighter or the Black face to be darker, binomial test: 61.7% (71/115), $p = .015$. Meanwhile, there was no significant difference in lightness judgments for participants in the new condition, binomial test: 50.0% (57/114), $p = 1$. This difference was marginally significant, $\chi^2(1, n = 229) = 3.2002, p = .073$.

Participants dissociated faces by race better than chance, old: 85.2% (98/115); new: 77.2% (88/114); binomial test: $ps < .001$. Assignment of race did not statistically differ between conditions, $\chi^2(1, n = 229) = 2.417, p = .120$. Pooling across conditions, participants were again more likely to rate the Black face as darker when they correctly differentiated race, correct differentiation: 61.8% (115/186), incorrect differentiation: 30.2% (13/43), $\chi^2(1, n = 229) = 14.143, p < .001$. Paired comparisons indicated an effect of racial categorization on lightness perception for the old condition, correct: 68.4% (67/98), incorrect = 23.5% (4/17), $\chi^2(1, N = 115) = 12.330, p < .001$, and a marginal effect in the new condition, correct: 54.5% (48/88), incorrect = 34.6% (9/26), $\chi^2(1, N = 114) = 3.189, p = .074$. Differences between conditions approached significance, $\chi^2(1, n = 186) = 3.753, p = .053$.

Discussion

Experiments 2a and 2b revealed no consistent tendency to judge one of the faces as lighter/darker than the other when

using the exclusion criteria applied in Firestone and Scholl. This was true for both the old stimuli of Experiment 1 and the new stimuli. In fact, combining data across Experiments 1, 2a, and 2b demonstrates that participants rated the lightness of nearly at chance (52.6% indicating Black was darker across 481 participants using Criterion 1, and 50.6% indicating Black was darker across 407 participants using Criterion 2). Thus, the tendency to label the Black face as darker reflects an inconsistent conclusion about a very subtle, near-threshold perceptual experience.

In addition, participants were significantly better than chance at differentiating race in forced-choice trials. Of particular note is the replication of the lightness illusion using blurry stimuli, revealing increased likelihood of judgments that the Black face was darker in participants who correctly classified race in a forced-choice trial. It is important to note that 54% of individuals in Experiments 2a and 2b perceived the Black face to be darker, but 81% judged the Black face to be racially Black in the forced-choice trial. Therefore, although there was a significant relationship between correct forced-choice attribution of race and lightness perception, it is possible that extremely subtle and inconsistently perceivable race information falls below a threshold necessary to influence lightness. Alternatively, the forced-choice measure of race may have been more sensitive to race-specifying information in the stimuli than was the lightness judgment. This may have occurred because the forced-choice measure was necessarily taken at the end of the experiment, when participants were more familiar with the faces.

Experiment 3

Firestone and Scholl (2014) contended that the Levin and Banaji (2006) faces contained sufficient differences in low-level features or contrast to give a perception of lightness that was independent of perceived race. Experiment 2 demonstrated that participants could detect the race of the blurred faces; however, it is possible that participants attributed race according to the perceived lightness of the faces rather than the form of the faces. Experiment 3 therefore presented participants with luminance-inverted photonegative versions of the blurred Firestone and Scholl stimuli. If participants in Experiment 2 had successfully detected race merely because low-level perceptual confounds made the White face look lighter for reasons independent of race-specifying form features that remained apparent in blurred faces, then photonegative versions of the blurred faces should invert this effect. If participants correctly categorize race in photonegative stimuli, then they cannot be making judgments based on the inference that faces characterized confound-induced lightness must be White.

Method

Participants

Participants ($N=154$) completed the experiment using Mechanical Turk. Recruitment procedures were identical to those in Experiment 1.

Stimuli and Procedure

To best compare to their findings, only the “old” Firestone and Scholl stimuli were used. The luminance of the faces was inverted using Photoshop CS6’s invert function, which simply subtracts all 8-bit grayscale values from 256 (see Fig. 3). The stimuli and procedure were otherwise identical to those in Experiment 2b.

Results

We again applied two exclusion criteria to better test for an effect of racial judgments on brightness perception. We excluded participants who explicitly distinguished faces by race or had participated in similar experiments (six excluded; $N=148$).

Participants did not judge the Black face to be darker or lighter than chance, suggesting the Black face was not physically lighter after inverting the luminance, binomial test: 47.4% (73/154) lighter; $p=.573$. Most participants reported that the faces were of the same race, consistent with Firestone and Scholl and Experiment 1, 83.1% (128/154) same race; binomial test, $p<.001$. However, participants were again significantly better than chance when forced to assign race to the faces, 74.7% (115/154) correct identification; binomial test: $p<.001$, and their level of accuracy was very similar to that observed in Experiments 2a and 2b for the original stimuli (73.6% in Experiment 2a, and 85.4% in Experiment 2b). Differences in lightness judgments

based on forced-choice racial identification were nonsignificant, correct race: 54.8% (63/115) reported Black as darker; incorrect race: 46.2% (18/39) reported Black as darker; $\chi^2(1, N=154)=0.870, p=.351$.

Discussion

Participant’s ability to identify race in the blurred faces is not an inference based on lightness. Participants were successful in detecting the race of the original blurred stimuli even when luminance was inverted. If participants in Experiment 2 inferred that the darker face was Black, then the luminance inversion should have reversed this judgment. We observed neither of these results. Participants did not significantly judge one face to be lighter than the other, and they again correctly identified race in the forced-choice judgment. This implies that the controls implemented by Firestone and Scholl were not sufficient to eliminate race in the stimuli. We should also note that in contrast to Experiments 2a and 2b, the relationship between correctly differentiating the race of the face and choosing the White face to be lighter was nonsignificant. It is, however, noteworthy that less than 27% participants incorrectly assigned race to the stimuli.

General Discussion

We responded to Firestone and Scholl (2014) in three experiments. In Experiment 1, we replicated the apparent confound observed in four experiments Firestone and Scholl, and then in Experiments 2 and 3 we observed and replicated three important phenomena. First, even with Firestone and Scholl’s stimuli, participants did not consistently judge the blurred Black face to be darker than the blurred White face: the judgment bias was absent in Experiment 2a, albeit with larger stimuli, and again in 2b when using Firestone and Scholl’s exact images and exclusion criteria. Furthermore, it was absent in stimuli created to minimize artifacts, and it was absent in Experiment 2 when using photonegative images. Thus, we conclude that any apparent brightness difference between the blurred stimuli is at best a near-threshold difference that can be observed only inconsistently. Second, any effects that might be observed can reasonably be explained by subtle signs of race that survived the blurring process. We consistently observed that participants could correctly identify the race of the stimuli across two sets of stimuli and three experiments using both normal and luminance-inverted faces. Finally, participants who correctly differentiated race were in at least some cases more likely to judge the Black face to be darker.

On the whole, we find Firestone and Scholl’s reevaluation of top-down effects to be commendable, but we would like to point out that their focus on “experience” means that a

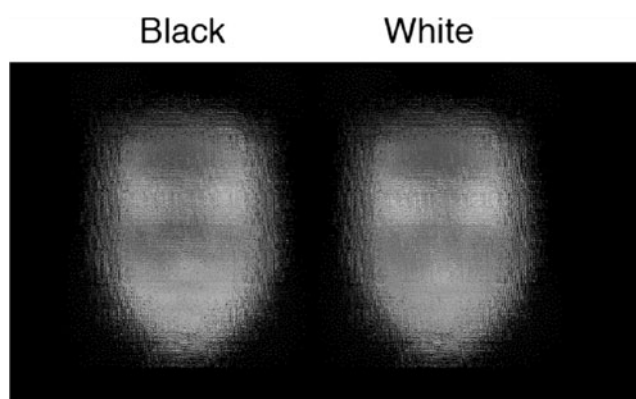


Fig. 3 Luminance-inverted photonegative stimuli used in Experiment 3

subjective and shifting criterion can be applied to selectively diminish the importance of experiments supporting a top-down hypothesis. For instance, the effect of racial categories on lightness judgments does not necessarily reflect the impact of transitory, abstract, online cognitions shaping perception independent of visual form-specifying information. More likely, these effects stem from aggregation of perceptual links between form and lightness that can be swayed by context—for instance, by interpreting a racially ambiguous face as Black or White in comparison to a different stimulus (similar to how ambiguous 3-D illusions shift depending on interpretation of contours; Adelson, 2000). We do not believe that one's perception of an unambiguously Black face can change simply by imagining the face is categorically White. This argument is clear in the original Levin and Banaji (2006) article and agrees with a large body of research. Researchers have demonstrated that experience plays a critical role in the object perception (Gauthier et al., 1999), through either abstract semantic knowledge or through midlevel visual influence (Bannert & Bartels, 2013; Vetter & Newen, 2014). Repeated experience linking the form of faces to their reflectance could lead to associations contained within midlevel vision. Just as top-down knowledge can alter perception of bistable illusions, so too could people's interpretation of race in ambiguous faces bias judgments of lightness, but knowledge cannot override a completely unambiguous perceptual input. We do not find the hypothesis that experience informs perception to be an extraordinary claim. However, we encourage continued exploration of this meaningful debate.

Acknowledgments This material is based upon work supported by NSF Graduate Research Fellowship #2013139545 to L. J. B.

References

- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, 262, 2042–2044.
- Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.
- Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology*, 23(22), 2268–2272.
- Firestone, C., & Scholl, B. J. (2014). Can you experience 'top-down' effects on perception? The case of race categories and perceived lightness. *Psychonomic Bulletin and Review*. doi:10.3758/s13423-014-0711-5
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6), 568–573.
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, 19(5), 847–857.
- Gilchrist, A. L., & Annan, V. (2002). Articulation effects in lightness: Historical background and theoretical implications. *Perception*, 31, 141–150.
- 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.
- MacLin, O. H., & Malpass, R. S. (2003). The ambiguous-race face illusion. *Perception*, 32, 249–252.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411–419.
- Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, 27, 62–75.