

BIAS IN THE FLESH

SKIN COMPLEXION AND STEREOTYPE CONSISTENCY IN POLITICAL CAMPAIGNS

SOLOMON MESSING*

MARIA JABON

ETHAN PLAUT

Abstract There is strong evidence linking skin complexion to negative stereotypes and adverse real-world outcomes. We extend these findings to political ad campaigns, in which skin complexion can be easily manipulated in ways that are difficult to detect. Devising a method to measure how dark a candidate appears in an image, this paper examines how complexion varied with ad content during the 2008 presidential election campaign (study 1). Findings show that darker images were more frequent in negative ads—especially those linking Obama to crime—which aired more frequently as Election Day approached. We then conduct an experiment to document how these darker images can activate stereotypes, and show that a subtle darkness manipulation is sufficient to activate the most negative stereotypes about Blacks—even when the candidate is a famous counter-stereotypical exemplar—Barack Obama (study 2). Further evidence of an evaluative penalty for darker skin comes from an observational study measuring affective responses to depictions of Obama with varying skin complexion, presented via the Affect Misattribution Procedure in the 2008 American National Election Study (study 3). This study demonstrates that darker images are used in a way that complements ad content, and shows that doing so can

SOLOMON MESSING is currently Director of Data Labs at the Pew Research Center, Washington, DC, USA, though the research was conducted while he was a Ph.D. candidate at Stanford University, before he joined the Pew Research Center. MARIA JABON is a software engineer at LinkedIn, Mountain View, CA, USA. ETHAN PLAUT is a fellow at Stanford University, Stanford, CA, USA. For helpful feedback and comments, the authors thank Jeremy Bailenson, Matthew DeBell, Justin Grimmer, Shanto Iyengar, Simon Jackman, Jonathan Wand, John Walker, Sean J. Westwood, and *POQ*'s editors, as well as the anonymous reviewers who provided valuable comments and suggestions. Thanks also to Taylor Young, Truc Nguyen, Alicia Solow-Niederman, Kimberly Rosenblum, and many Mechanical Turk workers who helped code advertisements. Functions to extract image data are available in the R package ImageMetrics. This work was generously supported by a Google grant [1139918-1-GBNKGW to S.M.]. All errors are the authors' alone. *Address correspondence to Solomon Messing, Pew Research Center, 1615L St. NW, Suite 800, Washington, DC 20036, USA; e-mail: smessing@pewresearch.org.

negatively affect how individuals evaluate candidates and think about politics.

Introduction

At the height of the 2008 primary season, the Hillary Clinton campaign aired an ad depicting Barack Obama with darkened skin and wider facial features than in the original footage. Political blogs leveled accusations of stereotype-consistent bias and discrimination, setting off a briefly lived scandal (Troutnut 2008). Of course, the darker portrayal of Obama's complexion and wider aspect ratio may very well have been a product of uploading the ad to the web, as Factcheck.org pointed out (Kolawole and Bank 2008). Nonetheless, the episode underscores a question of interest to those who study campaigns: How do stereotype-consistent portrayals of candidates affect how they are perceived by the voting public?

The fact that stereotypes have played a part in past political campaign narratives (e.g., the Bush Campaign's "Willie Horton" ad; see Jamieson [1993]) gives us reason to suspect that we might see direct manifestations of stereotype consistency in the context of the first general election campaign involving a Black candidate. At the same time, the strong link between skin complexion and stereotype activation found in the psychological literature (e.g., Blair et al. 2002; Maddox and Gray 2002) suggests that such portrayals may affect how voters respond to candidates. Indeed, scholars have shown that images can activate stereotypes relevant to electoral outcomes. For example, showing images of Black males with darker skin complexion makes negative stereotypes about Blacks more salient (Maddox and Gray 2002), and indeed, a series of experimental studies found that early in the 2008 campaign, viewing political advertisements with darker images of Obama negatively impacted respondents' preference for Obama as a presidential candidate (Iyengar et al. 2010), replicating past findings with hypothetical candidates (Terkildsen 1993). Additionally, Mendelberg (2001) showed that pairing a picture of a Black Willie Horton with the issue of crime in the presidential campaign of 1988 primed racial considerations in candidate evaluations and policy opinions, and Valentino, Hutchings, and White (2002) showed that visual references to Blacks significantly affect vote choice.

This work first outlines an original method to examine skin complexion and utilizes it to document how the complexion of presidential candidates in campaign advertisements varies with content and over time. Specifically, we measure how dark a candidate's skin appears in an image and interrogate how skin complexion fits into broader contextual schema conveyed in advertisements consistent with widely held stereotypes about Blacks. Study 2, which uses a word-completion task administered to respondents after viewing a light

or dark image of Barack Obama, demonstrates that darkened images do in fact increase the salience of negative stereotypes about Blacks. Finally, using data from the 2008 American National Election Study (ANES), which tested affective responses to photographs of each candidate using the Affect Misattribution Procedure (AMP), study 3 examines the public's affective responses to images of Obama with varying levels of skin-tone darkness.

We find that darker images were more frequent in negative ads—especially in those linking Obama to crime—which aired more frequently as Election Day approached. Further, our subtle darkness manipulation is sufficient to activate the most negative stereotypes about Blacks—even when the candidate is as famous and counter-stereotypical as Barack Obama. Regardless of intentionality, the 2008 campaign against Obama utilized message-consistent images that primed negative racial attitudes about Blacks in ads that were most likely to air close to Election Day. Our findings demonstrate that presenting an image of a Black candidate with darker complexion can shape how individuals respond to political advertisements and think about politics.

Skin Complexion and Stereotypes

People use physical features to gain access to a rich source of heuristic information about others based on stereotypes (Ashmore and Del-Boca 1979; Brewer 1988; Fiske and Neuberg 1990; Bodenhausen and Macrae 1998). People's physical features tell us about the categories to which they may belong, causing us to associate character traits with that person. These associations can become salient so quickly and automatically that we remain unaware of the process (Klatzky, Martin, and Kane 1982; Bargh, Chen, and Burrows 1996; Spencer et al. 1998). Indeed, there is evidence that our perceptual systems are biased to produce independent person and group representations of others (Zarate et al. 2008), and that outgroup members are more likely to be processed under the system that assigns group representations. Furthermore, these stereotypes may diverge from conscious attitudes, especially around socially sensitive topics (Hofmann et al. 2005).

Because 2008 saw the first ever successful Black candidate for president, we care especially about stereotypes related to skin complexion and race. The literature on minority groups and complexion generally finds that phenotypical features can activate minority stereotypes, especially in the case of darker skin. Maddox and Gray (2002) found that the participants listed significantly more stereotypical traits for darker-skinned Blacks than those with lighter skin and that stereotypical traits were nearly all negative (e.g., dirty, lazy, uneducated). In a series of studies, Blair et al. (2002) found that people with more Afrocentric features were more often attributed with negative Black stereotypical traits, regardless of their actual racial group membership—Whites with more Afrocentric features were also judged as more likely to have attributes

stereotypical of Blacks. Another study, [Livingston \(2002\)](#), found that Whites maintain more negative associations with Blacks than ingroup members, and this was particularly so for Blacks with more prototypical features, including darker skin. [Ronquillo et al. \(2007\)](#) showed that among Whites, the part of the brain associated with fear conditioning, the amygdala, shows greater activation in response to pictures of darker-skinned Blacks than lighter-skinned Blacks or Whites.

The evidence of the consequences of biases related to skin complexion is startling. For example, in the nation's courts, Black criminal defendants with more stereotypical facial features, including dark skin, were more likely to receive the death penalty ([Eberhardt et al. 2006](#)). College students and police officers were found to implicitly associate criminality more with Blacks than with Whites, and these associations were stronger for Blacks with darker skin and a more prototypical appearance ([Eberhardt et al. 2004](#)). Black perpetrators and their victims were also more memorable and produced the highest emotional concern among White subjects when the offender was darker skinned ([Dixon and Maddox 2005](#)). Even Black first graders were better able to remember stories that portrayed dark-skinned characters negatively and light-skinned characters positively ([Averhart and Bigler 1997](#)).

Despite the extensive work in psychology on stereotyping, only a handful of studies have investigated racial bias related to skin complexion in political settings. Early studies suggested that Whites are significantly less likely to vote for a Black candidate with darker skin tone ([Terkildsen 1993](#)). In more recent work, [Weaver \(2012\)](#) shows that voters generally prefer lighter-skinned candidates when given a choice between two Black candidates. After the 2008 election, a series of experimental studies found that viewing political advertisements with darker images of Obama had a negative impact on respondents' preference for Obama as a presidential candidate during the early stages of the campaign ([Iyengar et al. 2010](#)).

These findings suggest that ads that portray Black candidates with a darker complexion might prime negative stereotypes about Blacks, damaging the candidate's election prospects in a way that has nothing to do with political fitness for office and is difficult to detect. Indeed, many images from the 2008 presidential campaign appear to have been manipulated and/or selected in a way that produces a darker complexion for Obama—examples can be seen in the supplementary data online. However, we should not be terribly concerned about a few isolated dark images. Rather, we would need to see evidence that any image manipulation and/or selection was systematic. In fact, what would be most concerning is to find images of Obama with darker skin complexion in attack advertisements that seek to portray Obama according to stereotype-consistent narratives. Hence, we characterize skin complexion in advertisements with particular attention to how complexion varies with content, then show how manipulating skin complexion matters using an experiment and an observational analysis.

Study 1: Skin Tone and Visual Cues in Attack Advertisements

This first study examines skin complexion in actual campaign advertisements. We outline a method to quantify skin complexion in ads and utilize it to document how skin complexion varies with content, consistent with negative Black stereotypes suggested in the literature above, most notably content related to criminality. We also examine how skin complexion varies over the course of the campaign to interrogate whether stereotype-consistent depictions increase as the campaign develops and grows increasingly negative.

The data consist of an archive compiled by the Political Communication Lab (PCL) at Stanford University, and consisted of 126 English-language video ads produced by the Obama and McCain campaigns between July 1 and November 2, 2008. The data were obtained by monitoring candidate websites and YouTube channels throughout this period, gathering all ads directly posted by each campaign, in an attempt to collect a census of such ads. The *National Journal* website was also monitored, along with a variety of other news media sources, for references to other advertisements that were not posted, and web searches were performed in order to track down such ads (though this was rare). The vast majority of depictions of presidential candidates in ads consisted of still images, not video. Accordingly, we analyze still images, and on occasion, video still-image captures. In the sample of ads under investigation, there were 534 still images, 259 of Obama and 275 of McCain.¹

The aesthetic skin property of interest—darkness—corresponds to the value (brightness) measure that comprises one dimension of the HSV (hue, saturation, value) colorspace. An image of a completely white square has a value (V) equal to 1, while an image of a completely dark, black square has V equal to 0. Saturation corresponds to the presence or absence of color in an image, so, for example, a black-and-white image would have color saturation of 0, while a full-color image would have a saturation of roughly 1 (see [figure 1](#)). Hue corresponds to perceived location on the color spectrum. These measures should be interpreted as physical quantities, not as aesthetic properties. Brightness and color saturation combine with other elements in an image such as contrast, background, shadow, light diffusion, and other subjective elements in complicated ways to affect how humans perceive images. We utilize these metrics only as *indicators*, not measures, of stereotype consistency.

Neither hue (H), brightness (V), nor saturation (S) should be considered quantifications of overall image quality. Measurements of brightness, saturation, and hue spectrum location are not indicators of subjective aesthetics of an image, which are vastly more complicated than anything that can be described along these three dimensions.

1. We do not exclude any images from our analysis.

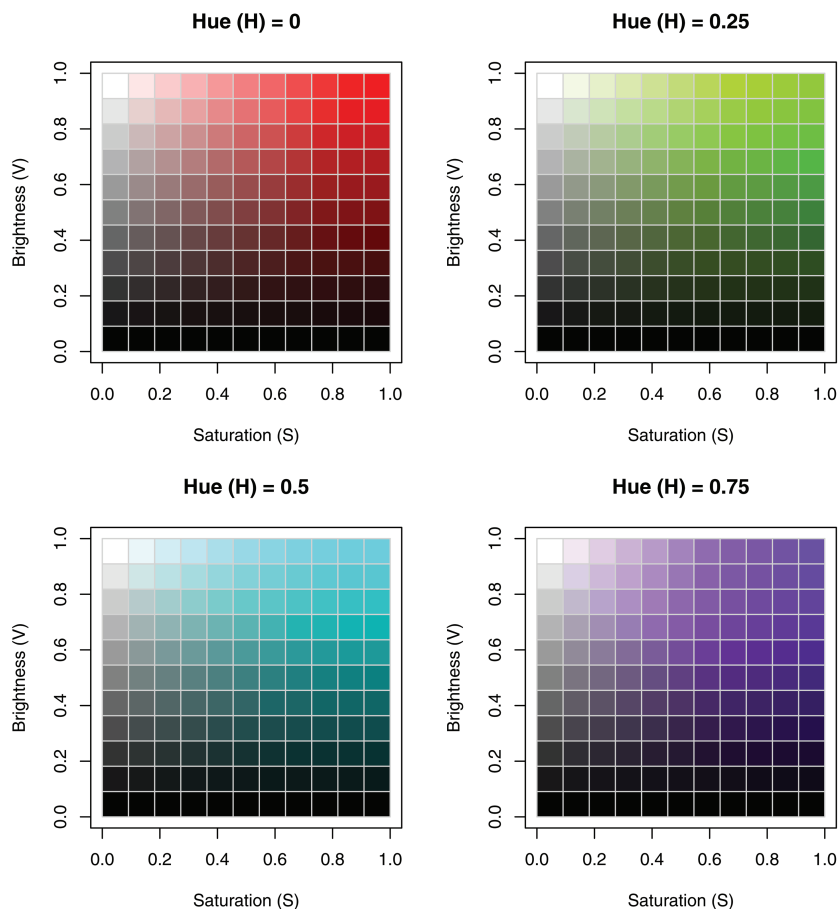


Figure 1. The Three-Dimensional HSV Color Space. Hue (H) captures the actual color, saturation (S) captures the amount of color present, and value (V) captures the relative lightness or darkness.

We extracted the relevant metrics from images by using commonly available open source tools, including Bio7, R, ImageJ, and EBIImage, to set up an interface that allows analysts to select part of an image and send the coordinates of the selection to a database for further analysis. Analysts drew a polygon around the candidate's face in each still (figure 2). Then, red-green-blue (RGB) metrics were calculated for each pixel located inside this facial polygon.²

2. Automated methods are under development (see OpenCV, for example) that can detect a face in an image and extract such metrics, though false positives are still a problem. For an interesting application to faces in *Time* magazine covers, see <http://s3.amazonaws.com/aws.drewconway.com/viz/time/index.html> and https://github.com/drewconway/shades_of_time for the underlying code.

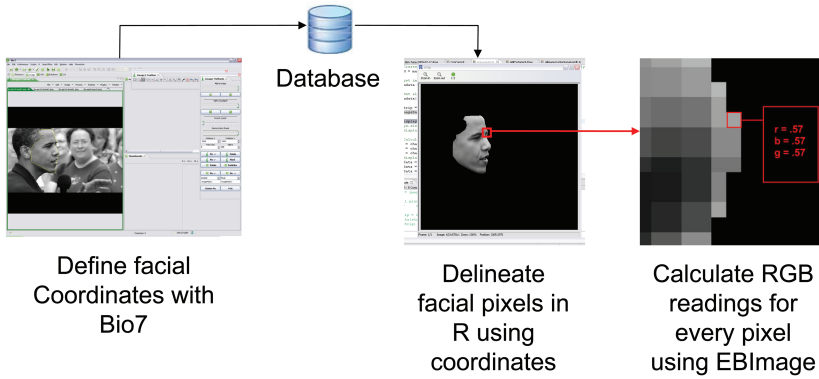


Figure 2. Data Pre-Processing.

In order to produce data that provided information about skin-tone darkness, V metrics were extracted from each pixel's RGB readings via a common mapping, as described below. Brightness simply maps to the highest RGB dimension, while color saturation corresponds to the widest difference between red, green, or blue (normalized by brightness):

$$V = \max(R, G, B) \quad (1)$$

$$S = \frac{V - \min(R, G, B)}{V} \quad (2)$$

The mean brightness V calculated for facial pixels serves as a measure of the image's depiction of a candidate's skin complexion along the relevant light-dark dimension.³ Of course, the metrics reflect that McCain has lighter skin than Obama: in facial pixels, McCain's average V reading is 0.62, while Obama's is 0.51.⁴ The distribution of V in each image is quite different across each campaign for each candidate, as shown in figure 3. Each candidate has on average lighter images of McCain, and the variance in both S and V is higher for each campaign's opponent. Interestingly, the Obama campaign's images of its own candidate are on average darker than the McCain campaign's. Of course, this could be due to a number of factors unrelated to the content of the ad (e.g., differences in recording and/or video processing equipment/software); instead, we care more about how skin complexion varies with content, which we examine below.⁵

3. We used the R package "maps" to return the exact pixels in each image that fell within the candidate's face polygon (figure 2).

4. The difference by candidate is highly significant: $T(531.36) = 9.55$, $P < 10^{-19}$, two-sided.

5. Each campaign used slightly less colorful images of its opponent (S), though we leave further examination of color saturation for future work.

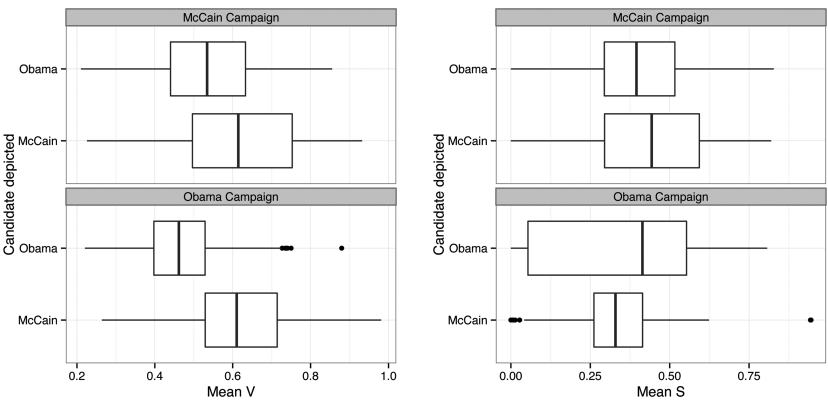


Figure 3. Skin Complexion in Ads for Each Candidate.

Each advertisement’s content was then classified along dimensions based on the literature. We sought to capture the overall “tone” of the ad (for negativity/attacks, e.g., [Ansolabehere et al. \[1994\]](#)), whether it contained attacks based on competence, character, or policy positions (based on appeals commonly referenced in the literature, e.g., [Popkin \[1994\]](#)), and in particular whether the ad attempted to associate a candidate with criminal activity (based on substantial evidence linking skin complexion and stereotypes related to crime, e.g., [Gilliam et al. \[1996\]](#); [Blair, Judd, and Chapleau \[2004\]](#); [Dixon and Maddox \[2005\]](#); [Eberhardt et al. \[2006\]](#)). We also recorded the presence of stereotype-inconsistent features, including a smile and formal attire, in each still (see [Dasgupta and Greenwald \[2001\]](#) and [Ito et al. \[2006\]](#) for examples of counter-stereotypical malleability with respect to race). We also coded aural and visual elements within each ad (to capture “dramatic” elements of advertisements; see [Jamieson \[1993\]](#)), whether the music sounded sinister, and whether the ad depicted visuals associated with children.

We refined the codebook after two pilot sessions with student coders, after which meetings were held to diagnose ambiguity and discuss disagreement. We made use of data from a third student coder who completed the coding task after training on the finalized codebook to avoid artificially inflating our agreement rates.

To attain measures of reliability, we compared codes attained from the trained student raters and from “master workers” on Amazon Mechanical Turk, a service often used for content analysis. Mechanical Turk serves as a market for tasks that can be done online, most often related to data collection and acquisition. Data obtained via Mechanical Turk has been shown to be a source of high-quality, reliable data (yielding measures of reliability comparable to traditional samples; see [Buhrmester, Kwang, and Gosling \[2011\]](#); [Sprouse \[2011\]](#)). Master workers in particular have been certified by Amazon

as having demonstrated excellence and accuracy⁶ for tasks including categorization (and command higher pay than ordinary workers). We had three master workers code each ad, took the modal code for each dimension, then computed Cohen's κ and Krippendorff's α to assess reliability between our trained undergraduate coders and our master workers from Mechanical Turk.⁷ Coders assessed all 126 ads and 534 still images.

We first examined message consistency between skin complexion and content. When comparing image-level content, we simply provide estimates of the mean V . When comparing ad-level content, we constructed an indicator variable that is non-zero if the ad contains an image that falls in the darkest (lowest V) quartile of the opponent's depictions utilized in the campaign, and use this variable to examine the probability that each ad contains a stereotype-consistent (dark) image of Obama.⁸ We use this indicator rather than simply taking the mean V for each ad because negative ads tend to display images that have been subjected to more modification in general than in other ads. Hence, the range of V is wider for these ads—and indeed, the most negative ads have a larger interquartile range (IQR).⁹ Yet, we are interested in whether an ad conveys a stereotype-consistent depiction of the author's opponent, not in the average depiction; it is unlikely that showing a high- V image depicting Obama in a “washed-out” photograph will somehow cancel out a stereotype-consistent image of Obama with darker skin (i.e., low V).¹⁰

RESULTS

The data show that the darkest images of Obama appear in the most negative, stereotype-consistent ads. First, we expected to see darker portrayals of Obama in ads that attempt to tie him to crime, based on the stereotyping literature reviewed above. Indeed, in attack ads that associated Obama with alleged criminal activity by leftists, the probability that the ad contained

6. Amazon does not publish these exact criteria, presumably to minimize the risk that workers will game the system.

7. Coding of specific content dimensions was robust: linking the candidate to the crime (both Cohen's κ and Krippendorff's α of .76); formal attire (button-down shirt) (κ = .62, α = .61); and smile (κ and α = .84). We excluded codes that did not attain sufficient reliability; for example, our “sinister music” measure attained κ = .31; our “policy attack” measure attained κ = 0.10, and our unprepared/competence attack dimension attained κ = 0.17).

8. Taking the number of images in an ad below the median for the candidate yields similar results. Taking the mean or median across the ad, the effects lose significance for depictions of Obama, due to higher variance in the most negative ads.

9. In attack ads that associate Obama with criminal activity, the average IQR for V is 0.091, versus 0.048 for other ads depicting Obama.

10. In addition, a large literature has found that negative attributions exert a stronger influence on relevant outcomes (including political outcomes) than positive or neutral attributions, especially in response to images (Lau 1982; Ansolabehere and Iyengar 1995; Ansolabehere, Iyengar, and Simon 1999; Spezio et al. 2008).

one of the darkest images is 0.86, compared to 0.30 for other ads ($W = 218$, $P = 0.006$, two-sided; see figure 4). In addition, counter-stereotypical images that depict Obama with a smile were marginally lighter ($\mu = 0.57$ versus $\mu = 0.53$, $T(71.16) = 1.84$, $P = 0.070$, two-sided), while the difference between images depicting Obama in formal attire ($\mu = 0.55$) compared to other images ($\mu = 0.49$, $T(16.39) = 1.70$, $P = 0.108$, two-sided) approaches significance.

The trend lines plotted in figure 5 suggest that as the election approached, attack ads featured images with darker depictions of Obama. Yet, as the OLS regression trend line indicates, on average the images did not change much. This is likely due to the fact that the images grew lighter as well, the higher variance being consistent with more exaggerated visual portrayals of Obama airing in advertisements airing closer to Election Day. At the same time, the

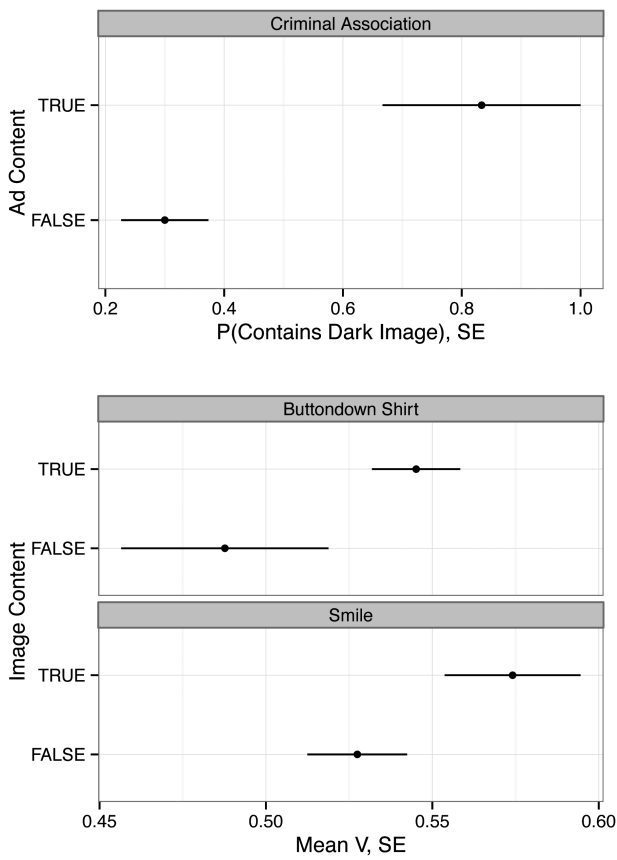


Figure 4. Message and Skin Tone Consistency in McCain Attack Ads Depicting Obama.

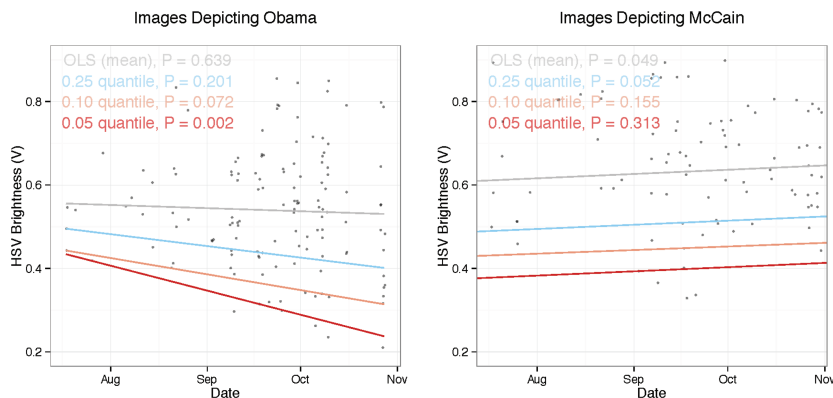


Figure 5. Key Longitudinal Relationship in Campaign Ads.

McCain campaign's own images of McCain grew on average lighter over time, suggesting that the aforementioned trends depicting Obama were not a relic of general trends toward darker campaign ads over time.¹¹ Because ads were more likely to contain stereotype-consistent images as Election Day approached, even short-lived effects would have been likely to be in play during the 2008 election.

LIMITATIONS AND CAVEATS

We have presented evidence that ad messages vary with visual content on dimensions related to stereotypes. In 2008, the variation was systematic, suggesting that the presence of darker images in certain advertisements is not likely to be due to chance or exogenous factors (e.g., a byproduct of how the opposing campaign uploaded images and video to the web). Of course, we cannot examine intentionality—it is impossible to determine whether stereotype-consistent images were included accidentally, purposefully, or incidentally, perhaps as a result of trying to make the opponent look bad.¹² This analysis also does not examine the question of differences in photographic contrast in the facial skin, which is thought to be an important element in attack advertisements (though one without clear implications for stereotype consistency). Of course, this analysis does not examine video clips, only still images and still captures from video, which, though not as common as still images, could be important when constituents judge candidates.

11. The same was true for color saturation: over time, the campaign used less saturated images to depict Obama, while using more saturated images of McCain.

12. The image quantities associated with the most negative ads here are fairly typical of most elections, according to Jamieson [1993], who notes that the use of black and white, dark colors, shadowed lighting, and stark contrasts are typically used in attack ads.

Though there is evidence that darker images of African Americans can activate stereotypes about Blacks, we do not know whether this effect extends to individuals as well known and as stereotype inconsistent as Barack Obama, which we address in study 2. Finally, despite evidence that darker complexion can affect judgments of target stimuli in the lab, we do not know whether the effect extends to evaluations of political candidates during campaigns, with so many other considerations present—a question that we address in study 3.

Study 2: Darkened Images of Obama and Stereotype Activation

Though we have shown that darker skin was associated with stereotype-consistent depictions of Obama in the 2008 campaign, the question of whether these darker portrayals of Obama are indeed more likely to activate negative stereotypes about Blacks remains open. The research cited above finds increased negative stereotype activation in response to Black target persons with darker skin (Blair et al. 2002; Livingston 2002; Maddox and Gray 2002; Ronquillo et al. 2007), but the targets in these studies are *not well-known counter-stereotypical exemplars* like Barack Obama, and *skin complexion itself is not manipulated* (rather, these studies use target persons with varying skin complexion). Indeed, there is evidence that the presence of a counter-stereotypical exemplar is often sufficient to regulate or prevent the activation of stereotypes (e.g., Ramasubramanian 2011). On the other hand, research examining the effect of skin complexion on how people perceive politicians (Terkildsen 1993; Iyengar et al. 2010; Weaver 2012) relies on vote choice and feeling thermometer ratings of the candidate, which could be capturing something other than stereotype activation.

In this study, we subject the skin-complexion hypothesis to an even more stringent test—whether a darkened image of a well-known candidate and counter-stereotypical exemplar, Barack Obama, can activate negative stereotypes about Blacks. We embed an image of Obama with either lightened or darkened skin in a survey that asked participants to consider the image and complete various words to measure stereotype activation, similar to Gilbert and Hixon (1991), Steele and Aronson (1995), Spencer et al. (1998), and Sinclair and Kunda (1999).¹³ The stimulus images (figure 6) depict Obama's face as either lighter ($V = 0.72$) or darker ($V = 0.53$) than in the original image ($V = 0.68$, not shown). The task comprised 11 words with missing blank spaces (e.g., L A _ _). Each fragment has as one possible solution a stereotype-related completion, along with non-stereotype-related completions. The complete list follows: L A _ _ (LAZY); C R _ _ _ (CRIME); _ _ O R (POOR); R _ _ (RAP); WEL _ _ _ _ (WELFARE); _ _ C E (RACE); D _ _ _ Y (DIRTY); B R _ _ _

13. These stimuli have been used previously in Iyengar et al. (2010).

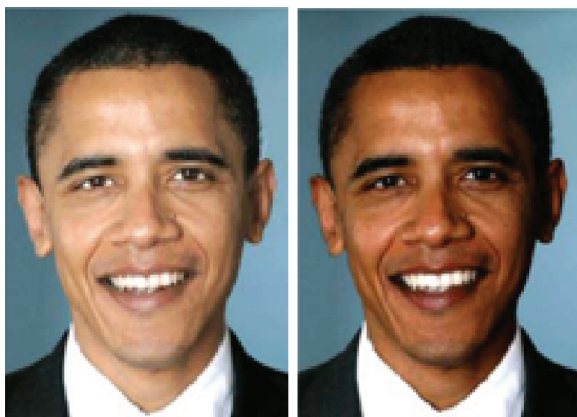


Figure 6. Stills Used in the Stereotype Activation Experiment.

__ (BROTHER); __ A C K (BLACK); M I ____ (MINORITY); D R __ (DRUG). These words were used in the stereotype-activation studies cited above. We focus on the three most unambiguously negative stereotypes in this study (LAZY, DIRTY, POOR), as prior research has shown that darker skin tends to activate the most negative stereotypes about Blacks (Blair et al. [2002]; Maddox and Gray [2002]; these variables also maximized interclass correlation). Participants also completed a standard battery of demographic measures.

Since we expect the effects of our manipulation to be more subtle than the effects documented in the lab studies cited above, we require more statistical power than a student sample can provide. Student samples at our West Coast university might also be expected to over-represent the young, liberals, and individuals who effortfully avoid stereotypical thinking, for whom the effect of a darker image of Obama might be substantially weaker than in the US population. While we would ideally like to obtain a nationally representative sample for this experiment, the resources necessary to attain a sufficient number of subjects via a firm such as Knowledge Networks or YouGov were not available.

As an alternative, we used Amazon's Mechanical Turk service to recruit participants. Berinsky, Huber, and Lenz (2012) show that this service provides a sample more representative than most in-person convenience samples (a finding replicated in Grimmer, Messing, and Westwood [2012]) and that Mechanical Turk experimental participants replicate experimental benchmarks. Our sample is also more diverse than a typical sample of college students, though not representative of the entire US population. Further, Grimmer, Messing, and Westwood (2012) show that the correlations among Mechanical Turk respondents are comparable to the correlations in benchmark survey data: Democrats, Republicans, liberals, and conservatives on Mechanical Turk respond like Democrats, Republicans, liberals, and conservatives in other studies. Studies from other



Figure 7. Stills Used in the AMP.

fields provide additional evidence of validity, showing that Mechanical Turk subjects are nearly indistinguishable from traditional laboratory samples—both in reproducing the results of classic studies (Buhrmester, Kwang, and Gosling 2011) and in replicating more recent experiments (Sproule 2011).

Our sample consisted of 630 Mechanical Turk workers, all of whom reported being over 18 and living in the United States. The average age reported was 34; 267 identified as female, while 363 identified as male; and 474 identified as White, 56 as Black, 31 as Hispanic, and 17 as other. Participants were paid \$0.50, which corresponds to an hourly rate of \$8.33 for the median participant. The full questionnaire is presented in the supplementary data online.

We took a variety of measures to ensure internal validity. Our between-subjects design minimized the possibility that participants might learn the purpose of the study and alter their behavior accordingly. We checked both the unique Amazon worker ID and each respondent's IP address against a database of all previous participants to ensure that each subject took the survey only once (using QualTurk, Kizilcec [2013]¹⁴), which not only maintains the integrity of our between-subjects design but also avoids analyzing repeat subjects in violation of the IID assumption. To further increase internal validity, we employed a series of standard “attention check” questions that assess whether the subjects were engaged with our questionnaire (the 257 who failed this check were excluded); removed subjects whose IP address did not resolve to a location within the United States (266); and removed subjects who reported not being a native English speaker (30). We also removed 73 participants who we suspected were not paying careful attention to the study itself, as reflected in survey completion times faster than those in the 95th percentile (1.92 minutes), and those who may have been preoccupied, as reflected by completion rates slower than those in the 5th percentile (8.13 minutes).¹⁵

RESULTS

Our results suggest that darker images of Obama can indeed activate negative stereotypes about Blacks, despite the fact that Obama is a counter-stereotypical exemplar. The mean number of fragments with stereotype-consistent completions was 0.33 in the “light” condition and 0.45 in the “dark” condition

14. QualTurk is a web-based software service designed to support Mechanical Turk in conjunction with Qualtrics.

15. Including these subjects in the analysis does not substantively alter the results.

($T(623.66) = 2.64$, $P = 0.008$, two-sided).¹⁶ In order to give the reader an intuition for the relative effect size, we provide an estimate of stereotype-consistent completions among conservatives: 0.55 compared to 0.36 for non-conservatives ($T(138.71) = 2.82$, $P = 0.005$).¹⁷ Those in the treatment were 36 percent more likely to complete an additional ambiguous word in a stereotype-consistent way, compared to 53 percent among conservatives. Thus, we find clear evidence that darker images of candidates can increase stereotype activation.

Study 3: Affective Responses to Different Images of Obama

Having established the causal effect of darker skin complexion on stereotype activation with an internally valid experiment, we turn to an additional source of data on how people respond to images of candidates with varying depictions of skin complexion, collected at the height of the 2008 presidential campaign. We use the Obama-McCain candidate Affect Misattribution Procedure (AMP), which exposed ANES participants to a variety of images of each candidate and collected affective responses. We emphasize that this study exploits the natural variation in skin complexion between these images, and hence should not be considered to be a randomized experiment.

The logic of the AMP follows: a respondent is asked to make a fast evaluative judgment about an ambiguous target (e.g., an abstract symbol, in this case a Chinese character) after being exposed to a prime for a split second. They are instructed to ignore the prime, and only to evaluate the symbol. However, exposure to each prime theoretically gives rise to a positive or negative evaluative reaction, and respondents' evaluations of original prime "transfer" to their reported ratings of the ambiguous target—that is, they misattribute their reaction from the original prime to the ambiguous object (Payne et al. 2005). Furthermore, this process tends to be resistant to corrective attempts.

The 2008–2009 ANES panel study, sponsored by the National Science Foundation (NSF), constitutes the data source for this study (Krosnick et al. 2009). The sample is based on random selection from a list of landlines to contact a sample of US citizens.¹⁸ The AMP was administered to respondents in waves 9 (September 2008, $N = 2,586$) and 10 (October 2008, $N = 2,628$) of the ANES; all interviews were conducted in English. We remove respondents who

16. Including the subset of completions that maximized alpha reliability (rather than interclass correlation)—lazy, black, poor, welfare, crime, and dirty—produces slightly noisier results: the means were 0.97 (light) versus 1.11 (dark), $T(626.72) = 1.77$, $P = 0.078$, two-sided.

17. Including the subset of variables that maximized alpha reliability (rather than interclass correlation) again produces noisier results: The means were 1.00 (non-conservative) versus 1.29 (conservative), $T(142.66) = 2.66$, $P = 0.009$, two-sided.

18. However, the user guide suggests that weights be used to generalize results to the US population. Weights are not applied in this analysis.

did not complete the AMP (wave 9, $N = 240$; wave 10, $N = 357$) or for whom AMP data were missing (wave 9, $N = 7$; wave 10, $N = 0$). When the analysis utilizes data on party identification, we remove respondents for whom such data are missing (wave 9, $N = 366$; wave 10, $N = 374$). As implemented in the 2008 ANES, the AMP measures affective responses 48 times per respondent to 16 different images of either Barack Obama or John McCain (8 images for each candidate, shown 3 times; see figure 7 for images). The images of Obama have a fairly wide range for the mean V measure, from 0.54 to 0.72.

The usual approach to the AMP is to take the mean of the responses from one set of primes and compare it to the other (e.g., pictures of Obama versus pictures of McCain) to get a sense of individual-level affect toward each set of targets. However, because this analysis aims to measure the variance in response to each photograph of each candidate, we use a multilevel modeling approach.¹⁹

We group the data by respondent, estimating a random intercept for each (which accounts for individual-level positive or negative evaluative tendencies), and then estimate fixed effects for the remaining parameters of interest (most notably the V skin-tone measure).²⁰

Formally, we consider the affective response as a binary outcome variable, y_{ij} , which represents how each respondent rated the ambiguous target (a Chinese character) in the AMP (0 = pleasant, 1 = unpleasant), for the i th subject ($i = 1, \dots, n$) on the j th occasion of measurement ($j = 1, \dots, J$). We estimate affective response as a logistic-normal mixed model, formally specified as follows:

$$\begin{aligned}\pi_{ij} | \alpha_i &= \text{logit}^{-1}(\beta_0 + \alpha_i + \beta_1 x_1 + \dots + \beta_p x_p) \\ \alpha_i &\sim N(0, \sigma^2)\end{aligned}\tag{3}$$

Where $\pi_{ij} | \alpha_i$ represents the probability that given respondent chooses the “unpleasant” response, logit^{-1} represents the logistic distribution function, α_i represents an intercept (random effect) for each respondent, each β term represents the coefficient (fixed effect) for each variable (1, ..., p) in the model.

RESULTS

Results modeling negative affect toward Obama show that respondents are more likely to rate the target “unpleasant” in response to darker images of Obama (table 1). Other model coefficients are in the expected direction, except that images with higher color saturation are oddly *more* likely to be rated as unpleasant. However, this could be because the AMP images of Obama where

19. We use the R package “lme4” for model estimation (Bates, Maechler, and Dai 2008).

20. In this case, the estimates are nearly identical when fitting a logit generalized linear model.

Table 1. Negative Affective Response to Obama

	(1)	(2)	(3)
(Intercept)	0.236 (0.130)	−0.844*** (0.216)	−1.567*** (0.230)
Mean V in photo	−1.101*** (0.192)	−1.092*** (0.195)	−0.596** (0.202)
Black respondent		−0.871*** (0.165)	−0.872*** (0.165)
Latino respondent		−0.294 (0.196)	−0.294 (0.196)
Female respondent		−0.185* (0.085)	−0.185* (0.085)
Respondent age		0.001 (0.003)	0.001 (0.003)
7 pt. Party ID		0.426*** (0.020)	0.427*** (0.020)
Mean S in photo			0.759*** (0.081)
Log-likelihood	−31444.164	−30148.739	−30105.748
Deviance	62888.328	60297.477	60211.496
AIC	62894.328	60313.477	60229.496
BIC	62921.493	60385.631	60310.669
<i>N</i> _{Obs}	63,264	61,032	61,032
<i>N</i> _R	2,636	2,543	2,543

NOTE.—Standard errors in parentheses.
p* ≤ 0.05; *p* ≤ 0.01; ****p* ≤ 0.001

his skin appears darker are also more saturated with color (the correlation between the S and V measures in this particular set of images is −0.26).

The relationships quantified in the above within-subjects analysis of the 2008 ANES AMP data are consistent with the effects we observe in study 2. Of course, this analysis of the AMP lacks the experimental controls we have in study 1—it is possible that skin tone was correlated with facial expression and/or lighting, for example, which probably also influences affective responses to a photograph. Nonetheless, we can conclude that differences between photographs can affect how a person responds to an image of the same candidate, even during the same sitting, and even when the candidate is well known.

Discussion

Using an original method to collect data on skin complexion, we demonstrate that campaign advertisements attacking Obama used darker images in the most negative, stereotype-consistent ads, and that these images were more frequent in ads that aired closer to Election Day. We then present an experiment

showing that darker portrayals of Obama are more likely to prime the most negative stereotypes associated with Blacks. We also find correlational evidence of the effect in an observational study that takes place at the height of the 2008 campaign. These findings help explain why darker depictions of Obama decreased support for his candidacy during the 2008 primary campaign (Iyengar et al. 2010) and why people tend not to prefer hypothetical Black candidates with a darker complexion (Terkildsen 1993). Together, the evidence we present shows that manipulating or selecting images of a Black candidate with a darker complexion can shape how individuals respond to political advertisements and think about politics.

Our findings underscore the importance of visual imagery in campaign advertisements. Past work has shown that campaign ads feature aesthetic qualities that match the ad's overall tone (Jamieson 1993), that ads often feature "implicit" racial appeals (Mendelberg 2001), and that racial imagery can affect preferences (Valentino, Hutchings, and White 2002). We show that campaign advertisements depict *candidates themselves* in ways that are consistent with the message conveyed in the ad. In this case, we find various instances in which the content in the ad and the visual depiction of the candidate are consistent with group-level stereotypes. Most notably, ads that tied Obama to crime also contained the darkest depictions of his skin complexion, both of which are linked to negative stereotypes about Blacks. Indeed, the most negative ads associating Obama with crime—including those attempting to tie Barack Obama to domestic terrorist Bill Ayers and allegations of misconduct related to the ACORN organization—contained the darkest images of Obama.

Our work also extends previous lab studies on how darkened images activate negative stereotypes about Blacks. Previous work has relied on student samples and stimuli limited to natural variation in skin complexion among targets that were unfamiliar to subjects. We extend these findings using an internally valid design that directly manipulates skin complexion, while using ecologically valid images of Barack Obama. The fact that a manipulated image of Obama can activate negative stereotypes about Blacks may come as a surprise in light of various studies suggesting that exposure to well-known racial exemplars tends to decrease subtle racial biases (for example, see Dasgupta and Greenwald [2001]).²¹ Our observational study from the 2008 ANES provides further evidence that people respond negatively to images that are consistent with stereotypes, despite the fact that subjects saw all pictures during the same sitting, and that he was so well known.

Though this evidence indicates that darker portrayals of Obama tend to appear in ads linking him to crime in study 1, we cannot conclude whether

21. In fact, Plant et al. [2009] found that during the 2008 presidential campaign, race IAT scores among study participants were not significantly different from zero. Of course, it seems quite plausible that the college student samples used in these studies were systematically different from the national sample analyzed with respect to whether and what proportion perceived Obama as a positive exemplar.

this pattern reflects a purposeful attempt to trigger stereotypes, an incidental result of simply trying to make Obama look bad, or something else altogether.²² Furthermore, we cannot assert that the darker images in question have been altered; instead, more shadowed images may have simply been selected. Yet, neither intent nor the question of manipulation versus selection is central here; rather, we focus on the effects of exposure to such images, showing that darker images of candidates can activate negative stereotypes (study 2) and providing further evidence based on natural variation between images (study 3). Of course, a common conclusion from the psychology literature on stereotyping is that the unintentional role that negative stereotypes about Blacks and other groups play in society is precisely what makes them so important to understand.

The fact that Obama won the 2008 and 2012 elections raises the question of the extent to which stereotype consistency “worked.” Of course, passing an electoral threshold does not preclude the possibility of a negative effect on electoral support. And, even if the effects of stereotype activation are short lived, the fact that darker images were more likely to appear in ads that aired immediately before the election suggests that any racial priming effects were likely in play during the election. Future work should exploit the increasingly high-quality data on what ads aired in what markets to attempt to identify the effect of such portrayals on turnout and vote margin at the district level. In the context of the rise of the Tea Party and Obama’s move away from a campaign message of bipartisan cooperation in 2012, we might expect to see even more evidence of “dirty politics” in that election. Regardless, as photographic appearance in campaigns continues to increase in importance (e.g., Polsby 1983) and more minority candidates face cameras in high-visibility races, research on the effects of stereotype consistency will not only benefit from richer data but also speak to an entrenched—yet in some sense increasingly urgent—problem in American politics.

Supplementary Data

Supplementary data are freely available online at <http://poq.oxfordjournals.org/>.

References

Ansolabehere, Stephen, and Shanto Iyengar. 1995. *Going Negative: How Political Advertisements Shrink and Polarize the Electorate*. New York: Free Press.

22. Indeed, there is evidence that strong Republicans tended to believe Obama’s skin tone was darker than did liberals during the 2008 campaign (Caruso, Mead, and Balcetis 2009), implying that the most loyal campaign managers may have simply used darker photographs in attack ads without conscious intent to create a stereotype-consistent narrative. This could help explain our results if, for example, stronger conservatives worked on later, more negative ads.

- Ansolabehere, Stephen, Shanto Iyengar, Adam F. Simon, and Nicholas A. Valentino. 1994. "Does Attack Advertising Demobilize the Electorate?" *American Political Science Review* 88:829–39.
- Ansolabehere, Stephen D., Shanto Iyengar, and Adam Simon. 1999. "Replicating Experiments Using Aggregate and Survey Data: The Case of Negative Advertising and Turnout." *American Political Science Review* 93:901–9.
- Ashmore, R. O., and F. K. Del-Boca. 1979. "Sex Stereotypes and Implicit Personality Theory: Toward a Cognitive-Social Psychological Conceptualization." *Sex Roles* 5:219–48.
- Averhart, Cara, and Rebecca Bigler. 1997. "Shades of Meaning: Skin Tone, Racial Attitudes, and Constructive Memory in African American Children." *Journal of Experimental Child Psychology* 67:363–88.
- Bargh, J. A., M. Chen, and L. Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71:230–44.
- Bates, Douglas, Martin Maechler, and Bin Dai. 2008. "lme4: Linear Mixed-Effects Models Using S4 Classes." <http://CRAN.R-project.org/package=lme4>.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20:351–68.
- Blair, Irene, Charles Judd, and Kristine Chappelle. 2004. "The Influence of Afrocentric Facial Features in Criminal Sentencing." *Psychological Science* 15:674–79.
- Blair, Irene V., Charles M. Judd, Melody S. Sadler, and Christopher Jenkins. 2002. "The Role of Afrocentric Features in Person Perception: Judging by Features and Categories." *Journal of Personality and Social Psychology* 83:5–25.
- Bodenhausen, G., and C. N. Macrae. 1998. "Stereotype Activation and Inhibition." In *Advances in Social Cognition*, edited by J. R. Wyer, 1–52. Mahwah, NJ: Erlbaum.
- Brewer, Marilynn B. 1988. "A Dual Process Model of Impression Formation." In *Advances in Social Cognition*, edited by Thomas K. Srull and Robert S. Wyer, vol. 1, 1–36. Hillsdale, NJ: Erlbaum.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6:3–5.
- Caruso, Eugene M., Nicole L. Mead, and Emily Balcetis. 2009. "Political Partisanship Influences Perception of Biracial Candidates' Skin Tone." *Proceedings of the National Academy of Sciences* 106:20, 168–20, 173.
- Dasgupta, Nilanjana, and Anthony G. Greenwald. 2001. "On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals." *Journal of Personality and Social Psychology* 81:800–814.
- Dixon, Travis L., and Keith B. Maddox. 2005. "Skin Tone, Crime News, and Social Reality Judgments: Priming the Stereotype of the Dark and Dangerous Black Criminal." *Journal of Applied Social Psychology* 35:1555–1570.
- Eberhardt, Jennifer, Paul Davies, Valerie Purdie, and Sheri Johnson. 2006. "Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes." *Psychological Science* 17:383–86.
- Eberhardt, Jennifer, Phillip Goff, Valerie Purdie, and Paul Davies. 2004. "Seeing Black: Race, Crime, and Visual Processing." *Journal of Personality and Social Psychology* 6:876–93.
- Fiske, Susan T., and Steven L. Neuberg. 1990. "A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation." *Advances in Experimental Social Psychology* 23:1–74.
- Gilbert, Daniel T., and J. Gregory Hixon. 1991. "The Trouble of Thinking: Activation and Application of Stereotypic Beliefs." *Journal of Personality and Social Psychology* 60:509–17.

- Gilliam, Franklin D., Shanto Iyengar, Adam Simon, and Oliver Wright. 1996. "Crime in Black and White: The Violent, Scary World of Local News." *Harvard International Journal of Press/Politics* 1:6–23.
- Grimmer, Justin, Solomon Messing, and Sean J. Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." *American Political Science Review* 106:703–19.
- Hofmann, Wilhelm, Tobias Gschwendner, Brian A. Nosek, and Manfred Schmitt. 2005. "What Moderates Implicit—Explicit Consistency?" *European Review of Social Psychology* 16:335–90.
- Ito, Tiffany A., Krystal W. Chiao, Patricia G. Devine, Tyler S. Lorig, and John T. Cacioppo. 2006. "The Influence of Facial Feedback on Race Bias." *Psychological Science* 17:256–61.
- Iyengar, Shanto, Solomon Messing, Jeremy Bailenson, and Kyu Hahn. 2010. "Explicit Racial Cues in Campaign Advertising: The Case of Skin Complexion in the 2008 Campaign." Presented at the Annual Meeting of the American Political Science Association, Washington, DC.
- Jamieson, Kathleen H. 1993. *Dirty Politics: Deception, Distraction, and Democracy*. New York: Oxford University Press.
- Kizilcec, Rene. 2013. QualTurk. A web-based software service designed to support Mechanical Turk in conjunction with Qualtrics.
- Klatzky, R., G. Martin, and R. Kane. 1982. "Influence of Social-Category Activation on Processing of Visual Information." *Social Cognition* 1:95–109.
- Kolawole, Emi, and Justin Bank. 2008. "Did Clinton Darken Obama's Skin?" <http://www.factcheck.org/2008/03/did-clinton-darken-obamas-skin/>.
- Krosnick, Jon A., Arthur Lupia, Vincent L. Hutchings, Matthew DeBell, and Darrelle Donakowski. 2009. Advance release of the 2008–2009 ANES panel study [data set]. Stanford University and the University of Michigan.
- Lau, R. R. 1982. "Negativity in Political Perception." *Political Behavior* 4:353–78.
- Livingston, Robert W. 2002. "The Role of Perceived Negativity in the Moderation of African Americans' Implicit and Explicit Racial Attitudes." *Journal of Experimental Social Psychology* 38:405–13.
- Maddox, Keith B., and Stephanie A. Gray. 2002. "Cognitive Representations of Black Americans: Re-Exploring the Role of Skin Tone." *Personality and Social Psychology Bulletin* 28:250–59.
- Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, NJ: Princeton University Press.
- Payne, B. Keith, Clara M. Cheng, Olesya Govorun, and Brandon D. Stewart. 2005. "An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement." *Journal of Personality and Social Psychology* 89:277–93.
- Plant, E. Ashby, Patricia G. Devine, William T. L. Cox, Corey Columb, Saul L. Miller, Joanna Goplen, and B. Michelle Peruche. 2009. "The Obama Effect: Decreasing Implicit Prejudice and Stereotyping." *Journal of Experimental Social Psychology* 45:961–64.
- Polsby, Nelson. 1983. *Consequences of Party Reform*. Oxford: Oxford University Press.
- Popkin, Samuel L. 1994. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.
- Ramasubramanian, Srividya. 2011. "The Impact of Stereotypical Versus Counterstereotypical Media Exemplars on Racial Attitudes, Causal Attributions, and Support for Affirmative Action." *Communication Research* 38:497–516.
- Ronquillo, Jaclyn, Thomas F. Denson, Brian Lickel, Zhong-Lin Lu, Anirvan Nandy, and Keith B. Maddox. 2007. "The Effects of Skin Tone on Race-Related Amygdala Activity: An fMRI Investigation." *Social Cognitive and Affective Neuroscience* 2:39–44.
- Sinclair, Lisa, and Ziva Kunda. 1999. "Reactions to a Black Professional: Motivated Inhibition and Activation of Conflicting Stereotypes." *Journal of Personality and Social Psychology* 77:885–904.

- Spencer, Steven J., Steven Fein, Connie T. Wolfe, Christina Fong, and Meghan A. Duinn. 1998. "Automatic Activation of Stereotypes: The Role of Self-Image Threat." *Personality and Social Psychology Bulletin* 24:1139–1152.
- Spezio, Michael L., Antonio Rangel, Ramon M. Alvarez, John P. O'Doherty, Kyle Mattes, Alexander Todorov, Hackjin Kim, and Ralph Adolphs. 2008. "A Neural Basis for the Effect of Candidate Appearance on Election Outcomes." *Social Cognitive and Affective Neuroscience* 3:344–52.
- Sprouse, Jon. 2011. "A Validation of Amazon Mechanical Turk for the Collection of Acceptability Judgments in Linguistic Theory." *Behavior Research Methods* 43:155–67.
- Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69:797–81.
- Terkildsen, Nayda. 1993. "When White Voters Evaluate Black Candidates: The Processing Implications of Candidate Skin Color, Prejudice, and Self-Monitoring." *American Journal of Political Science* 37:1032–1053.
- Troutnut. 2008. "Hillary's Ad: Debate Footage Doctored to Make Obama Blacker." <https://web.archive.org/web/20151017114413/http://www.dailykos.com/story/2008/3/3/14550/75567/858/467989>.
- Valentino, Nicholas A., Vincent L. Hutchings, and Ismail K. White. 2002. "Cues That Matter: How Political Ads Prime Racial Attitudes During Campaigns." *American Political Science Review* 96:75–90.
- Weaver, Vesla M. 2012. "The Electoral Consequences of Skin Color: The Hidden Side of Race in Politics." *Political Behavior* 34:159–92.
- Zarate, Michael A., Colby J. Stoeber, M. Kimberly MacLin, and Clarissa J. Arms-Chavez. 2008. "Neurocognitive Underpinnings of Face Perception: Further Evidence of Distinct Person and Group Perception Processes." *Journal of Personality and Social Psychology* 94:108–15.