When a face type is perceived as threatening: Using general recognition theory to understand biased categorization of Afrocentric faces

Heather M. Kleider-Offutt^{1,2} · Alesha D. Bond^{1,2} · Sarah E. Williams^{1,2} · Corey J. Bohil^{1,2}

Published online: 7 March 2018
© Psychonomic Society, Inc. 2018

Abstract

Prior research indicates that stereotypical Black faces (e.g., wide nose, full lips: Afrocentric) are often associated with crime and violence. The current study investigated whether stereotypical faces may bias the interpretation of facial expression to seem threatening. Stimuli were prerated by face type (stereotypical, nonstereotypical) and expression (neutral, threatening). Later in a forced-choice task, different participants categorized face stimuli as stereotypical or not and threatening or not. Regardless of prerated expression, stereotypical faces were judged as more threatening than were nonstereotypical faces. These findings were supported using computational models based on general recognition theory (GRT), indicating that decision boundaries were more biased toward the threatening response for stereotypical faces than for nonstereotypical faces. GRT analysis also indicated that perception of face stereotypicality and emotional expression are dependent, both across categories and within individual categories. Higher perceived stereotypicality predicts higher perception of threat, and, conversely, higher ratings of threat predict higher perception of stereotypicality. Implications for racial face-type bias influencing perception and decision-making in a variety of social and professional contexts are discussed.

Keywords Decision making · Face processing · Perception · Social cognition

Biased treatment has plagued Black men across a variety of contexts, including employment and education, and throughout the legal system from law enforcement to juror decisions. Such treatment is attributed, in part, to cultural/social stereotypes associating Black men with crime (e.g., Cooper, 2015; Fiske, 1998). Negative stereotypes and judgments are especially evident for a subgroup of Black men who have Afrocentric features, also called a stereotypical Black face type (e.g., some combination of darker skin, wide nose, and full lips; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). These individuals are considered characteristic of the category “Black male” and most closely tied to the criminal Black male stereotype (Kleider, Cavrak & Knuycky, 2012). The question considered in the present study is whether there is a perceptual-based or decision-based relationship between stereotypical face type and threat, such that Black men who

have Afrocentric features are inherently perceived as threatening, or whether biases occur at a separate decisional stage. Are Afrocentric Black faces perceived as threatening regardless of their facial expression or intent? Potentially, scenarios wherein facial expression (i.e., demeanor) is evaluated (e.g., police questioning, job interviews), a misinterpretation of facial expression could be interpreted as a cue to aggressive behavior and can negatively impact communication.

A growing body of research has shown that emotional expressions interact with social categories, such as race, to influence how emotional cues are interpreted (see, for review, Smith, LaFrance, & Dovidio; 2017). For example, Black faces are identified faster as angry than are White faces (Hugenberg, 2005; Hugenberg & Bodenhausen, 2003), and bodily expressions of anger compared to happiness facilitate the categorization of Black targets (Hinzman & Kelly, 2013). Moreover, Black men are stereotypically associated with anger (see Bijlstra, Holland, Dotsch, Hugenberg & Wigboldus; 2014; Eberhardt, Goff, Purdie, & Davies, 2004, for discussion), largely stemming from the association between Black men and assumed aggression/crime (Correll, Park, Judd, & Whittenbrink; 2002; Devine, 1989). This link likely underpins the expectation of anger even in nonverbal cues that support the interpretation of emotion and intent. That is, even when

✉ Heather M. Kleider-Offutt
hoffutt@gsu.edu

¹ Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, GA, USA

² University of Central Florida, Orlando, FL, USA

facial expression is ambiguous, Black men are considered more hostile and threatening than are White men (Hugenberg & Bodenhausen, 2003). Furthermore, the distorted construal of Black men as physically threatening results in their being judged as physically larger than they are (Cottrell & Neuberg, 2005; Wilson, Hugenberg & Rule, 2017), which translates to hypervigilance in their assessment because they are considered potentially aggressive (Haselton & Buss, 2003; Wilson et al., 2017).

Activation of the criminal Black male stereotype may occur automatically (Devine, 1989; Eberhardt et al., 2006; Eberhardt et al., 2004) such that not applying the stereotype requires engagement of cognitive resources (Devine & Elliot, 1995; Fiske, 1998, 2004; Macrae, Milne, & Bodenhausen, 1994; Stewart & Payne, 2008). When people categorize others, they make quick judgments and look to confirm their initial inclination and seek other information if motivated to do so (Fiske, 2004). Facial cues to emotion provide a quick indication of intent, and racial stereotypes may be implicitly activated by subtle cues related to face processing. Using a gun/tool classification task, Kubota and Ito (2014) primed participants with Black and White male faces posing with angry, happy, or neutral expressions just before making the classification decision. Generally, findings showed that a Black angry prime elicited negative stereotypes, whereas a Black smiling prime diminished negative responses, suggesting that emotional expression can attenuate automatic racial stereotyping, thereby diffusing expectations of threat, similar to what may occur in everyday encounters.

Threat is not conveyed solely by displays of emotion (and the interaction between emotion and race); craniofacial structure can also provide cues to threat and indicate perceived behavior patterns. For example, facial width and eyebrows positioned low on the face can suggest aggressive tendencies even when the facial emotion is neutral (see Deska, Lloyd, & Hugenberg, 2017; Shasteen, Sasson, & Pinkham, 2015) and may be associated with conclusions about aggressive behavior (e.g., Carré, McCormick, & Mondloch, 2009; Carré, Morrissey, Mondloch, & McCormick, 2010). Thus, facial structure in and of itself may signal potential threat. It is possible that the facial structure of Afrocentric faces (i.e., stereotypical Black faces) activates expectations of threat.

Research on face processing suggests that certain faces may seem more characteristic of a race compared to others and that the degree to which a face is emblematic of race influences the judgments made about that face (Colarelli, Poole, Unterborn, & D'Souza, 2010; Knuycky, Kleider, & Cavrak, 2014; Ma & Correll, 2011; Strom, Zebrowitz, Zhang, Bronstad, & Lee, 2012; Wilkins, Kaiser, & Rieck, 2010). Possessing Afrocentric features is considered archetypal of the Black male face (Knuycky et al., 2014). Afrocentric Black face types are more readily associated with negative racial stereotypes than are non-Afrocentric faces because they are closer to what is considered

illustrative of the category “Black” (Knuycky et al., 2014; Valentine, 1991). Specifically, men with Afrocentric facial features are more likely to be miscategorized into negative, criminal-type roles (Kleider, Cavrak, & Knuycky, 2012), misidentified from a lineup (Knuycky et al., 2014), misidentified and falsely incarcerated (Kleider-Offutt, Knuycky, Clevinger & Capadanno, 2017a), prescribed longer sentences (Blair et al., 2004), and have an increased likelihood of being given the death penalty in murder cases compared to Black men with non-Afrocentric faces (Eberhardt et al., 2006; however, also see Wilson & Rule, 2016, for other findings). Although to a lesser extent than in Black men, people also attribute Afrocentric features to White male faces. White men with Afrocentric facial features are also more likely to be ascribed negatively biased judgments than are White men with non-Afrocentric features (Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004; Kleider et al., 2012). Thus, the association between Afrocentric facial features and assumed criminality extends beyond solely skin tone; it rather is a unique constellation of features that influences subsequent judgment (Blair, Chapleau, & Judd, 2005; Blair, Judd, Sadler, & Jenkins, 2002; Hagiwara, Kashy, & Cesario, 2012).

If Afrocentric features activate negative racial stereotypes, it may be especially difficult to interpret facial expressions as non-threatening. As such, the close alignment of Afrocentric features to the category Black male has instantiated the face type as a harbinger for aggressive behavior. Responses to face type as a perceptual cue may occur automatically. Moreover, if this is true, then regardless of the intended expression, neutral or happy, Afrocentric faces may be misinterpreted as threatening.

Although the literature summarized here is consistent in linking facial features to biased interpretation of emotional expression and threat, it is unclear whether visual perception or separate decision processes underlie these judgments. Does the perception of Afrocentric facial features simultaneously co-occur with the perception of threatening expression (i.e., is any facial expression actually *perceived* as more threatening in the context of Afrocentric features)? Or does the relationship between these dimensions depend on (possibly unconscious) *decisions* that take an unbiased perceptual representation as input (e.g., decisions based on expectations)? The behavioral patterns reported in the literature summarized above may be the product of biased decision-making that is potentially independent of the perceptual representation.

What is needed to answer these questions is an analytic approach that can assess the separate contributions of perception and decision processes, and in addition provide insight into dimensional interactions between face type and emotional expression. We address this goal in the current research by employing general recognition theory—a multidimensional generalization of signal detection theory (Ashby & Townsend, 1986)—to examine perceptual and decisional contributions to the relationship between face type and emotional expression.

General recognition theory (GRT)

In this section we summarize the conceptual foundations of GRT, which will provide a basis for understanding our modeling results and conclusions. Following this section, we consider some recent applications of GRT to the classification of face stimuli, followed by details of the current study.

Like signal detection theory, GRT provides separate measures of perceptual and decisional contributions to classification judgments. GRT greatly expands on this, though, by providing measures of perceptual and decisional interactions between dimensions, both within each individual category as well as across the different categories at each feature level.

GRT commonly models the perceptual representation of an individual category as a multivariate normal distribution. Figure 1 shows an example with four categories derived from factorially combining two levels of two dimensions. Because the category means, variances, and covariances are sufficient to communicate all important information about dimensional interactions, communication is simplified by depicting categories with equal-likelihood contours (i.e., slicing the distributions at a single height and viewing from above the plane; see Fig. 1, right, and Fig. 2). The strength of the GRT approach is that it enables the researcher to explore questions about integration of stimulus dimensions. GRT allows insight into the extent to which dimensions interact at a perceptual level, at a decisional level, or both.

One important question answered by GRT is whether perceptual interaction occurs between dimensions within a single category. *Perceptual independence* (PI) reflects the fact that there is perceptual variability among stimuli within a single category. For example, if a face sampled from a category is perceived as high on the stereotypicality dimension—relative to other stimuli from that *same* category—then does that stimulus also tend to be perceived as more threatening? In other words, within a category, does perceived stereotypicality predict perceived emotion (i.e., is there a nonzero correlation between perceptual dimensions within a category)? PI holds when the dimensions *do not* interact within a single category. In Fig. 2, PI holds in the bottom

contours of equal likelihood. Even though the variance along each dimension is unequal for these categories (i.e., they are not circular), the dimensions are perceived as orthogonal to the coordinate axes, and thus nonpredictive of each other. The top contours of equal likelihood in Fig. 2 display violations of PI because the contours are noncircular and also convey a positive correlation between dimensions.

Before continuing, a few points of clarification may be helpful. Throughout this article, we use the term *stereotypicality* to refer to a continuous-valued perceptual dimension that reflects the degree to which faces possess Afrocentric features. The more apparent these features appear, the more stereotypically Black we consider the face. This is not to be confused with the notion of “typicality,” which is widely used in the categorization literature referring to closeness of a category member to the category centroid or prototype. This definition of typicality characterizes virtually any category that has some degree of central tendency. In the current research, two of the categories are high—relative to the other two categories—on the stereotypicality dimension and two are comparatively low along this dimension, and the stimuli within each category vary along this dimension.

Another important point is that the assumed distributions of perceptual effects in GRT (as well as in signal detection theory) shown in Figs. 1 and 2 are often described and tested in terms of noisy perception of individual stimuli. Repeated presentation of a single stimulus is thought to result in a normal distribution of perceptual effects. Indeed, many tests of GRT (referenced below) are carried out using a four-stimulus identification task that factorially combines two levels of two stimulus dimensions. A great deal of categorization research, however—with many distinct stimuli per category—adopt these same distributional assumptions to represent categories (e.g., Ashby & Gott, 1988; Ashby & Lee, 1991; Maddox & Ashby, 1993). Because the combined distribution of perceptual effects of each stimulus can be thought of as a probability mixture, the model is formally equivalent for identification and categorization if the category stimuli are normally distributed in stimulus space. Traditionally, GRT-based categorization

Fig. 1 Multivariate normal distributions representing categories in 2-D perceptual space (i.e., along two stimulus dimensions, each with two levels factorially combined to create four categories). Left: view of

category distributions. Middle and right: alternate view of categories using contours of equal likelihood

Fig. 2 Hypothetical contours of equal likelihood from a complete factorial categorization experiment along two stimulus dimensions (A and B). Marginal distributions portray category means and variances for both levels of each dimension. The horizontal line illustrates a bound consistent with decisional separability (DS), while the slanted vertical line illustrates one possible violation of DS along Dimension A. A violation of perceptual separability (PS) is shown by the vertical non-alignment of the two right-side contours. The (noncircular, tilted) top row contours illustrate violations of perceptual independence

studies use artificial stimuli whose dimensional values are directly sampled from normal distributions. Recent research, however, using complex real-world stimuli (cited below), has assumed this representation for a large collection of photographs (as does the current research). In the Method section for Study 1, we provide additional details that seem to support adoption of the standard GRT assumptions about category distributions.

Another important question answered by GRT is whether the level of one dimension influences the perceived level of another dimension *across separate* categories. For example, are faces showing a “neutral” emotional expression perceived as neutral both for stereotypical faces and nonstereotypical faces (i.e., across category levels of the other dimension)? If *perceptual separability* (PS) holds, then neutral expressions would be perceived the same way regardless of the level of the other dimension. Violations of PS can appear as mean shifts in contour location across levels of the orthogonal dimension, or as unequal variances across levels of the orthogonal dimension. For example, the right side of Fig. 2 depicts a violation of PS based on mean shift across levels of Dimension B (although the variances at level A₂ are the same across levels of B). PS holds on the left side of Fig. 2 because the means and variances at level A₁ are the same across levels of Dimension B. Another example is the violation of PS based on unequal variances for the two categories

at level B₁ across levels of Dimension A (i.e., the bottom two contours).

GRT also distinguishes between perceptual and decisional contributions to classification judgments. GRT analysis allows insight into dimensional interactions in terms of decision criterion placement. For example, an observer might use the same decision criterion to separate neutral and threatening facial expressions at both levels of the orthogonal dimension (e.g., stereotypical or nonstereotypical facial features). So, like PS, *decisional separability* (DS) pertains to dimensional interactions across different categories. DS holds if the decision criterion used on one dimension is unchanged at each level of the orthogonal dimension. In Fig. 2, DS holds on Dimension B (i.e., the same criterion is used at each level of Dimension A), depicted by a linear decision bound that is orthogonal to the coordinate axes. Violation of DS is reflected in any decision bound that is nonorthogonal to the coordinate axes, reflecting decisional integration of category dimensions. Figure 2 depicts a violation of DS along Dimension A, such that a different criterion is used for categories at levels B₁ and B₂. GRT assumes that decision bound placement can be essentially independent of perception of stimuli and perceptual interactions.

In sum, GRT accounts for classification choices by parsing the distinct contributions of perceptual processing from decision processes. By specifying stimulus levels along two dimensions, GRT provides much more information than traditional signal detection analysis. GRT analysis provides insight into dimensional interactions both within and between categories at both a perceptual and decisional level.

Applications of GRT

Several applications of GRT have focused on interactions between facial dimensions (e.g., “holistic” processing of—or lack of selective attention to—separate face dimensions; Cornes, Donnelly, Godwin, & Wenger, 2011; Richler, Gauthier, Wenger, & Palmeri, 2008; Thomas, 2001; Wenger & Ingvalson, 2002, 2003). Much of the research using the GRT approach has focused on identification tasks composed of only four face stimuli (e.g., Soto, Vucovich, Musgrave, & Ashby, 2015). In most cases, this research has been conducted with artificial stimuli or otherwise highly controlled laboratory conditions. Recently, researchers have begun capitalizing on the GRT approach to understand classification judgments using less controlled, more naturalistic stimuli and viewing conditions.

An example of this is the work of Treat and colleagues (Farris, Viken, & Treat, 2010). Farris et al. (2010) sought to explore interaction of perceptual and decisional contributions to young men’s errors in interpreting sexual interest. They created categories of photos depicting women along dimensions of clothing style (provocative or not) and facial affect (friendly or sexually interested). Based on GRT analysis, they found that men at

higher risk for sexual coercion tended to perceive an illusory correlation between facial affect and clothing style. In order to collect classification data, the researchers first conducted a photo-rating study to find stimuli to represent each category. The research we report here is in some ways modeled on the research reported by Farris et al.

In the current study, GRT modeling is employed to understand whether perception of facial expression is to some degree determined by perception of facial features. Moreover, we wish to examine whether the dimensions have independent decisional or perceptual effects on judgment, or whether they are perceived as a whole (nonindependent) with biases largely driven by decision cutoff placement. We are testing the extent to which the interaction of face type and threat, as summarized in the literature, lies only in perception, as it has commonly been described in the literature. This study will add insight into how much face-type bias found with judgments about Afrocentric features is decision based rather than perceptual.

Overview of the experiment

In a pilot study, faces were rated by separate groups of participants on facial-feature stereotypicality or emotional expression to determine their relative position along the two perceptual dimensions of interest: stereotypicality and emotion. A subset of these pre-rated faces were then chosen as members of the four categories used in Study 1: stereotypical/neutral, stereotypical/threatening, nonstereotypical/neutral, and nonstereotypical/threatening. In Study 1, a separate group of participants classified the subset of faces as threatening or not and Afrocentric (i.e., stereotypically Black) or not.

Pilot study

Method

Participants The pilot study participants consisted of 209 Georgia State University (GSU) undergraduate students. All of the students participated for course credit and self-identified their age (range: 18–60 years), gender (174 female, 35 male), and race (94 Black, 42 White, 73 other race).

Procedure In a between participants design, participants rated 224 photos of Black faces on typicality (i.e., Afrocentricity) and attractiveness (not reported in detail in this article) or emotional expression (either threatening or neutral). Each participant rated only one variable of interest (e.g., only stereotypicality or only degree of threat). All of the faces were taken from online actor databases (also used in Kleider et al., 2012), previously published facial stimuli databases (Ma, Correll, & Wittenbrink, 2015), and mugshot databases

(Florida Department of Corrections). All selected faces were listed in mugshot and actor databases by ethnicity, and those selected were identified as Black; faces labeled as biracial on the respective source were not used in this study. The faces were presented one at a time in random order. Participants were instructed to use their own subjective criteria for their ratings (see Blair et al., 2002). Participants were then asked to indicate either how stereotypically Black (i.e., “Using your own opinions of what you think a stereotypically Black face means, how stereotypically Black do you perceive this face to be?”) or how attractive (i.e., “Using your own opinions of what you think an attractive face means, how attractive do you perceive this face to be?”) or how threatening [or neutral] (i.e., “Using your own opinions of what you think a face depicting a threatening [or neutral] attitude means, how threatening [or neutral] do you perceive this face to be?”) he or she perceived that face to be. The attractiveness question was included as a control of sorts to rule out halo effects, there was no hypothesis about this rating. Thus, different groups of participants rated each variable of interest. Ratings were reported on a 1–7 Likert scale (1 = *not at all stereotypical/attractive/neutral/threatening*, 7 = *very stereotypical/attractive/neutral/threatening*). This procedure and category options were successfully used in our previous work (Kleider et al., 2012). Outcomes of the pilot study informed selection of the stimulus faces used in Study 1.

Pilot results

Average attractiveness, stereotypicality, and emotion expression ratings were calculated for each face (averaged ratings across participants in each rating group). Based on a median split along the dimensions, each face was categorized into a face-type group (*nonstereotypical*, range: 2.10–4.93, $n = 115$; *stereotypical*, range: 4.94–6.44, $n = 109$) and attractiveness groups (*not attractive*, range: 1.10–2.00, $n = 113$; *attractive*, range: 2.00–5.86, $n = 111$). Each face was also categorized as either threatening (*not threatening*, range: 1.41–3.51, $n = 114$; *threatening*, range: 3.52–6.00, $n = 110$) or neutral (*not neutral*, range: 2.57–3.75, $n = 117$; *neutral*, range: 3.76–6.38, $n = 107$) via median split.^{1,2}

We used these median-split-based groupings to reduce the number of face stimuli to consider when constructing the categories for Study 1. As detailed below, from these groupings

¹ Nonattractive, nonstereotypical faces ($M = 4.59$) were significantly more stereotypical than attractive, nonstereotypical faces ($M = 3.89$). There was no significant stereotypicality difference between nonattractive, stereotypical faces and attractive, stereotypical faces, $F(1, 56) = 7.61$, $p = .008$, $\eta_p^2 = .12$.

² Nonattractive, neutral faces ($M = 3.64$) were significantly more threatening than attractive, neutral faces ($M = 2.65$). There was no significant threat difference between nonattractive, threatening faces and attractive, threatening faces, $F(1, 56) = 4.28$, $p = .04$, $\eta_p^2 = .07$.

we selected a subset of stimuli with the largest average ratings on each rated dimension.

Study 1

Method

Participants The participants included 178 GSU undergraduate students. All of the students participated for course credit and self-identified their age (range: 18–50 years), gender (147 female, 31 male), and race (74 Black, 30 White, 74 other race).

Materials The stimulus set included 60 photographs. Four categories were created based on the ratings from the pilot study. The categories (nonstereotypical/neutral; nonstereotypical/threatening; stereotypical/neutral; stereotypical/threatening; see Fig. 3) represented quadrants of the two-dimensional perceptual space (along dimensions of emotional expression and stereotypicality of facial features). Of the original 224 photographs assessed in the pilot study, only the most highly rated photographs were used for each quadrant (i.e., 15 nonstereotypical/neutral faces; 15 nonstereotypical/threatening faces; 15 stereotypical/threatening faces; 15 stereotypical/neutral faces). Table 1 reports the range of

Table 1 Range of stereotypical, threat, and neutrality ratings across dimensions of selected face categories

Category	Stereotypicality	Threat or neutrality
Stereotypical/Threatening	5.38–6.04	3.67–5.22
Stereotypical/Neutral	5.00–6.44	3.76–5.86
Nonstereotypical/Threatening	3.69–4.90	3.52–5.15
Nonstereotypical/Neutral	2.60–4.21	4.71–6.05

average pilot ratings for these stimuli. Note that in order to be selected as a category member, a stimulus had to be high (above the median split) on both stimulus dimensions.

Because category membership was based on rating-scale data averaged across participants, we do not attempt to make strong claims about the distribution of stimuli within each category. However, examining the distribution of ratings within each category seems to suggest that the stimuli were approximately normally distributed (or at least does nothing to contradict this assumption). Observing the shape of kernel density plots makes clear that each distribution along both dimensions was approximately unimodal and symmetric. Ratings along each dimension (within each category) passed dip tests for unimodality (p value $> .15$ for all dimensions; Hartigan & Hartigan, 1985). And there was no significant correlation between dimensions for any category (all p values $> .43$).

Procedure Participants viewed a three-block repeated series (randomly ordered) of 60 equally cropped and sized pictures of target faces (prerated from the pilot study; categories described above) in the center of the computer screen for 3 seconds. Each picture was presented individually on a trial, once per block. After each individual picture presentation, participants were instructed to provide a forced-choice response: “Using your own opinion of what you perceive a stereotypically Black face to look like, is this face stereotypical or not? (Yes or No) and “Is this person’s attitude threatening or neutral?” These questions were counterbalanced such that half the time people were asked about the stereotypicality first and half the time people were asked about emotionality first. Participants were then asked demographic questions including age, gender, and race.

GRT modeling analysis Traditionally, GRT analysis has proceeded by computation of summary statistics based on hit and false alarm rates in classification responding (referred to as multidimensional-signal detection analysis), or by fitting a series of models to response data instantiating different hypotheses about PS, DS, and PI (see Ashby & Soto, 2015, for a detailed explanation of both approaches). However, recent research has highlighted a limitation of the traditionally used analytic approach (Mack, Richler, Gauthier, & Palmeri, 2011;

Fig. 3 Black male target face type and emotionality examples: stereotypical/neutral (upper left); stereotypical/threatening (upper right); nonstereotypical/neutral (bottom left); nonstereotypical/threatening (bottom right)

Silbert & Thomas, 2013; Soto et al., 2015). Specifically, if DS is violated, then it is not possible to uniquely identify the source of interaction between dimensions (i.e., either perceptual, decisional, or both).

In GRT research exploring dimensional interactions, the simplest, and most widely used, design is a 2×2 factorial combination of stimulus dimensions. This results in four categories, responses to which are arranged in a 4×4 confusion matrix (4 categories $\times$ 4 category response options). In this design, fitting models to test for violations of PS, DS, or PI requires estimating more free parameters than there are degrees of freedom in the confusion matrix. As a result, simplifying assumptions are enacted by fixing the values of some model parameters. In particular, the modeling approach has tended to assume DS holds. But this approach cannot guarantee that DS holds, and so conclusions based on results are weakened.

Recently, Soto et al. (2015) proposed a solution to this problem that entails fitting a single GRT model to all participants' confusion matrices simultaneously. The logic is similar to that underlying multidimensional scaling analysis with individual differences (INDSCAL: Carroll & Chang, 1970). The model, referred to as GRT-wIND, assumes that a single perceptual representation of the stimulus space can provide a basis for modeling individual differences in stimulus perception. However, personalization of the perceptual representation for each participant is modeled by unique attention weights on each dimension for each participant. The model also estimates unique decision bounds on each dimension for each participant. Using this approach, it is possible to verify whether DS held for each stimulus dimension and enabling identification of perceptual and decisional contributions to classification performance. To take advantage of this recent development in modeling, we fit the GRT-wIND model to the data reported here.

Results

Classification accuracy A 2 (stereotypicality level: nonstereotypical, stereotypical) $\times$ 2 (facial expression: neutral, threatening) repeated-measures ANOVA, with face type and facial expression as the independent variables and classification accuracy as the dependent variable, was conducted to test response accuracy based on the categories derived from the pilot study. For the image presented on a trial, participants gave separate stereotypicality (stereotypical Black, nonstereotypical) and emotion (neutral, threatening) responses. If both of these matched the category values for the sampled image, then this was considered a “correct” response. The accuracy analysis was based on this definition of a correct classification response. Data were collapsed across the three training blocks because there were no significant differences between blocks. Overall classification accuracy—across all

four categories—was 39% ($SD = 7\%$). This accuracy rate was significantly higher than chance (25%), $t(176) = 25.01$, $p < .001$. Although there was no learning component to our classification task (i.e., no training feedback was provided), there was some degree of correspondence between the pilot study ratings and stimulus-category assignments.

We also examined accuracy differences between each of the four categories, which factorially combined two levels of stereotypicality and emotion. We conducted a within-factor ANOVA, with two levels of stereotypicality (nonstereotypical, stereotypical) and two levels of emotional expression (neutral, threatening). There were significant main effects (i.e., differences between levels) on both stimulus dimensions. Average accuracy for neutral expression faces ($M = 45\%$, $SD = 33\%$), collapsed over levels of stereotypicality, was higher than for threatening faces ($M = 33\%$, $SD = 31\%$), $F(176) = 50.47$, $p < .001$, $\eta_p^2 = .22$. Accuracy for stereotypical faces ($M = 54\%$, $SD = 29\%$) was higher, collapsed over levels of emotion, than for nonstereotypical faces ($M = 24\%$, $SD = 29\%$), $F(176) = 105.93$, $p < .001$, $\eta_p^2 = .376$. There was also a significant interaction between stereotypicality and emotion level, $F(176) = 47.29$, $p < .001$, $\eta_p^2 = .212$. Figure 4 displays average accuracy rates for the four categories.

It is clear from these results that accuracy was higher for the stereotypical categories than for the nonstereotypical faces, and that emotional expression did not seem to influence classification accuracy for the stereotypical faces. On the other hand, accuracy was lower overall for nonstereotypical faces, and much lower for nonstereotypical faces with a threatening expression. Understanding the basis for these differences is difficult based solely on accuracy data. In order to understand the relationships between levels of each factor, we conducted signal detection analysis and general recognition theoretic analysis. Both of these provide deeper insight into the classification response patterns across stimulus dimensions.

Fig. 4 Mean accuracy (proportion correct) for each category. Error bars depict standard error of the mean

Signal detection analysis To better understand the ANOVA results, we conducted signal detection analysis on each dimension. Specifically, we computed signal detection measures of discriminability (d') and decision criterion ($\ln\beta$) for the emotion dimension, separately for the nonstereotypical faces and for the stereotypical faces. Similarly, we computed signal detection measures for the stereotypicality dimension separately for the neutral-expression faces and the threatening-expression faces.

The left side of Fig. 5 illustrates average decision criterion placement ($\ln\beta$) and ability to discriminate (d') between emotions for nonstereotypical (top) and stereotypical (bottom) faces. Participants were better able to discriminate emotional expressions for nonstereotypical (d' : $M = 1.68$, $SD = 1.01$) than for stereotypical (d' : $M = 1.18$, $SD = 0.83$) faces, $t(176) = 5.32$, $p < .001$. For stereotypical faces (bottom), decision criterion placement tended to favor the “threatening” response ($\ln\beta$: $M = 0.36$, $SD = 1.83$) more than for nonstereotypical faces ($\ln\beta$: $M = 1.55$, $SD = 2.29$), $t(176) = 6.86$, $p < .001$.

The right side of Fig. 5 illustrates average decision criterion placement ($\ln\beta$) and ability to discriminate (d') between levels of stereotypicality for neutral (top) and threatening (bottom) expressions. Participants were better able to discriminate stereotypicality for neutral (d' : $M = 1.03$, $SD = 0.99$) than for threatening (d' : $M = 0.72$, $SD = 0.98$) expressions, $t(176) = 3.71$, $p < .001$. For faces with threatening expressions

(bottom), decision criterion placement tended to favor the “stereotypical” response ($\ln\beta$: $M = -1.05$, $SD = 1.92$) more than faces with a neutral expression ($\ln\beta$: $M = -0.79$, $SD = 2.13$), although this difference was not quite large enough to reach statistical significance, $t(176) = 1.45$, $p = .07$.

Relating signal detection measures to classification accuracy

The signal detection analysis sheds light on the accuracy results reported above. The clearest interpretation comes from looking at decision criterion differences across category dimensions. Higher accuracy for stereotypical than for nonstereotypical faces seems to be due to an overall bias (across emotional expressions) toward giving the “stereotypical” response. This response tendency is especially pronounced for the faces showing a threatening expression (i.e., threatening expressions are seen as more stereotypical). In terms of emotion classification, the decision criterion values indicate an overall bias toward responding that a “neutral” expression is present. In the case of stereotypical faces, however, this bias is lessened; for stereotypical faces, participants were more likely to respond “threatening” than for nonstereotypical faces. This would account for the higher accuracy rate for “neutral” than for threatening nonstereotypical faces.

Although the signal detection analyses provides additional insight into the accuracy results, there are questions still left

Fig. 5 Signal detection measures of discriminability (d') and decision criterion ($\ln\beta$) for each dimension. Left: Discriminability and response criterion on emotion for nonstereotypical (top) and stereotypical (bottom)

faces. Right: Discriminability and response criterion on stereotypicality for neutral (top) and threatening (bottom) facial expressions

unanswered. For example, to what degree—if at all—are the stereotypicality and emotion dimensions perceived independently or holistically (for discussions of holistic processing, see Richler et al., 2008; Wenger & Ingvalson, 2002). What is the relationship between signal detection results across feature levels? The signal detection analyses reported here only show discriminability and response bias conditioned upon the individual levels of the orthogonal dimension. GRT enables us to understand how these dimensions interact across levels of both dimensions, both between categories and within each category. We report the results of GRT analysis next.

GRT analysis We performed a GRT with individual differences (GRT-wIND) analysis to further explore the interactions between face stereotypicality and emotional expression. The analyses were carried out in R using the *grtools* package written by Soto (Soto, Zheng, & Ashby, 2017). GRT analysis is concerned with violations of PS, DS, and PI (see Introduction for details). The GRT-wIND model was fit to the data from the 177 individual confusion matrices (simultaneously), and model parameters were adjusted to maximize log-likelihood fit. The model accounted for 98% of the variance in the data ($R^2 = .983$). Figure 6 plots the resulting perceptual distributions associated with each category. The ellipses depict contours of equal likelihood representing the joint perceptual distribution for each category.

The GRT-wIND model uses the data from all participants at once to estimate category means along each dimension as well

Fig. 6 Best fitting configuration of perceptual distributions from Experiment 1. Ellipses are contours of equal likelihood. Marginal distributions illustrate discriminability across factor levels. Scatterplot shows observed response probabilities against model-predicted response probabilities (clustered around a line representing perfect model fit), with each dot representing a single participant's data

as estimates of category variance and covariance. This shared perceptual representation is personalized for each participant, i , by estimating two parameters related to perceptual experience: κ_i , which scales the entire space to reflect global attention to the task (i.e., $\kappa_i > 0$; larger values increase perceptual discriminability by shrinking the category variances to reflect greater attention to the task), and λ_i , which reflects relative attention to the two dimensions (i.e., $0 < \lambda_i < 1$; $\lambda_i > .5$ suggests relatively more attention to emotion, whereas $\lambda_i < .5$ suggests relatively more attention to stereotypicality). For each participant, the model also estimates four decision bound parameters (a slope and intercept for the bound along each dimension) to assess DS. To test for violations of PS, DS, and PI, several versions of the model are fitted that place constraints on parameters to instantiate each type of violation. The most parsimonious model is determined via likelihood ratio tests on the fit values. (Additional model-fitting details for the GRT-wIND model can be found in Soto et al., 2015.)

A critical assumption of the GRT-wIND model is the shared perceptual representation that provides a basis for individual-level modeling results. For each of the individual-participant-level parameters, we verified that all distributions were unimodal (as multiple modes might suggest distinct subgroups of participants in terms of basic perceptual representation of the stimulus set) by conducting dip tests (smallest p value = .29). Across participants, the median value of κ_i was 2.45, the median value of λ_i was .53, and there was a weak positive correlation between κ_i and λ_i values, $r(175) = .16$, $p = .04$. This positive correlation, along with the median $\lambda_i > .5$, suggests relatively greater attention to the emotion dimension, particularly for those participants who were highly focused on the task. However, a chi-squared test on λ_i values indicated there was no reliable difference in the number of participants that paid relatively more attention to the emotion ($n = 96$) or stereotypicality ($n = 81$) dimension, at least in terms of perceptual experience, $\chi^2(1) = 1.27$, $p = .26$.

Figure 6 indicates that the distributions are shifted to the right at the stereotypical-face level for both neutral and threatening expressions. This shift indicates a violation of PS on the emotion dimension, $\chi^2(4) = 27.49$, $p < .001$. In other words, judgments of whether a facial expression was perceived as neutral or threatening depended strongly on whether the face was considered stereotypical or not. The model indicates that stereotypical faces were perceived as more threatening, even if the face image was drawn from the “neutral expression” category. DS was violated on the emotion dimension, $\chi^2(177) = 957.86$, $p < .001$, indicating that decision criterion placement was influenced by level of face stereotypicality. On the emotion dimension, most participants (87%) used a decision bound with a slightly negative slope (average deviation from the DS bound was -18 degrees), indicating that the “threatening” response was more likely for stereotypical than for nonstereotypical faces, $\chi^2(1) = 96.96$, $p < .001$. In other

words, the bias to consider stereotypical faces as more threatening than nonstereotypical faces is driven both by perceptual integration of facial features and emotional expression, as well as a decision bias toward the threatening response for stereotypical faces (relative to nonstereotypical faces). The decision bound slope values seem to corroborate the signal detection results, but the perceptual shift toward “threatening” of the stereotypical categories would not be visible in traditional signal detection analysis.

The GRT-wIND analysis also indicates there was a significant violation of PS on the stereotypicality dimension, $\chi^2(4) = 69.35$, $p < .001$. Participants perceived threatening-expression faces to be more stereotypical than neutral-expression faces (contours on the right side of Fig. 6 are shifted upward regardless of category of origin). The GRT-wIND analysis also indicated a violation of DS on the stereotypicality dimension, $\chi^2(177) = 650.59$, $p < .001$. However, there was not a clear trend in the direction of decision bias (51% had a positively sloped bound), $\chi^2(1) = .05$, $p = .82$. This result suggests that judgments of stereotypicality for different levels of emotional expression may have been more consistently driven by perceptual feature integration than by decision bias.

Finally, the GRT-wIND analysis indicates that there was a violation of PI within each the four categories, $\chi^2(4) = 71.29$, $p < .001$. This is visible in the noncircular shapes of each category ellipse. Although the perception of stimulus dimensions within each category exhibited some nonzero correlation, there was not a clear systematic trend across the categories.

General discussion

The question asked in this study is whether people are biased to perceive a stereotypical Black face as threatening, regardless of the intended emotion expressed, as well as whether there is a decision-level (nonperceptual) component to this bias. Results suggest that stereotypical Black/Afrocentric faces are entangled with threat such that people integrate the two factors when classifying a face. Signal detection analyses (average decision-criterion values) indicated that classification responding for stereotypical faces was more biased toward the “threatening” response than for nonstereotypical faces. General recognition theoretic analysis makes clear that there is a strong perceptual interaction between face type and emotion. This was particularly clear in the rightward shift in perceiving stereotypical faces as “threatening” (Fig. 6), even if they were sampled from the neutral-expression category. The GRT analysis also indicated that decision criteria were more biased toward the “threatening” response at the stereotypical face category level.

There was also a trend toward perceiving faces as more stereotypical when they expressed a threatening emotion. Because there was no clear decisional explanation for this trend, it seems largely driven by visual perception. We believe this is a novel conclusion within the face-type bias literature resulting from the use of GRT analysis. It raises the question as to whether the presentation of stereotypically White faces that show threatening expressions would be rated as having more Afrocentric features.

These results suggest that Black men with Afrocentric features may be perceived as threatening regardless of their facial expression or their demeanor. This biased assessment seems to be embedded in the way face types are interpreted such that people may have difficulty interpreting such faces as not threatening. That is, the more stereotypically Black the face is perceived to be, the more threatening it appears, regardless of expression. This result may be a step toward understanding one reason why the intentions of some people may be misinterpreted leading to reprimands by teachers or even law enforcement officers.

The results in this study support previous findings suggesting that Afrocentric faces are associated with criminality and violence and as such are more likely than non-Afrocentric faces to receive biased judgments (Kleider-Offutt, Bond, & Hegerty 2017b). News media and entertainment outlets often portray Black men as criminals, which may promote the negative association of Black men and crime. However, biased judgment facilitated by these negative images does not explain differences in how people perceive face types of Black men (Afrocentric, non-Afrocentric); negative portrayal of Black men extends to all Black men, not just those with Afrocentric features. What this study reveals is that expressions of anger, which are interpreted as threatening, may be intrinsic in the Afrocentric face. The findings here suggest that men with Afrocentric faces are more likely to be judged as threatening, regardless of their intended emotion, than are men with non-Afrocentric faces. The GRT analysis supports the hypothesis that people are biased to interpret Afrocentric faces as threatening and that threat and this face-type may be perceived holistically (rather than being the product of separate decisions about independently perceived dimensions). However, alternative explanations are possible. The perceptual interactions observed here could be the result of perceptual learning—changes in perceptual sensitivity through direct experiences with threatening expressions, rather than something inherent to Afrocentric faces—or else by some top-down (cognitive) influence of classifying faces on the perceptual experience (Goldstone, 1995, 1998).

It is important to point out that face processing also includes assessing relevant social cues as a means of making behavioral forecasts about others in order to respond appropriately while using few cognitive resources (Adams, Albohn, & Kveraga, 2017). Attunement to facial cues are of such

primary importance that people often make assumptions about personality or other traits based on transient expressions (Knutson, 1996). Moreover, people are so predisposed to respond to facial cues that they will read expression information from neutral faces that are devoid of expression (Zebrowitz, Kikuchi, & Fellous, 2010). Potentially, people unwittingly used face-type cues when judging the emotional expression of faces in this study. Although when rated in the pilot study people determined the emotionless Afrocentric and non-Afrocentric faces to be neutral, when presented for classification in Study 1, the neutral faces may now more often have been judged threatening when also stereotypical. This may occur because people use relative comparison wherein the neutrality of an Afrocentric face is questioned when viewed in light of other neutral non-Afrocentric faces. That is, compared to having just seen and rated a non-Afrocentric face, the Afrocentric face is inherently more threatening.

In the current study, the interpretation of the Afrocentric faces as being threatening may be underpinned by the social category of Black men as criminal; in addition, this face type is characteristic of the category Black male (which may bias the decisional aspect of judgments as found in the current study), which further supports the categorization. The combination of social categorization, and possibly the facial structure of the Afrocentric face being consistent with an expression of anger, may work together to inform the observer that threat is imminent, cues that are present to a lesser degree in non-Afrocentric faces. Research suggests that craniofacial structure can be inherently threatening, (e.g., small eyes, wide nostrils, a pronounced jaw). The reasons for the association to craniofacial structure and apparent aggression is unclear, but theories argue for level of testosterone as a face-type-related indicator (Lefevre, Lewis, Perrett, & Penke, 2013) tied to masculinity and others suggest that certain facial characteristics are similar to facial expressions of anger (Fox & Damjanovic, 2006; Said, Sebe, & Todorov, 2009). Emotion cues that signify anger or assist in avoidance of imminent danger are adaptive. Certain craniofacial features are consistently associated with aggressive attitudes and thus influence the speed and accuracy of emotion recognition. Facial height-to-width ratio (FHWR), for example, influences how readily faces are judged as angry such that high compared to low FHWR facilitates aggression judgments (Deska et al., 2017). Although it is clear that social categories can influence how people interpret facial expressions (Hugenberg & Wilson, 2013), facial structure may also facilitate certain interpretations. Afrocentric faces may retain some characteristics that exist in other measures of face structure that also are interpreted as a cue of aggressive behavior. The GRT analysis seems to indicate that this is largely the result of integration of feature information at a perceptual level but also that there is a decisional component for interpreting emotions at different degrees of Afrocentric stereotypicality. This decisional bias

did not seem to exist along the stereotypicality dimension, because threatening faces seemed to be perceived as more stereotypical.

Results reveal that people differentially interpret the facial expression of subgroups of Black men based on facial structure and that this biased interpretation is driven by distinct perception and decision contributions. Perceptual and decisional components of the observed bias may be differentially responsive to raising conscious awareness of the bias or via experience-based perception that may impact everyday communications. From an applied standpoint, Black men with Afrocentric features may be viewed as aggressive in a variety of contexts, which could influence their treatment by teachers, employers, law enforcement officers, and, potentially, jurors. Because interpreting facial cues is such an automatic part of communication, poor first impressions could negatively influence many aspects of everyday life.

Direct application to everyday communication processes could be limited because this task was laboratory-based with no context in which to judge facial expressions. Context may influence the findings such that faces presented in a positive setting (e.g., classroom, church) may be less likely to be judged as threatening, regardless of face type. Future directions include testing whether people judge Afrocentric face types as angry as opposed to threatening, which may speak to the craniofacial cues that may also exist in these faces. This study is among the first to employ advanced modeling techniques to suggest that perceptual cues to threat are inherent in Afrocentric faces.

A potential limitation on our current conclusions stems from the relatively low accuracy rates achieved in the classification task. We accepted this in the current study, though, as we hoped to use the GRT model to measure the influence of distinct perceptual and decisional processes on a stimulus set composed of noisy, real-world images. Another potential limit on our conclusions is that perceptual bias (interaction between facial features and emotional expression) may have influenced the pilot study ratings and then carried over into category construction in Study 1. If there is perceptual integration of facial features and emotional expression, then it seems likely this would be reflected in the ratings (even though the pilot study participants rated only a single dimension). In a future study, it might be possible to generate a much larger set of rating stimuli and take steps to remove the interaction from the underlying category distributions. But it would likewise be difficult to know how these highly controlled categories relate to real-world face classification. Our position is that the GRT analysis should describe interactions that exist between stimulus dimensions, and in addition help us to understand whether the interactions are based in perception, decision-making, or both. The current results help provide insight into the origin of these response-level interactions.

References

- Adams, R. J., Albohn, D. N., & Kveraga, K. (2017). Social vision: Applying a social-functional approach to face and expression perception. *Current Directions in Psychological Science*, 26(3), 243–248. <https://doi.org/10.1177/0963721417706392>
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1), 33–53. <https://doi.org/10.1037/0278-7393.14.1.33>
- Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, 120(2), 150–172. <https://doi.org/10.1037/0096-3445.120.2.150>
- Ashby, F. G., & Soto, F. A. (2015). Multidimensional signal detection theory. In J. R. Busemeyer, J. T. Townsend, Z. J. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 13–34). New York, NY: Oxford University Press.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2), 154–179.
- Bijlstra, G., Holland, R. W., Dotsch, R., Hugenberg, K., & Wigboldus, D. J. (2014). Stereotype associations and emotion recognition. *Personality and Social Psychology Bulletin*, 40(5), 567–577. <https://doi.org/10.1177/0146167213520458>
- Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal Of Personality And Social Psychology*, 83(1), 5–25. <https://doi.org/10.1037/0022-3514.83.1.5>
- Blair, I. V., Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology*, 35(1), 59–68.
- Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, 15(10), 674–679.
- Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, 87(6), 763–778. <https://doi.org/10.1037/0022-3514.87.6.763>
- Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, 20(10), 1194–1198.
- Carré, J. M., Morrissey, M. D., Mondloch, C. J., & McCormick, C. M. (2010). Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic? *Perception*, 39(3), 356–377. <https://doi.org/10.1068/p6543>
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n -way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3), 283–319.
- Colarelli, S. M., Poole, D. A., Unterborn, K., & D’Souza, G. C. (2010). Racial prototypicality, affirmative action, and hiring decisions in a multiracial world. *International Journal of Selection and Assessment*, 18(2), 166–173.
- Cooper, H. L. (2015). War on drugs policing and police brutality. *Substance Use & Misuse*, 50(8/9), 1188–1194.
- Cornes, K., Donnelly, N., Godwin, H., & Wenger, M. J. (2011). Perceptual and decisional factors influencing the discrimination of inversion in the Thatcher illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 645–668.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer’s dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314–1329. <https://doi.org/10.1037/0022-3514.83.6.1314>
- Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to ‘prejudice’. *Journal of Personality and Social Psychology*, 88(5), 770–789. <https://doi.org/10.1037/0022-3514.88.5.770>
- Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2017). The face of fear and anger: Facial width-to-height ratio biases recognition of angry and fearful expressions. *Emotion*. <https://doi.org/10.1037/emo0000328>
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5–18. <https://doi.org/10.1037/0022-3514.56.1.5>
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, 21(11), 1139–1150.
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, 17(5), 383–386.
- Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personal Social Psychology*, 87(6), 876–893. <https://doi.org/10.1037/0022-3514.87.6.876>
- Farris, C., Viken, R. J., & Treat, T. A. (2010). Perceived association between diagnostic and non-diagnostic cues of women’s sexual interest: General recognition theory predictors of risk for sexual coercion. *Journal of Mathematical Psychology*, 54(1), 137–149.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). New York, NY: McGraw-Hill.
- Fiske, S. T. (2004). Intent and ordinary bias: Unintended thought and social motivation create casual prejudice. *Social Justice Research*, 17(2), 117–127. <https://doi.org/10.1023/B:SORE.0000027405.94966.23>
- Fox, E., & Damjanovic, L. (2006). The eyes are sufficient to produce a threat superiority effect. *Emotion*, 6(3), 534–539. <https://doi.org/10.1037/1528-3542.6.3.534>
- Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, 6(5), 298–304.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49(1), 585–612.
- Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). The independent effects of skin tone and facial features on Whites’ affective reactions to Blacks. *Journal Of Experimental Social Psychology*, 48(4), 892–898. <https://doi.org/10.1016/j.jesp.2012.02.001>
- Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13, 70–84.
- Haselton, M. G., & Buss, D. M. (2003). Biases in social judgment: Design flaws or design features? In J. P. Forgas, K. D. Williams, W. von Hippel, J. P. Forgas, K. D. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 23–43). New York, NY: Cambridge University Press.
- Hinzman, L., & Kelly, S. D. (2013). Effects of emotional body language on rapid out-group judgments. *Journal of Experimental Social Psychology*, 49(1), 152–155. <https://doi.org/10.1016/j.jesp.2012.07.010>
- Hugenberg, K. (2005). Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion*, 5(3), 267–276. <https://doi.org/10.1037/1528-3542.5.3.267>
- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, 14(6), 640–643. <https://doi.org/10.1046/j.0956-7976.2003.psci.1478.x>
- Hugenberg, K., Wilson, J. P., See, P. E., & Young, S. G. (2013). Towards a synthetic model of own group biases in face memory. *Visual*

- Cognition*, 21(9–10), 1392–1417. <https://doi.org/10.1080/13506285.2013.821429>
- Kleider, H. M., Cavrak, S. E., & Knuycky, L. R. (2012). Looking like a criminal: Stereotypical black facial features promote face source memory error. *Memory & Cognition*, 40(8), 1200–1213. <https://doi.org/10.3758/s13421-012-0229-x>
- Kleider-Offutt, H. M., Knuycky, L. R., Clevinger, A. M., & Capodanno, M. M. (2017a). Wrongful convictions and prototypical black features: Can a face-type facilitate misidentifications? *Legal and Criminological Psychology*, 22(2), 350–358. <https://doi.org/10.1111/lcrp.12105>
- Kleider-Offutt, H. M., Bond, A. D., & Hegerty, S. E. (2017b). Black stereotypical features: When a face type can get you in trouble. *Current Directions in Psychological Science*, 26(1), 28–33.
- Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20(3), 165–182. <https://doi.org/10.1007/BF02281954>
- Knuycky, L. R., Kleider, H. M., & Cavrak, S. E. (2014). Line-up misidentifications: When being 'prototypically Black' is perceived as criminal. *Applied Cognitive Psychology*, 28(1), 39–46. <https://doi.org/10.1002/acp.2954>
- Kubota, J. T., & Ito, T. A. (2014). The role of expression and race in weapons identification. *Emotion*, 14(6), 1115–1124. <https://doi.org/10.1037/a0038214>
- Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: Facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, 34(4), 273–279. <https://doi.org/10.1016/j.evolhumbehav.2013.03.005>
- Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology*, 47(2), 391–396.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4), 1122–1135
- Mack, M. L., Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Indecision on decisional separability. *Psychonomic Bulletin & Review*, 18(1), 1–9.
- Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, 66(1), 37–47. <https://doi.org/10.1037/0022-3514.66.1.37>
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Attention, Perception, & Psychophysics*, 53(1), 49–70.
- Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 328–342.
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9(2), 260–264. <https://doi.org/10.1037/a0014681>
- Shasteen, J. R., Sasson, N. J., & Pinkham, A. E. (2015). A detection advantage for facial threat in the absence of anger. *Emotion*, 15(6), 837–845. <https://doi.org/10.1037/emo0000090>
- Silbert, N. H., & Thomas, R. D. (2013). Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review*, 20(1), 1–20.
- Smith, J. S., LaFrance, M., & Dovidio, J. F. (2017). Categorising inter-sectional targets: An 'either/and' approach to race- and gender-emotion congruity. *Cognition and Emotion*, 31(1), 83–97. <https://doi.org/10.1080/02699931.2015.1081875>
- Soto, F. A., Vucovich, L., Musgrave, R., & Ashby, F. G. (2015). General recognition theory with individual differences: A new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review*, 22(1), 88–111.
- Soto, F. A., Zheng, E., & Ashby, F. G. (2017). Testing separability and independence of perceptual dimensions with general recognition theory: A tutorial and new R package (grtools). *Frontiers in Psychology*, 8, 696. <https://doi.org/10.3389/fpsyg.2017.00696>
- Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, 34(10), 1332–1345. <https://doi.org/10.1177/0146167208321269>
- Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M., & Lee, H. K. (2012). Skin and bones: The contribution of skin tone and facial structure to racial prototypicality ratings. *PLOS ONE*, 7(7), e41193. <https://doi.org/10.1371/journal.pone.0041193>
- Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Attention, Perception, & Psychophysics*, 63(4), 625–650.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43(2), 161–204.
- Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5), 872–892.
- Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1106–1118.
- Wilkins, C. L., Kaiser, C. R., & Rieck, H. (2010). Detecting racial identification: The role of phenotypic prototypicality. *Journal of Experimental Social Psychology*, 46(6), 1029–1034.
- Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, 113(1), 59–80. <https://doi.org/10.1037/pspi0000092>
- Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological and Personality Science*, 7(4), 331–338.
- Zebrowitz, L. A., Kikuchi, M., & Fellous, J. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, 98(2), 175–189. <https://doi.org/10.1037/a0017990>