American Psychological Association
Journal of Experimental Psychology: General

0096-3445/18/$12.00
2018, Vol. 147, No. 5, 683–701
http://dx.doi.org/10.1037/xge0000443

Rapid Visual Perception of Interracial Crowds: Racial Category Learning From Emotional Segregation

Sarah Ariel Lamer and Timothy D. Sweeny
University of Denver

Michael Louis Dyer
Hamilton College

Max Weisbuch
University of Denver

Drawing from research on social identity and ensemble coding, we theorize that crowd perception provides a powerful mechanism for social category learning. Crowds include allegiances that may be distinguished by visual cues to shared behavior and mental states, providing perceivers with direct information about social groups and thus a basis for learning social categories. Here, emotion expressions signaled group membership: to the extent that a crowd exhibited *emotional segregation* (i.e., was segregated into emotional subgroups), a visible characteristic (race) that incidentally distinguished emotional subgroups was expected to support categorical distinctions. Participants were randomly assigned to view interracial crowds in which emotion differences between (black vs. white) subgroups were either small (*control condition*) or large (*emotional segregation condition*). On each trial, participants saw crowds of 12 faces (6 black, 6 white) for roughly 300 ms and were asked to estimate the average emotion of the entire crowd. After all trials, participants completed a racial categorization task and self-report measure of race essentialism. As predicted, participants exposed to emotional segregation (vs. control) exhibited stronger racial category boundaries and stronger race essentialism. Furthermore, such effects accrued via ensemble coding, a visual mechanism that summarizes perceptual information: emotional segregation strengthened participants' racial category boundaries to the extent that segregation limited participants' abilities to integrate emotion *across* racial subgroups. Together with evidence that people observe emotional segregation in natural environments, these findings suggest that crowd perception mechanisms support racial category boundaries and race essentialism.

Keywords: crowd perception, ensemble coding, essentialism, racial categorization, social vision

Supplemental materials: http://dx.doi.org/10.1037/xge0000443.supp

The tendency to sharply distinguish among different 'categories' of people, such as racial categories, fundamentally shapes the thoughts and feelings people have about one another. People quickly classify individuals as Black or White, woman or man, child or adult, for example, and draw inferences about those individuals accordingly (Ito & Urland, 2003; Montepare & Zebrowitz, 1998; Wiese, Schweinberger, & Neumann, 2008). The tendency to quickly and rigidly categorize others appears to be especially strong for some characteristics, and people often believe that those social categories describe the inherent "essence" of people. Why?

One possible explanation is that people have biologically inherited a tendency to make distinctions about specific human characteristics. Yet although humans may inherit tendencies to categorize others on a few dimensions (e.g., age, sex; Kinzler, Shutts, & Correll, 2010; Kurzban, Tooby, & Cosmides, 2001; Rhodes & Gelman, 2009), most of the categorical distinctions that people make—including racial distinctions—are unlikely to be inherited. For example, our human ancestors traveled by foot and were thus unlikely to encounter individuals of other races, so an evolutionarily inherited mechanism for categorizing race is unlikely (Eerkens, 1999; Kelly, 1995, 2003; Weber et al., 2011).

Instead, developmental and evolutionary psychologists have argued that most social categorical distinctions are scaffolded by existing cognitive mechanisms (i.e., exapted; Gould & Vrba, 1982) for the purpose of group living (Bigler & Liben, 2007; Cosmides, Tooby, & Kurzban, 2003). We argue here that one class of mechanism operates on human crowds to select specific characteristics as bases of social categorization. This mechanism capitalizes on the association between visible cues and group membership, and on the rich information about social

Sarah Ariel Lamer and Timothy D. Sweeny, Department of Psychology, University of Denver; Michael Louis Dyer, Department of Neuroscience, Hamilton College; Max Weisbuch, Department of Psychology, University of Denver.

This research was supported by an NSF GRFP (DGE-1104602).

Correspondence concerning this article should be addressed to Sarah Ariel Lamer, Department of Psychology, University of Denver, 2155 South Race Street, Denver, CO 80210. E-mail: sarahariellamer@gmail.com

groups uniquely available in crowds.[1] We theorize that rapid visual processing of human crowds provides an efficient means for people to learn social category distinctions, and in turn, to infer that certain people are inherently different from each other. In what follows, we develop this theory, test it with regard to racial categorization, and examine whether such patterns may exist in interracial crowds.

Crowd Perception: Relevance to Intergroup Relations

Our theory is rooted in the idea that crowd perception is an especially powerful means for learning about small groups and alliances. First, crowd perception enables perceivers to see people engage in behavior or communication together, even simultaneously, at the level of the collective (Elias, Dyer, & Sweeny, 2017; Haberman & Whitney, 2007, 2009; Phillips, Weisbuch, & Ambady, 2014; Sweeny & Whitney, 2014). For example, behavioral coordination is typical of groups (Campbell, 1958) and allegiances (Barsade, 2002; Bernieri & Rosenthal, 1991; Likowski, Mühlberger, Seibt, Pauli, & Weyers, 2008) and such coordination can be observed from a glance at a crowd (e.g., Sweeny, Haroz, & Whitney, 2013). By observing patterns of coordination, perceivers may learn to distinguish among different groups in a crowd. Yet behavioral coordination is only one of several properties that define social groups to perceivers (Campbell, 1958; Lickel et al., 2000), and crowds provide visible cues to these properties. For example, Turner (1984) suggests that social groups can be defined by homogeneity among its members and heterogeneity from other collections of persons. Crowd perception enables people to evaluate these patterns of homogeneity directly and rapidly, without requiring inspection and deliberate comparison of individuals one at a time (Phillips et al., 2014; Whitney, Haberman, & Sweeny, 2014). Hence, the properties that help define groups to perceivers may be visible in crowds. Finally, Tajfel (1969; see others, including Bigler, Jones, & Lobliner, 1997; Kurzban et al., 2001) argued that people will treat visible characteristics as group identities to the extent that those characteristics consistently distinguish one social group from another. For example, perceivers may notice that behavioral coordination in crowds covaries with race, and consequently learn that racial categories are indicative of meaningful social groups. Hence, crowd perception uniquely enables perceivers to immediately and directly perceive behavioral coordination and homogeneity, and to *additionally* learn the stable social cues (e.g., race) that distinguish people who behave together versus separately.

The importance of crowd perception to social categorization is foreshadowed in Oakes and colleagues' (1991) treatment of social identity theory: "social categorizations fit those aspects of social reality characterized by the distinctive, *emergent* properties of group relationships and collective action" (p. 126; italics added). Emergent properties of groups (properties that can be observed in a group but not an individual) can be efficiently perceived within crowds. For example, separate groups within a crowd can exhibit unique patterns of coordinated behavior, expressions, and other cues. If those emergent group patterns (e.g., coordinated expression of emotion) are associated with an incidental but visible characteristic (e.g., racial cues), perceivers might then attribute the possession of that characteristic to group membership. We theorized that this process could occur with respect to the grouping of emotion expressions in crowds.

Emotional Segregation in Crowd Perception: A Basis for Social Category Distinctions

We focused here on group emotion expressions because emotion expressions provide perceivers with information about social grouping. Specifically, laypersons (a) assume that shared mental states (e.g., emotions) are characteristic of groups (Lickel et al., 2000), (b) are more likely to adopt the emotions of ingroup than outgroup members (Barsade, 2002; Weisbuch & Ambady, 2008), and (c) assign shared group identity to crowds that exhibit similar facial emotion (Magee & Tiedens, 2006). Equally as important, visual processes involved in crowd perception are sensitive to emotion expressions (e.g., Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2009, 2010).

To the extent that crowds are consistently segregated into emotional subgroups (i.e., the crowd exhibits *emotional segregation*), visible characteristics (e.g., race) that coincide with those subgroups should become a basis for perceivers to identify meaningful social groups. Extensive exposure to emotional segregation by race may thus cause perceivers to believe that racial categories represent group "entities" suggestive of a deep essence (Campbell, 1958; Lickel et al., 2000; Yzerbyt, Corneille, & Estrada, 2001) naturally inherited by group members. Hence, repeated perceptions of emotional segregation between different races should lead perceivers to believe (a) that racial identities constitute mutually exclusive categories and (b) that groups have naturally inherited allegiances.

The main purpose of the current research was to test these predictions in what we believe to be the first experiment examining how crowd perception can shape beliefs about race. Earlier we argued that crowd characteristics can be observed in a glance, and such efficient processing may provide a functional benefit for social category learning. Accordingly, the influence of crowd perception on racial cognition should emerge even when participants only have enough time to glance at a crowd. This leads to our main hypotheses:

> *Hypothesis 1:* Brief exposures to emotional segregation (by race) will cause perceivers to make sharper categorical distinctions between races.
>
> *Hypothesis 2:* Brief exposures to emotional segregation (by race) will cause perceivers to essentialize racial categories.
>
> *Hypothesis 3:* The relationship described in Hypothesis 1 will mediate the relationship described in Hypothesis 2.

[1] We here define crowds as groups of more than two individuals. Although it could be argued that a crowd includes as few as two people, it could also be argued that dyadic interactions relevant to mating or mutual exchange have unique qualities distinct from crowds. So although it is possible that our results would apply to dyads, we aimed to avoid conflating dyads and crowds in our materials, and therefore used the more conservative definition of $n > 2$.

Ensemble Coding and the Gestalt Approach to Visual Perception

Hypotheses 1–3 regard the influence of emotional segregation on racial categorization and essentialism. The theorizing that underlies these hypotheses includes an assumption that when emotional segregation is present in crowds, people should be able to see it. We thus sought to strengthen our theory by exploring the role of a specific visual mechanism in mediating the above hypothesized effects. We theorized that those effects would emerge by virtue of an efficient mechanism involved more broadly in visual grouping. Indeed, the question of how people group perceptual information can be traced to Aristotle and was first empirically examined with the emergence of Gestalt principles of perception (Wertheimer, 1938). Gestalt principles are well known for their utility in explaining when individual objects are likely to be perceived as a group (Palmer, 1999; Wagemans et al., 2012; Wertheimer, 1938). Although originally applied to the perception of objects (e.g., patterns, shapes), Gestalt principles of grouping can also be mapped onto "people perception" (Phillips et al., 2014). Campbell (1958), for example, argued that homogeneity in human movement (*common fate*), appearance (*similarity*), and location (*proximity*) defined social groups, and research (e.g., Lickel et al., 2000) suggested that laypersons use Campbell's principles to linguistically characterize social groups. Campbell's principles should also apply to the visual perception of crowds, and although a variety of visual mechanisms may explain the influence of crowd perception on racial cognitions, here we isolate one: *ensemble coding*. We explore ensemble coding as a likely candidate for mediating our proposed effects because (a) it is known to operate rapidly and (b) it is sensitive to principles of grouping (Corbett, 2017). Indeed, ensemble coding is sensitive to crowd characteristics that are vital for evaluating groups of people, such as homogeneity in movement (Sweeny, Haroz, et al., 2013), identity, and facial expression (Whitney et al., 2014).

Ensemble coding is a mechanism that extracts statistical summaries of large amounts of perceptual information, enabling people to understand a crowd of people or a collection of objects in terms of its overall or gist properties (Alvarez, 2011; Phillips et al., 2014; Whitney et al., 2014). Research on ensemble coding extends beyond Gestalt principles to examine not only which objects should be grouped, but also how information from those individuals is pooled into a coherent percept. Ensemble coding is flexible, occurring for a range of simple and complex features across the visual hierarchy. For example, perceivers are able to determine the average location (Hess & Holliday, 1992; Morgan & Glennerster, 1991; Whitaker, Mcgraw, Pacey, & Barrett, 1996), orientation (Dakin, 2001), motion (Watamaniuk, Sekuler, & Williams, 1989), size (Ariely, 2001), facial emotion (Haberman et al., 2009; Haberman & Whitney, 2009), gaze direction (Sweeny & Whitney, 2014), and biological motion (Sweeny, Haroz, et al., 2013) of briefly presented crowds.

Critically for the current research, ensemble coding is also rapid—it can generate high-fidelity summary percepts of a crowd in as little as one-twentieth of a second (Haberman & Whitney, 2009). A consequence of this process is that information about individuals is typically lost for conscious report (e.g., Haberman & Whitney, 2007, 2009), albeit for the good of gaining a strikingly precise representation of the group (Sweeny et al., 2013). In fact, by averaging noisy representations of each member of a set, people achieve ensemble judgments of a group or crowd that are often even more precise than judgments of its individuals (Alvarez, 2011; Elias et al., 2017; Haberman & Whitney, 2007; Sweeny, Suzuki, Grabowecky, & Paller, 2013).

Ensemble Coding of Emotionally Segregated Crowds

The engagement of ensemble coding mechanisms during crowd perception may be important for perceivers to learn that categorical distinctions (e.g., race) are associated with emotionally segregated subgroups. Indeed, ensemble codes are capable of directly representing a crowd's variability (Haberman, Lee, & Whitney, 2015), and the precision of summary representations (e.g., a crowd's average) can depend on heterogeneity naturally introduced by the presence of distinct subgroups (de Gardelle & Summerfield, 2011; Hubert-Wallander & Boynton, 2015; Marchant, Simons, & de Fockert, 2013; Maule, Witzel, & Franklin, 2014; McDonnell, Larkin, Dobbyn, Collins, & O'Sullivan, 2008; Sweeny, Haroz, et al., 2013; Sweeny & Whitney, 2014; Utochkin, 2015; Utochkin & Tiurina, 2014). In this way, the summary percepts that emerge from ensemble coding are very much like the statistical mean of a data sample, and in both cases error surrounding the mean increases with heterogeneity. Accordingly, when the only source of crowd heterogeneity is subgrouping, the precision of summary percepts can provide an index of the extent to which perceivers process subgroups.

We capitalized on this relationship in the current experiment. Specifically, we ensured that subgrouping was the only source of differences in crowd heterogeneity between experimental conditions. One experimental condition—the *control* condition—included emotional subgroups which were fairly similar to each other in emotional intensity. The other experimental condition—the *emotional segregation* condition—included emotional subgroups which were quite different from each other in emotional intensity. Critically, within-subgroup heterogeneity was equivalent between the two conditions. The emotional segregation condition thus differed from the control condition in between-subgroup heterogeneity but not within-subgroup heterogeneity. In an additional experiment, we controlled for the possibility that the increased overall heterogeneity in the emotional segregation condition could be sufficient for influencing perceivers' racial category beliefs (see the Discussion section). However, in Study 1, we tested the idea that systematic covariation between racial characteristics and shared emotion influences perceivers' racial category beliefs.

Intentional Versus Unintentional Subgrouping

This indirect measure of subgrouping in summary perception was preferred over a simpler measure. Specifically, we could have asked participants for judgments of the emotional subgroups, rather than the entire crowd. This would have yielded direct measures of subgroup summary percepts. However, this procedure would have introduced intentional subgrouping rather than incidental subgrouping. By asking participants for judgments about the entire crowd instead, we could assume that if participants engaged in any cognition or learning about the association between race and subgrouping, it would likely have been incidental.

Ensemble Coding Hypotheses

In summary, we designed our experimental conditions in a manner that forced us to make a specific interpretation of participants' summary percepts. Specifically, the design ensured that the precision of summary percepts inversely indexed the degree to which participants encoded emotional differences (i.e., heterogeneity) between subgroups. If ensemble coding does indeed play a role in the hypothesized effects of crowd perception on racial cognition, then we should observe that (a) ensemble coding mechanisms are active during interracial crowd perception, (b) summary percepts are less precise for the emotional segregation condition than for the control condition, and most importantly, (c) reduced precision of ensemble coding in the emotional segregation condition partially explains the effect of emotional segregation on racial category distinctions and race essentialism. Although "a" and "b" describe basic properties of ensemble coding, we list them as Hypotheses 4–5 below, to test whether ensemble coding processes are operating as expected. Hypothesis 6 is the hypothesis most unique to our theory and regards the indirect effect of condition through visual processes ("c" above).

> *Hypothesis 4:* Perceivers will exhibit ensemble coding of interracial crowds of faces, with higher fidelity visual representations of group than individual emotion expressions.

> *Hypothesis 5:* Perceivers will generate lower-fidelity summary percepts for interracial crowds with greater emotional segregation.

> *Hypothesis 6:* Emotional segregation will have an indirect effect on perceivers' racial category distinctions via summary perceptual accuracy (i.e., ensemble coding will partially mediate the effect of emotional segregation on racial category distinctions and race essentialism).

Current Research

We theorized that the perception of emotional segregation in crowds would influence perceivers' social category representations. We tested hypotheses drawn from this theory in the context of racial categorization. Hence, to the extent that racial cues covary with emotional segregation, and to the extent that ensemble coding processes are sensitive to emotional subgroups, perceivers may strengthen their belief that race is categorical and distinguishes groups based on meaningful characteristics.

We chose to use race as an incidental visible cue to pair with emotional segregation because (a) racial categories are not biologically given either in fact (Goodman, 2000; Graves, 2001; Lewontin, 1972; Long & Kittles, 2009; Nei & Roychoudhury, 1993) or in perception (e.g., Kurzban et al., 2001) so learning must contribute to racial categorization; (b) people vary in the degree to which they categorize by race (Eberhardt, Dasgupta, & Banaszynski, 2003; Fazio & Dunton, 1997; Stangor et al., 1992); (c) people vary in the degree to which they essentialize race (Chao, Hong, & Chiu, 2013; Plaks, Malahy, Sedlins, & Shoda, 2012), that is the degree to which they believe racial identity is biological in nature, cannot be changed, is universally recognized, and has existed throughout human history (Williams & Eberhardt, 2008); and (d) racial categorization is an antecedent to racism and racial conflict, both of which are relevant to contemporary societal issues.[2]

We thus conducted a test of our theory on the role of crowd perception in social categorization by focusing on expressive (emotional) cues to alliances and visible cues to race. All participants viewed 216 crowds of 12 faces, each consisting of two emotional subgroups that were distinguished by race (Black vs. White). Participants were randomly assigned to a control condition or an emotional segregation condition—these conditions describe the degree of emotional segregation presented to participants (the control condition is a conservative control, in that it contains slight emotional segregation). On each trial, participants identified the average emotion of the *entire* crowd of faces. Following this procedure, participants completed measures of racial categorization and race essentialism. We expected participants in the emotional segregation condition to draw sharper distinctions between races (Hypothesis 1) and endorse race essentialism to a greater degree (Hypothesis 2) than participants in the control condition, with the latter effect mediated by the former (Hypothesis 3). Further, we expected ensemble coding mechanisms to effectively operate on interracial crowds (Hypothesis 4), to produce less precise summary percepts for the emotional segregation than control condition (Hypothesis 5), and for this latter effect to serve as a precondition for (to mediate) the indirect effect of emotional segregation on race essentialism via racial categorization (Hypothesis 6).

Study 1: Influence of Crowd Perception on Racial Cognition

Method

Participants and setting. Participants were recruited from an undergraduate participant pool and received partial class credit for their participation. The study was approved by the institutional review board at the University of Denver. The experiment was conducted on computers using MATLAB and the Psychophysics ToolBox (Brainard, 1997), with each computer located in its own room. The final sample consisted of 148 participants (70% women), including 115 White, 13 Asian, three Latina(o), eight mixed-race, four Black, and four Middle Eastern participants (one participant declined to list race) ranging in age from 18–35.[3] This sample size reflects a priori power analyses. Specifically, we examined effect sizes observed in studies with similar racial categorization tasks ($.61 < d < .78$; Castano, Yzerbyt, Bourguignon, & Seron, 2002; Peery & Bodenhausen, 2008) and measures of race essentialism (i.e., the Biological Conceptions of Race scale: $.57 < d < .96$; Sanchez, Young, & Pauker, 2015; Young, Sanchez, & Wilton, 2013). These studies yielded medium to large conditional

[2] Although people may be especially ready to categorize on the basis of cues that look inheritable (e.g., race; Gil-White, 2001), thus limiting our ability to draw firm conclusions from the current study about more transitory cues (e.g., shirt color; Kurzban et al., 2001), the fact that racial categories are well-learned in the tested culture may work *against* our hypotheses in the sense of being potentially more difficult to change than other types of categories.

[3] Not included in this final sample were participants who were mistakenly-assigned a different condition in each block ($n = 3$), were minors ($n = 1$), or who did not finish the study ($n = 6$).

effect sizes. Because the experimental manipulation was subtle (see below), we erred on the side of the smaller of the observed effect sizes and thus conducted a power analysis with a medium effect size (d = .50). Using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), this analysis suggested a sample size of 128 participants randomly assigned to two between-subjects conditions would achieve 80% power. However, because we intended to conduct a bootstrapped mediation analysis, we sought a sample size that would accommodate such mediation at 80% power. With two medium effect sizes, the recommended sample size to achieve 80% power with a bias-corrected bootstrapped analysis is 71 per condition or 142 participants (Fritz & Mackinnon, 2007). We just met this larger sample size with exclusions. Additionally, we report analyses on the first 128 participants in the supplementary materials (results and significance levels are nearly identical to that observed with the full N).

Stimulus generation. We selected faces from the NimStim set of facial images, which includes adults of different races posing a number of different facial emotions (Tottenham et al., 2009). For each face, we generated a "morph wheel" with prototypical angry, fearful, and happy expressions as anchor points ("parent" faces; see Figure 1). We then created a sequence of morphs with 50 increments from a neutral expression to the full expression of each emotion. For each identity, each emotion sequence progressed from neutrality to either angry, fearful, or happy, with 50 images per sequence. We selected these three emotions to control for emotional arousal (all emotions were high arousal) and to include emotions that minimize facial artifacts during morphing. We selected closed-mouth expressions to minimize ghosting effects caused by visible teeth during the morphing process. For this face set we held gender constant and restricted the stimulus set to male faces.

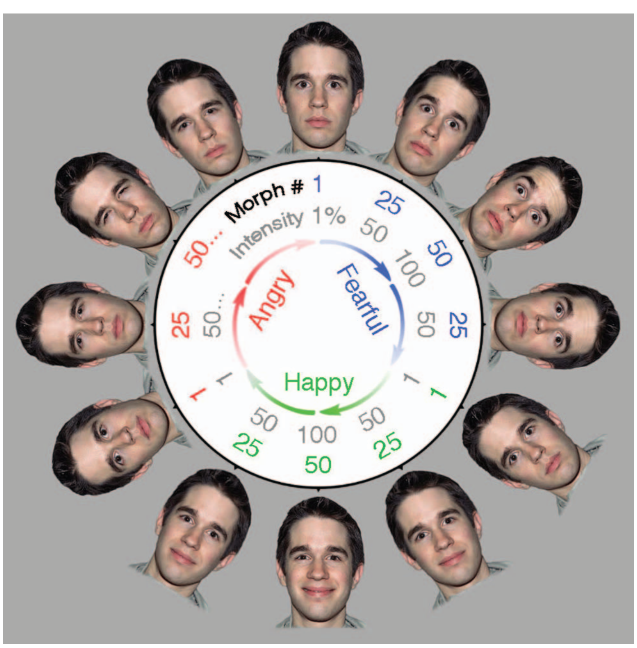


Figure 1. Example of the emotion morph space for one actor. The wheel contained 300 faces in total. See the online article for the color version of this figure.

For each identity, we created emotional morphs from neutral to happy, neutral to angry and neutral to fearful. We first replaced the background of each parent face with uniform gray (R/B/G = 170/170/170). Then, using Fantamorph software (v5.4.2), we morphed between (a) neutral and fearful (N-F; b) neutral and happy (N-H), and (c) neutral and angry (N-A). We matched analogous points on two parent faces (e.g., corners of the mouth, tip of the nose for both neutral and angry) of the same identity and 48 morphs were linearly interpolated. By creating 50 images between each morphing pair, we had a sufficient number of faces to allow for close perceptual matching of emotional intensity between identities. Each transition from neutral toward each parent face could be mirror reversed (e.g., from N to A and from A back to N). With the inclusion of these duplicate, mirror-reversed faces, each facial identity wheel included 300 morphed images (see Figure 1). Morph wheels are advantageous in method-of-adjustment tasks in which a participant responds by adjusting a facial image (Haberman & Whitney, 2007). With no end-points, participants must cycle through the entire wheel and cannot use a single face as a frame of reference. Furthermore, the lack of endpoints reduced the risk of response compression (Sweeny, Suzuki, et al., 2013). With eight identities, our face set included a total of 2,400 images. Three hundred faces made up the face wheel for each of eight identities, and 1,184 of these 2,400 faces were unique: one neutral, 49 angry, 49 fearful, 49 happy × 8 identities.

Pretest: Matching on emotional intensity. We first evaluated our stimulus set by measuring the change in perceived emotional intensity over change in objective morph units. In a pretest, eight participants viewed randomly selected faces from the full set, one at a time, for 506 ms each. The procedure for this pretest is detailed in Figure 2. The critical outcome for each identity from the set was the slope derived from the linear relationship between perceived emotion intensity (1–10) and objective morph unit (1–50). For each of eight identities, we fit linear functions to the relationship between perceived and objective emotion intensity, separately for each emotion. Critically, all slopes were positive: emotion intensity ratings increased by 1% for every ~1.5% increase in objects' morph units (slopes were greater than .06 and less than .14). indicating that (for each identity and emotion) as morph units increased, so too did perceived intensity. To examine any differences by emotion and race, we also fit linear functions to the data collapsed across race and emotion, respectively. The confidence intervals for the slopes and intercepts of happy, angry, and fearful morphs overlapped, as they did for the black and white morphs, indicating that there were no differences in perceived intensity by race or emotion. Based on the resulting slopes describing the relationship between objective and subjective emotion expression, we identified the two most expressive identities (one black, one white) to use as the response faces for the trials. Specifically, these two identities had the steepest slopes across the three emotions and included equivalent and wide ranges of perceived emotional intensity. Use of these identities as the response faces thus limited floor and ceiling effects. The remaining six unique identities were used to generate the crowd on each trial.

The response face (which participants used to make judgments about the crowd) was thus always different from the identities of the crowd members. This design feature ensured that participants would not be able to use distinctive static features of the response face (e.g., the position of a freckle) to match it to a particular crowd face, and would instead make their selection based on facial emotion alone.

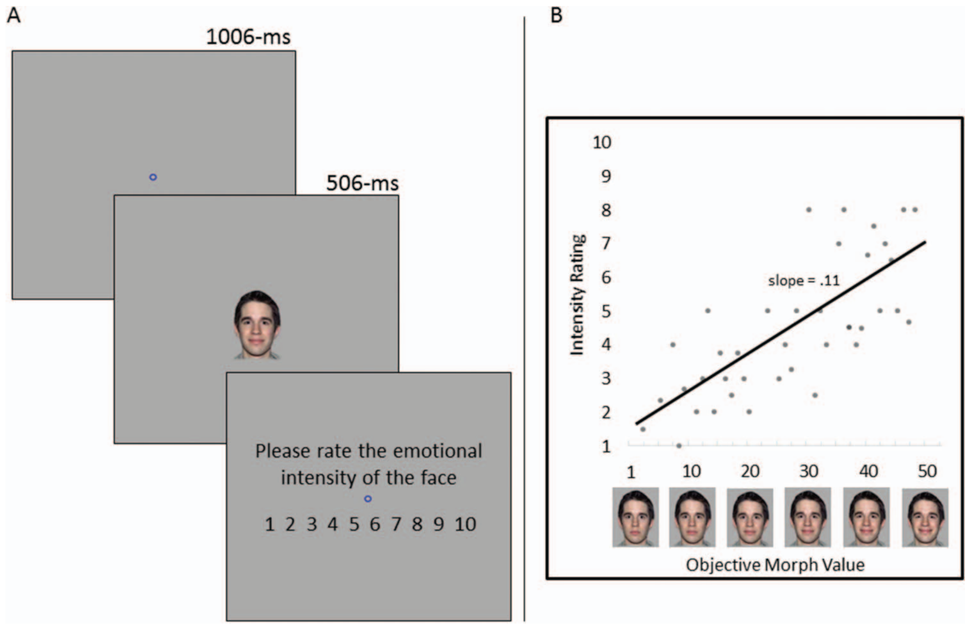


Figure 2. Trial structure for the pretest. Participants each completed 400 trials (Panel A). On each trial, a random face from the set of 1,184 unique faces was shown, without replacement. For each face, participants rated the intensity of the emotional expression (1–10). We pooled data cross all participants and plotted changes in perceived intensity (1–10) as a function morph intensity (1–50), separately for each set of emotion morphs (Panel B). We fit linear and logistic functions to each set of emotion morphs separately for each of the eight identities in our stimulus set. Linear fits were more appropriate to the data than logistic fits, $\Delta SSE = -8.96$, $\Delta R^2 = .06$. See the online article for the color version of this figure.

Crowd perception task. We used a crowd perception task to (a) measure the sensitivity of participant's summary percepts and (b) to manipulate emotional segregation. We here describe the task and how we indexed ensemble coding, and then describe how we manipulated emotional segregation.

On most of the 216 crowd perception task trials, participants viewed a crowd of 12 faces for 306 ms. On these crowd trials, all faces subtended a visual angle of 3.43° × 3.85°, on average. Crowd trials featured 12 faces arranged around a fixation point. The 12 faces were composed of six unique identities repeated twice in each crowd. Random variation was introduced in the 12 possible locations such that the centroid of each position varied by 1 to 15 pixels in either direction along the horizontal and vertical axes. We randomly jittered the location of each face to make it difficult to predict the location of a given face on each trial, and therefore encourage a relatively global spread of attention. Prior to the random position variation, the centroids of adjacent faces were 10.8° away from each other along the horizontal axis, and 8.1° away from each other along the vertical axis. The other 20% of trials presented only a single face, and those trials enabled an important control for evaluating whether participants used multiple faces to make crowd judgments (i.e., was ensemble coding active?). Single-face trials featured just one face in one of the 12 possible crowd-member locations.

The brief duration of each crowd trial (305 ms) made it unlikely that participants would have serially inspected individual faces before making judgments about the crowds. Rather, the brief duration encouraged a strategy of evaluating the entire crowd in parallel. All faces in a given crowd displayed different intensities of one facial emotion (e.g., happy). Each crowd included two emotion subgroups, such that the average emotion of one subgroup was more intense than the average emotion of the other subgroup. Critically, one of the emotion subgroups was populated entirely by six Black faces and the other was populated entirely by six White faces. Figure 3 includes an example of a trial but note that, because of legal permissions, the crowd image is shown here with only two unique identities and the response face is included in the crowd, whereas on actual trials, each crowd image included six unique identities and the response face was never included in the crowd. Whether the Black or White subgroup exhibited more intense emotion was counterbalanced across trials.

Once the crowd disappeared, each face from the crowd was backward-masked with a scrambled, inverted facial image for 1 s. Then each participant adjusted a response face's emotion (presented at the center of the screen) by moving a cursor to the left or right to adjust emotional intensity, sequentially progressing through morphs on the emotion wheel (see Figure 1) to match the expression of the group. Recall that participants' task on each trial was to identify the average emotion of the entire crowd (there was no mention of race or subgroups in the instructions). The response face had one of two identities (a black or white face, counterbalanced across trials) that never appeared within any crowd. The emotional starting point of the response face was randomly drawn from a uniform distribution on the emotional response wheel on each trial (see Figure 3). We also included single-face trials that

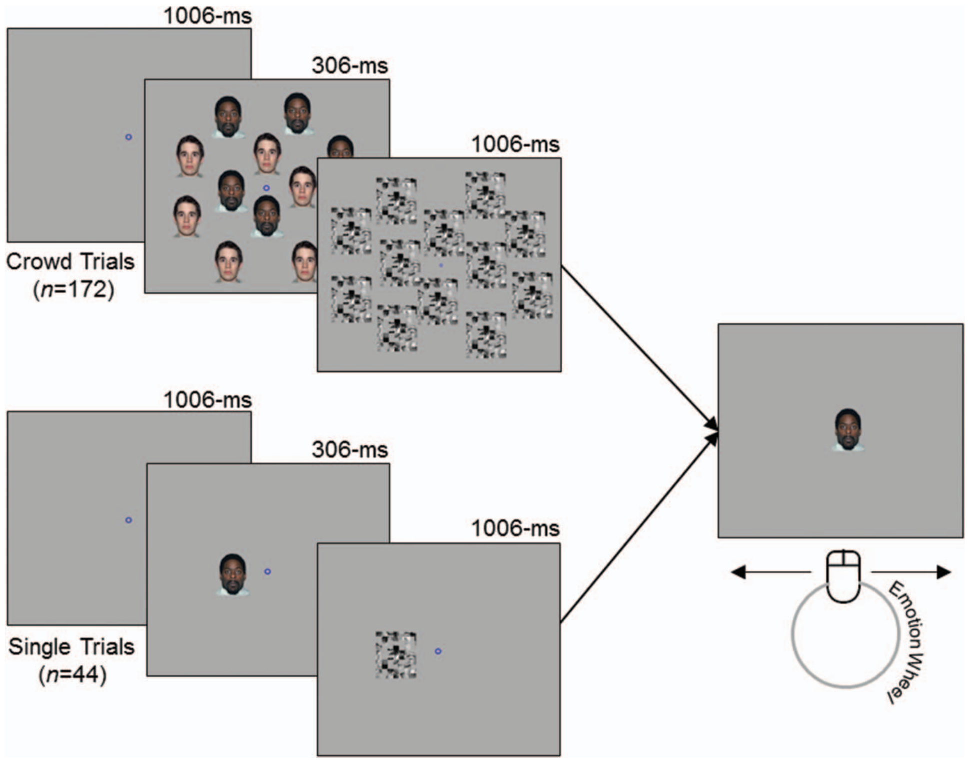


Figure 3. Trial structure for the crowd perception task. Real emotional faces from six different identities were used in every trial. However, to comply with face-model confidentiality, only two NimStim identities appear in the crowd image. Moreover, the response face identify was never included in the crowd and was never depicted as the face on single trials. On each of the 172 trials (80% of the experiment), a crowd of 12 faces was visible for 306 ms and was immediately backward-masked with inverted, scrambled facial images. Participants' task was to indicate the average emotion of the crowd they had just seen. They did so by adjusting the emotion on a novel face (a "response face") that had not been part of any crowds. The response face appeared immediately after the masks, and participants moved their computer mouse left or right to scroll through the emotion wheel. An additional 44 trials (20%) did not present crowds, but rather a single face. On these trials, a single face was visible for 306 ms before it was backward-masked. Following the mask, participants identified the emotion and intensity of the face they had just seen by using a computer mouse to scroll through the emotion wheel of a novel face. See the online article for the color version of this figure.

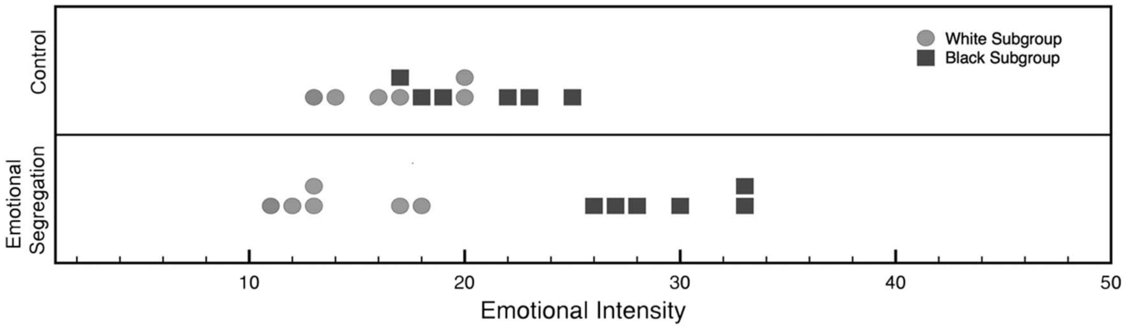
were otherwise identical to the crowd trials. Their importance is described in more detail later (see the Results section).

Measurement of summary percepts. For each trial, we calculated an absolute error score—the difference between the actual crowd average and the participant's perceived crowd average. As in prior research (Elias et al., 2017), we excluded trials in which participants selected an emotional response that was a categorically different emotion from that of the crowd. Difference scores from categorically separate portions of the emotion wheel were not meaningful.[4] For each participant, we then aggregated a mean value of the absolute difference scores. The crowd error score inversely quantifies a participant's accuracy in extracting the mean emotion of the entire crowd.

Manipulation of emotional segregation. On a between-subjects basis, we varied the emotional segregation between black and white subgroups. In the *control condition*, participants saw crowds in which there was only a small difference between the average emotion expressions of the black and white subgroups, whereas in the *emotional segregation* condition, participants saw crowds in which there was a large difference between the average emotion expressions of the black and white subgroups. Specifically, in the control condition, the mean emotion intensities of the two subgroups (races) were set to be four morph units apart (on average across trials). In the emotional segregation condition, the mean emotion intensities of the two subgroups were set to be 15 morph units apart (see Figure 4). The mean intensities of the subgroups never exceeded the middle 58% of any one emotion range.

In each condition, there were an equal number of trials in which the white subgroup was more emotional as there were trials in which the black subgroup was more emotional. For each trial, the emotion intensities of the six facial images in each subgroup were randomly selected from a normal sampling distribution, with the

[4] Inclusion of these categorically incorrect responses does not alter pattern of results reported herein. Further, inspection of categorical errors revealed that they primarily occurred in the neutral portion of any given emotion, where "happy," "fearful," and "angry" faces look similar.



	White Subgroup								Black Subgroup							
	M	*SD*	Emotion of Individual Faces						*M*	*SD*	Emotion of Individual Faces					
Control	17	2.94	13	14	16	17	20	20	21	3.14	17	18	19	22	23	25
Emotional Segregation	14	2.83	11	12	13	13	17	18	30	3.02	26	27	28	30	33	33

Figure 4. The emotional intensities of black and white faces in two example trials are shown. Each square (black subgroup) or dot (white subgroup) represents the emotion of a single individual on one example trial. In the control condition (upper row), the means of the black and white subgroups are similar, whereas in the emotional segregation condition (lower row), the means of the black and white subgroups are more distinct. For both conditions, the means of the black and white subgroups were assigned ahead of time. On a given trial, each actor was randomly assigned an intensity within the given emotion category to construct a crowd that matched the means we assigned. We set the standard deviation *within* each subgroup to be equivalent across condition.

distribution of each subgroup characterized by a standard deviation of three morph units (for an example of crowd sampling, see Sweeny et al., 2013). The sampling distribution for each of the two subgroups was centered at each subgroup's emotional mean and the difference between these means (four or 15 morph units) varied according to whether a participant was in the control condition or the emotional segregation condition. Importantly, then, within-race variability in emotion intensity was equivalent between the two conditions—the between-condition difference in crowd variability was thus attributable entirely to the degree of emotional segregation.

Variability within each subgroup of the crowd ensured that any single face would not be informative about an entire subgroup or the entire crowd. In other words, to make a precise judgment about the entire crowd, participants had to rely on ensemble coding—pooling information from multiple faces—a process we confirmed by comparing errors on crowd and single-face trials (see the Results section).

Finally, we included multiple emotions to ensure that results were not specific to one emotion, as our theory is not specific to any one emotion. Faces were happy on one third of trials, angry on one third of trials, and fearful on one third of trials.

Racial categorization measure. After completing the crowd perception task, participants completed a racial categorization task. On each of 21 trials, participants were asked to identify a face as "Black," "Biracial," or "White." This task assesses the extent to which participants draw a rigid categorical distinction between races and thus assign a monoracial identity to each person, as opposed to identifying people as having mixed-race heritage. We therefore compared participant usage of monoracial versus biracial labels, as in prior research (Chen & Hamilton, 2012; Pauker, Ambady, & Apfelbaum, 2010; Slepian, Weisbuch, Pauker, Bastian, & Ambady, 2014). The faces used in this task were created via FaceGen (Blanz & Vetter, 1999) and were extensively pretested for use in a different set of studies (Pauker, Ambady, & Freeman, 2013). This face set included seven black, seven biracial, and seven white faces that were previously pretested to ensure that there was consensus among raters regarding face race (i.e., black, biracial, white). These faces did not, however, differ significantly on perceived emotional valence or attractiveness.

In the racial categorization task, faces were presented in a random order using MATLAB and the Psychophysics ToolBox (Brainard, 1997) and participants were asked to categorize each face as quickly and accurately as they could using a keypress. Participants selected "Biracial" by using the space bar but on a between-subjects basis, we counterbalanced the assignment of the "f" (vs. "j") key for "Black" or for "White" (given that leftward horizontal spatial location signals agency for English speakers; Maass, Suitner, Favaretto, & Cignacchi, 2009). We report the percent of faces participants identified as biracial. There was no effect of counterbalanced keys on the number of biracial categorizations made, $p = .712$.

Race essentialism. To measure race essentialism, we used the Race Conceptions Scale, a reliable and construct-valid 22-item questionnaire (M. J. Williams & Eberhardt, 2008). The scale includes items such as "racial groups are primarily determined by biology" and "it's possible to be a member of more than one race" (reverse-scored) rated by participants on a 7-point Likert scale. The scale items had good reliability (α = .81) and scale scores were computed by reverse-scoring the relevant items and taking an average score from the 22 items. Scores were normally distributed by condition (defined as skew falling between −1 and 1).

Procedure. Participants first completed 216 trials in the crowd perception task (72 in each of three blocks) followed immediately by the Race Conceptions Scale and the Racial Categorization Task. Participants also completed several questionnaires (see Appendix A) for a separate study before filling out demographic information and being debriefed.

Results

Experimental manipulations did not statistically interact with participant race (*p*s > .50) or gender (*p*s > .53) in predicting any outcome variable, and we therefore collapse across participant race and gender in analyses reported below. We detail results of the main hypotheses first (i.e., Hypotheses 1–3) followed immediately by tests of ensemble coding hypotheses (i.e., Hypotheses 4–6).

Hypothesis 1: Influence of emotional segregation on racial categorization. We expected exposure to race-based emotional segregation to lead to sharp racial category distinctions. Specifically, we expected participants in the emotional segregation condition to be more reluctant than participants in the control condition to describe a person as mixed-race (vs. *either* White or Black). Indeed, participants in the emotional segregation condition identified *fewer* faces (one face, on average) as biracial (M = 26.8% of faces, SD = 9.9%) than those in the control condition (M = 31.1%, SD = 10.6%), $t(146) = -2.57$, p = .011, d = .43, 95% CI for the difference between conditions [−7.62%, −.95%] (see Figure 5A). These findings are consistent with the postulate that brief exposure to race-based emotional segregation in crowds sharpens racial-category distinctions.

Hypothesis 2: Influence of emotional segregation on race essentialism. We expected exposure to race-based emotional segregation to cause participants to think of racial categories as biologically driven, universal, and mutually exclusive. Specifically, we expected participants to more strongly endorse essentialist beliefs about race in the emotional-segregation condition versus the control condition. Indeed, participants in the emotional segregation condition endorsed essentialist beliefs about race (M = 4.21, SD = .67) more than did those in the control condition (M = 3.94, SD = .79), $t(146) = 2.22$, p = .028, d = .37, 95% CI for the difference between conditions [.03, .51] (see Figure 5B). These findings are consistent with the postulate that exposure to race-based emotional segregation fosters race essentialism among perceivers.

More generally, seeing emotional segregation in an interracial crowd increased the extent to which participants (a) perceived race in terms of distinct categories and (b) thought of racial categories as reflecting underlying essences.

Hypothesis 3: Mediation of emotional segregation. We expected weaker racial category boundaries, indexed via the use of biracial (vs. monoracial) categorizations, to be a precondition for the effects of emotional segregation on participants' racial essentialism. Participants should not logically believe that racial categories are biologically driven and universal (e.g.) if they do not identify race as categorical in the first place. We therefore expected racial categorization to mediate the relationship between perceived emotional segregation and racial essentialism. Using PROCESS (Hayes, 2013), we ran a simple mediation model testing the effect of emotional segregation on race essentialism through biracial categorizations. In other words, we tested whether

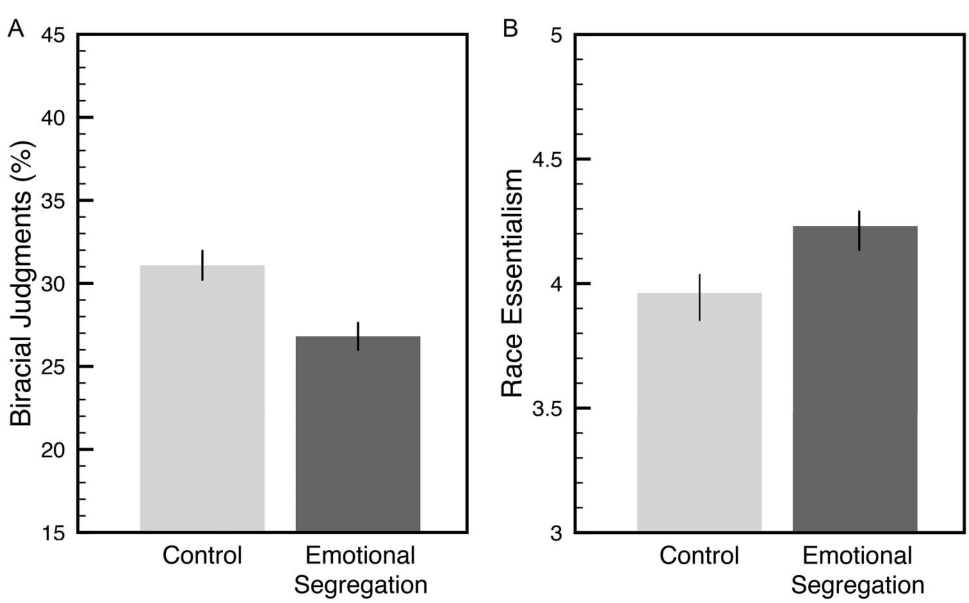


Figure 5. Effects of emotional segregation on biracial categorizations (Panel A) and race essentialism (Panel B). Error bars depict the 95% confidence intervals. Those in the emotional segregation condition made fewer biracial judgments and had stronger race essentialism scores than those in the control condition, *p*s < .05.

the influence of crowd perception on race essentialism depended on the degree to which participants subsequently categorized others as biracial or instead used more rigid and distinctive monoracial categories.

As hypothesized, there was a significant indirect effect of emotional segregation on essentialism through biracial categorizations (see Figure 6). First, and as in the analyses reported above, experimental condition predicted the use of biracial categorizations, $b = -.91$, $t(146) = -2.57$, $p = .011$, $d = .34$, 95% CI [−7.62%, −.95%]. Second, decreased use of biracial categorizations predicted increased race essentialism, $b = -.05$, $t(146) = -1.96$, $p = .052$, $d = .34$, 95% CI [−.11, .001]. We tested the significance of the indirect effect using bootstrapping procedures. Unstandardized indirect effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect testing the mediating role of racial categorization (i.e., condition—biracial categorizations—essentialism) was .05, and the 95% confidence interval did not cross 0 (.003 to .14). Thus, the indirect effect was statistically significant, which suggests that emotional segregation in interracial crowds may have influenced race essentialism via strengthened racial category boundaries.

Hypothesis 4: Rapid summary perception of interracial crowds. We next turned our attention to ensemble coding. Although Hypotheses 4 and 5 are precursors to the ensemble coding hypothesis in which we had the most interest (Hypothesis 6), they are nevertheless important because they allowed us to confirm that participants engaged ensemble coding and that the sensitivity of summary representations followed predictable patterns. That is, one should not assume that ensemble coding has occurred simply because participants completed a crowd perception task—participants could have simply sampled a single face from the crowd, for example, and then used that face to generate their response. To evaluate the operation of ensemble coding, we included *single-face trials* (20% of all trials) to simulate how participants would have performed on crowd trials if they had made evaluations based on the emotion of just one face. On these trials, we still generated crowds of 12 faces with a variety of emotional intensities, but we only displayed a single randomly selected face from this crowd. We recorded the average emotion expression of the crowd even though participants were not permitted to see the entire set. Accordingly, on single-face trials, participants had no choice but to base their responses on the emotion of the single face that they were permitted to view. It is important to note that the purpose of this condition was not to ask participants to make evaluations about faces that they could not see. Instead, this condition simulates what performance on actual crowd trials would have looked like had participants not engaged ensemble coding and instead based responses on a single random face from the crowd. For each single-face trial, then, we compared the participant's response with the average emotion of the crowd (which participants could not see). We expected errors calculated against this crowd-average to be relatively high.

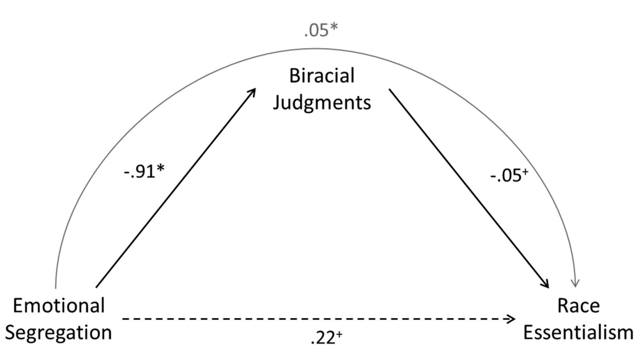


Figure 6. Depiction of the simple mediation model testing the indirect effect of visual grouping on essentialism via biracial categorizations. * $p < .05$. $^{+}$ $p < .10$.

For each participant, we aggregated error separately for single-face trials and crowd trials. We then compared average errors on single-face trials to average errors on crowd trials. If participants used multiple faces to make their evaluations on crowd trials (i.e., ensemble coding was active), their responses should have been closer to the mean of the crowd than their responses on single-face trials. This is exactly what we found. Specifically, participants produced significantly less error on average on crowd trials ($M = 10.52$, $SD = 1.57$; see Figure 7) than on single-face trials ($M = 11.11$, $SD = 2.09$), $t(147) = -4.47$, $p < .001$, $d = .74$, 95% CI [−.85, −.33], confirming they were integrating multiple faces into the crowd average, consistent with ensemble coding.[5] This pattern was true across emotion such that when we examined error in a 2 (Trial Type: Crowd, Single) × 3 (Emotion: Angry, Happy, Fearful) repeated measures analysis of variance, the effect of trial type persisted; participants produced less error on crowd than single-face trials, $F(1, 147) = 20.65$, $p < .001$, $\eta^2 = .141$. Moreover, there was no emotion by trial type interaction, $F(2, 294) = .60$, $p = .552$, $\eta^2 = .004$, indicating the pattern in error did not significantly differ across emotions. The analyses that follow are therefore collapsed across emotion.

Hypothesis 5: Influence of emotional segregation on ensemble coding. The between-race difference in emotion expression that was presented to participants was larger in the emotional segregation condition than in the control condition. As described in the Introduction, and as in prior work (e.g., Marchant et al., 2013), we expected such increases in crowd heterogeneity to lead to increased error in summary percepts. Indeed, even if participants were equally efficient (between the two conditions) in engaging ensemble coding to integrate information from multiple crowd members, the overall heterogeneity difference between the two conditions should have produced less precise summary percepts for participants in the emotional segregation condition than the control condition. Indeed, the summary percepts of participants in the emotional segregation condition included more error (less precision; $M = 10.78$, $SD = 1.59$) than did the summary percepts of participants in the control condition ($M = 10.25$, $SD = 1.52$), $t(146) = -2.05$, $p = .043$, $d = .34$, 95% CI [−1.03, −.02].

Tests of Hypotheses 4 and 5 suggest that (a) participants actively engaged ensemble coding processes and that (b) participants

[5] As expected, judgments of single faces were less representative of their respective crowd averages in the emotional segregation condition compared with the control condition, $t(146) = 3.53$, $p = .001$, $d = .58$, 95% CI [−1.83, −.51].

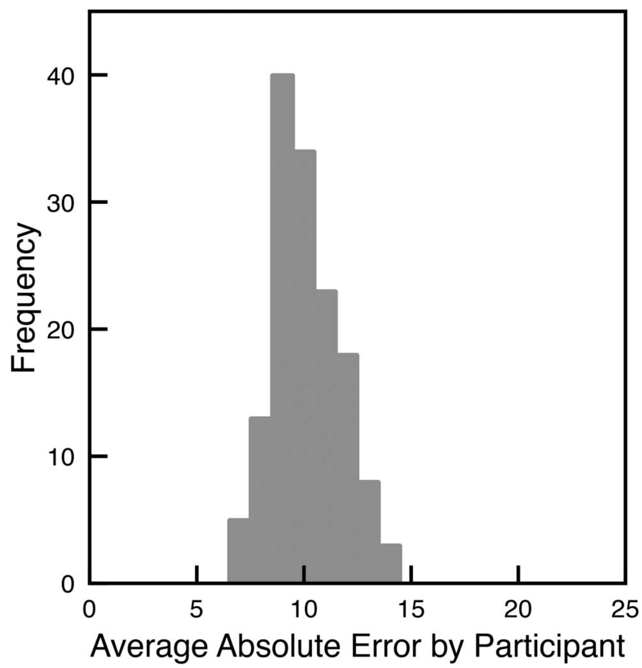


Figure 7. Histogram illustrating participants' average absolute error across all 172 crowd trials. Units on the *x* axis refer to objective morph units.

encoded a greater emotion discrepancy between races in the emotional segregation condition than in the control condition. These tests set the stage for Hypothesis 6, which examines the role of ensemble coding in the influence of crowd perception on racial cognitions.

Hypothesis 6: Mediation of racial categorization by summary perception. We expected visual processing of emotional segregation, indexed via the precision of summary percepts, to be a precondition for the effects of emotional segregation on participants' racial cognition. We thus expected the influence of exposure to emotionally segregated races on racial cognition to be mediated by the precision of summary percepts.

Using PROCESS (Hayes, 2013), we ran a sequential multiple mediation model testing the effect of emotional segregation condition on racial cognition through the precision of summary percepts. Specifically, we tested the effect of emotional segregation (0 = control) on race essentialism through summary percept precision (i.e., error) and biracial categorizations. In other words, we tested whether the influence of crowd segregation on race essentialism depended on how effectively participants used ensemble coding to evaluate the interracial group as a single entity, and the degree to which summary perceptual error strengthened participants' racial category distinctions. As hypothesized, the relationship between emotional segregation and essentialism was mediated by the precision of summary representation and biracial categorizations (see Figure 8).

As in the analyses reported above, emotional segregation predicted summary perception error such that participants in the emotional segregation condition less accurately estimated the emotion of the entire interracial crowd, as compared with those in the control condition, $b = .52$, $t(146) = 2.05$, $p = .043$, $d = .34$, 95% CI [.02, 1.03]. Second, increased error in summary percepts predicted biracial categorizations such that lower fidelity of summary percepts was predictive of fewer biracial categorizations, $b = .32$, $t(145) = 2.87$, $p = .005$, $d = .48$, 95% CI [.10, .54]. Finally, the number of biracial categorizations predicted race essentialism such that fewer biracial categorizations was predictive of increased race essentialism, $b = .06$, $t(144) = 2.08$, $p = .040$, $d = .35$, 95% CI [.003, .12].

We tested the significance of all indirect effects in the model using bootstrapping procedures. Unstandardized indirect effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed. The bootstrapped unstandardized indirect effect testing the mediating role of summary perception through biracial categorizations (i.e., condition–error–monoracial categorizations–essentialism) was .01, and the 95% confidence interval ranged from .001 to .04 ($p < .025$). Thus, the indirect effect was statistically significant, which indicates that the perception of emotional segregation in interracial crowds influenced race essentialism via summary percepts that strengthened racial category boundaries.

In summary, we observed that exposures to emotional segregation in interracial crowds increased the degree to which participants treated racial categories as mutually exclusive and endorsed race essentialism. These effects were mediated, in part, by ensemble coding.

Study 2: Do People Encounter Emotional Segregation in Natural Environments?

Study 1 provides proof of concept for the mechanisms described in our theory, and such proof of concept is the main conclusion to be drawn from this article. Yet apart from the question of whether people possess crowd perception mechanisms capable of informing their intergroup cognitions, there is the question of whether contemporary life provides sufficient learning opportunities for the crowd perception effects observed in Study 1 to meaningfully influence modern humans' intergroup cognitions. It is important that brief exposures to crowds (1/3 second each) influenced racial cognition in a predictable manner, but this occurred after exposure to a number of trials (i.e., 216 trials) and the statistically significant effects on racial categorization were on the order of only 5% of total biracial categorizations (i.e., 1 of 21 faces). This effect,

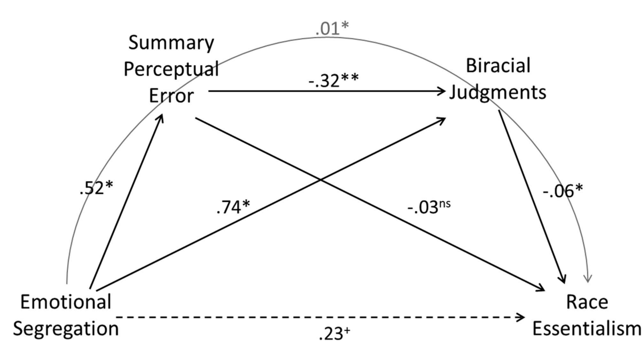


Figure 8. Depiction of the sequential multiple mediation model testing the indirect effect of visual grouping on essentialism via summary perceptual error and biracial judgments. * $p < .05$. ** $p < .001$. + $p < .10$. ns $p > .10$.

however, is still meaningful and not unusual for this area of research (Chen & Hamilton, 2012; Hoffman, Trawalter, Axt, & Oliver, 2016; Kubota, Peiso, Marcum, & Cloutier, 2017; Lloyd, Kunstman, Tuscherer, & Bernstein, 2017). Applying the mechanism we isolated in Study 1 to a lifetime of exposure to crowds could, in theory, have much larger influences on racial category boundaries. This statement of course rests on the assumption that our study population is frequently exposed to emotional segregation in contemporary life. We here provide a preliminary test of this assumption. Specifically, do modern humans actually encounter emotional segregation in their social environments?

Method

Stimulus selection. We sought a stimulus sample relevant to the population of participants in the main study, who were college students. Specifically, we archived the Instagram feeds of 25 American colleges and universities between January and April 2016. Instagram is a popular photo sharing application used by individuals and organizations alike, and colleges use this social media app to visually communicate with potential and current students, parents, and alumni. We purposefully sampled a range of geographic locations, school types (private or public), acceptance rates, sizes, and racial diversity (see Appendix B). We downloaded all images that featured between three and 12 visible faces and defined visible faces as those that showed at least half of the face, yielding 238 crowd images containing a total of 1,243 faces.

Emotion coding. We cropped each individual face out of each crowd image and saved that face as its own image (N = 1,243 individual faces). Each of 573 Mechanical Turk workers then rated individually presented faces (M = 14 faces per MTurk worker; participants were not told that some individual faces came from the same crowd). For each facial image, participants rated emotion on valence from 1 (*extremely negative*) to 9 (*extremely positive*).

Analytic strategy. The number of raters for each face ranged from 27 to 38, and for each face, we took an average across all ratings. To estimate emotional segregation for each crowd image, we compared between-race variability in emotion to within-race variability in emotion. To ensure that these forms of variability were comparable (i.e., on the same metric), we computed absolute deviation scores. Specifically, we first computed (a) the absolute difference between the average emotions of two racial subgroups within a single crowd and (b) the average absolute difference within each racial subgroup in a crowd. The between-race absolute difference was an absolute difference score between the average emotion of majority race individuals and the average emotion of minority race individuals. The within-race average absolute deviation was calculated by averaging the absolute difference between each individual's emotion and their racial ingroup's average emotion. Thus, comparing these two values yields the degree to which emotional segregation was present in the Instagram images. To the extent that emotional segregation was observed in this environment, we hypothesized that between-race deviation would be larger than within-race deviation. To approximate the materials in the main study, we focused on images in which majority race (White) individuals were depicted with minority race individuals (n = 116). We followed up these analyses by examining emotional segregation in a smaller sample of images which included only White and Black crowd members (n = 53).

Results

The average facial emotion in crowds was somewhat positive (M = 6.05, SD = 1.48). More importantly, facial emotion in crowds differed more between races (M = .73, SD = .70) than within races (M = .59, SD = .41), $t(112) = 2.01$, $p = .047$, d = .38, 95% CI [.002, .27]. This is a small-to-medium effect but given the considerable degree of naturally occurring perceptual noise in these images, it is noteworthy that emotional segregation was observed. We observed a similar effect size when specifically comparing black and white emotional subgroups; emotional valence was marginally more similar within (M = .59, SD = .38) than between race (M = .79, SD = .78), $t(52) = 1.79$, $p = .080$, d = .50, 95% CI of difference between values [−.02, .42].

These findings provide preliminary evidence that people (college students) who encounter interracial crowds might often encounter emotional segregation, even if only by way of social and mass media. In this respect, it is notable that the changing demographics of the United States imply that people who do not frequently encounter interracial crowds likely will in the near future. Both public and private college campuses are growing increasingly diverse (National Center for Education Statistics, 2016). For example, the number of Black students attending public institutions has increased by 57% between 2000 and 2014 relative to only a 7% increase among White students. Such increases in racial diversity do not guarantee increases in interracial crowds but evidence does suggest that increased diversity leads to more frequent interracial interactions (Antonio, 1998, 2001; Jayakumar, 2008; Pike & Kuh, 2006). In the current study, for example, interracial crowds made up 51% of the images available to purveyors of university Instagram feeds. Thus, the experimental manipulation in the main study is consistent with evidence that emotional segregation is conveyed via facial emotion in the natural world, and that people may—over time—extensively and frequently encounter emotional segregation.

Discussion

The main purpose of this research was to evaluate if and how perceiving emotionally segregated interracial crowds influences perceivers' racial cognitions. Consistent with our hypotheses, seeing the emotional segregation of interracial crowds shaped participants' racial category boundaries and essentialist beliefs, even though each crowd of 12 was seen for only one third of a second. Ensemble coding processes accounted, in part, for these effects. This evidence is some of the first to demonstrate that visual mechanisms which support perception of crowds can also shape social categorization. Moreover, we presented evidence for the ecological validity of emotional segregation, suggesting that crowd perception may contribute broadly to beliefs that support physical segregation (e.g., race essentialism).

Crowd Perception in Racial Categories and Race Essentialism

We argued that crowd perception plays an important yet underappreciated role in social categorization. Broadly, we theorized that people can learn about social categories by the rapid perception of alliances within small crowds. We examined this broad theory with respect to shared emotion and racial categories. Crowds provide perceivers with information about collective emotion, and we theorized that perceivers categorize people according to (racial) cues that reliably distinguish emotional collectives. We further hypothesized that visual processing of crowds supports this learning process, and we specifically identified ensemble coding as a viable mechanism.

We designed an experiment to test hypotheses derived from our theory, and results supported our predictions. Consistent with our primary hypotheses, perception of emotionally segregated interracial crowds strengthened the boundaries of racial category representations and strengthened beliefs in race essentialism relative to perception of unsegregated interracial crowds. These effects occurred even though participants had very little time to gather information about emotional segregation—each crowd was visible for only 306 ms per trial. Additionally, these effects were mediated by the operation of ensemble coding mechanisms, thus suggesting the existence of a cognitively efficient (visual) mechanism through which people learn social categories. We argue that there are important implications of this work for research on intergroup relations, racial cognition, vision science, and face perception. We describe these implications below along with a review of the limitations of this work and a preliminary study exploring the prevalence of emotional segregation.

Limitations

Single study articles. Widespread and intense debate in psychological science on the "replication crisis" might cause readers to be skeptical of a single experiment article. For example, our sample was college-aged and almost entirely American, the study occurred in a single city and state, emotional segregation was manipulated via eight specific identities, emotions such as disgust and sadness were excluded, and stimuli were limited to disembodied faces of two racial categories. Although we expect these results to be replicable, our goal was not to examine the extent to which the results generalize but rather to develop a mechanistic theory and to design an experiment to test the theory with respect to several key hypotheses (see Mook, 1983). The experimental design enabled us to test several hypotheses which were collectively unique to our theory, and we thus believe the data are in keeping with the theory. Hence, even though it would be inaccurate to draw the conclusion from this one experiment that our theory is "correct" and applies broadly to all humans, the evidence is consistent with our theory, and we thus later describe implications of this work for both method and theory in intergroup relations, social perception, and vision science more generally.

Crowd heterogeneity. In the study design, trials in the emotional segregation condition featured more heterogeneity in emotion than did trials in the control condition. Thus, it is possible that seeing crowds with greater heterogeneity (even if uncoupled from race) may lead to racial category distinctions and to race essentialism. In this section we present two forms of evidence against this hypothesis. First, this hypothesis is inconsistent with findings from prior research. Second, we conducted a control experiment to explicitly test this hypothesis.

Prior findings. Recent research has examined the social impact of exposure to random variability in facial emotion, finding that such variability had no influence (p = .88) on race essentialism (Weisbuch, Grunberg, Slepian, & Ambady, 2016). Hence, in a context where participants saw emotion heterogeneity but not emotion segregation, there was no influence on racial cognitions. Additionally, in the studies reported by Weisbuch and colleagues, increased emotional variability led people to believe that social constructs such as personality and intelligence were less stable. In other words, seeing random emotional variability caused participants to believe that there was actually *less* (not more) categorical structure in the social world.

Control experiment. We more directly ruled out the alternative hypothesis that increased heterogeneity among facial emotions influenced participants' racial category boundaries or essentialism by conducting a follow-up study. Specifically, in this control experiment, emotional heterogeneity again varied by condition, but was uncoupled from race. Below, we summarize this experiment (for more details, please see supplemental materials).

One hundred ninety-eight participants (56 more than in Study 1) completed a series of laboratory tasks identical to that in Study 1, but with one critical difference: In the control experiment, each *subgroup* contained both white and black faces. In the *control condition*, participants saw crowds in which there was only a small amount of heterogeneity between subgroups (both subgroups were interracial), whereas in the *emotional segregation* condition, there was substantial heterogeneity between subgroups (both subgroups were interracial). The control experiment, was otherwise identical to Study 1, including but not limited to the degree of overall emotional heterogeneity, the number of black and white faces in the crowds, and the number of trials.

If increases in overall heterogeneity lead to stronger racial category boundaries, we would expect participants in the emotional segregation condition to exhibit stronger race essentialism and to categorize more individuals as biracial. Instead the number of faces categorized as biracial did not significantly differ between the emotional segregation condition (M = 29.0% of faces, SD = 12.5%) and the control condition (M = 30.0%, SD = 8.9%), $t(196) = -.69$, p = .494, d = .10, 95% CI for the difference between conditions [−4.10%, 1.98%]. Similarly, essentialist beliefs about race did not significantly differ between the emotional segregation condition (M = 4.07, SD = .75) and the control condition (M = 4.08, SD = .64), $t(196) = -.15$, p = .879, d = .02, 95% CI for the difference between conditions [−.21, .18]. See the supplementary materials for additional details and analyses about ensemble coding. The results of this control experiment converge with previous work to suggest that increased heterogeneity did not influence beliefs about race.

The relationship between categorization and essentialism. We predicted and found a mediation of the effect of emotion segregation on race essentialism through racial categorization. Clearly, however, the relationship between racial categorization and race essentialism is correlational and the effects on racial categorization were small. Accordingly, a more conservative interpretation of these results is that perceiving emotional segregation influences both racial categorization and race essentialism through a shared pathway. Notably, this explanation is not contrary to hypotheses but rather reflects the shared conceptual meaning of racial categorization and race essentialism, as measured here.

Racial identities and emotional segregation. Returning to our main study, the crowds that participants saw were limited in racial identities (Black and White only). In the United States, people with Black racial identities have been oppressed for hundreds of years from enslavement up to the culturally aggregated oppression of the current day (e.g., pervasive disparities in income, health care access, and rates of imprisonment). People of both races are largely aware of this history, potentially making "Black versus White identity" an especially salient cue for American perceivers.

It is thus possible that the perception of emotionally segregated crowds simply accentuates existing social category boundaries and their meaning, rather than *creating* those categories in the first place. This limitation would not invalidate the theory presented but would describe important boundary conditions (i.e., strengthening rather than initial creation of boundaries), so future experiments should examine (a) whether the effects of emotional segregation are or are not limited to visible cues which perceivers already use in social categorization, and (b) whether similar results are observed before participants reach middle-childhood—a critical time for the development of racial cognition (Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008; Pauker et al., 2010).

Despite this limitation, we do not believe the results are likely to be restricted to race per se. Like most other social categories, racial categories are learned and their usage does not simply reflect universal visual salience of race (Bigler & Liben, 2007; Dunham, Stepanova, Dotsch, & Todorov, 2015; Williams, Sng, & Neuberg, 2016). Thus, even if the results of the current experiment depend upon existing racial category usage, our theory would suggest that perceptions of emotional segregation would strengthen category boundaries and usage for any well-learned social category. Nonetheless, and especially with respect to examining boundary conditions, future studies should examine the extent to which these findings extend from race to other social categories.

Implications for Intergroup Relations and Racial Cognition

In the current work, we theorized that crowd perception enables people to quickly discern members of different social groups and then distinguish those groups on the basis of other visible features. Our findings were consistent with this theory. It is our hope that this article motivates social psychologists to extend their study of intergroup relations to crowd perception and motivates vision scientists to extend their study of crowd perception to intergroup relations. In this section, we focus on implications for social psychological theory and vision science.

Social category learning. Our theory of social category learning integrated several postulates from social identity theory and evolutionary psychology (Cosmides et al., 2003; Kurzban et al., 2001; Oakes, Turner, & Haslam, 1991; Tajfel, 1969; Turner, 1984). First, social identity theory describes principles by which people might learn social categories but is somewhat agnostic with respect to how those principles are instantiated. For example, social identity scholars have argued that observing homogeneity *within* a collection of persons and heterogeneity *between* collections of persons causes perceivers to (a) regard those collections of persons as different groups (e.g., Turner, 1984), to (b) use social categories to distinguish those groups (e.g., Oakes et al., 1991), and to (c) use social characteristics which distinguish said groups as the basis of social categories (e.g., Tajfel, 1969).

The application of these principles to perceivers has largely regarded sequential perception of individuals (Freeman, Pauker, & Sanchez, 2016; Johnson, Freeman, & Pauker, 2012; Kurzban et al., 2001), which surely plays a role in social category learning, but for which emergent group features may be difficult to track. Indeed, the current work suggests that these social identity principles may be instantiated in rapid and parallel visual processing of human crowds, yielding changes to social category structure and beliefs. Advances in social identity theory may therefore be achieved by further explorations of crowd perception, including the specific crowd features and visual/cognitive processes that give rise to social categories. We focused on emotional segregation in the current work (see below), and because emotion implicates shared mental states it may be an especially powerful grouping cue. Yet according to some social identity treatments, as well as Bigler's influential developmental approach, intergroup heterogeneity more generally drives social category learning, so crowd cues may play a role even when they do not typically imply shared mental states. Proximity, for example, was nominated by Campbell (1958) as a defining feature of groups, and Bigler and colleagues (e.g., Bigler et al., 1997; Bigler & Liben, 2007) have observed that children learn social categories by observing physical segregation. More generally, social identity theories could be enriched by integrating crowd perception processes, as illustrated here.

Similarly, an understanding of visual mechanisms involved in crowd perception might help to advance evolutionary theorizing with respect to social category learning but also more broadly in terms of coalitional psychology. Alliances and bands, for example, can be directly observed in crowds, such that crowd perception might provide a rapid and low cost means of learning about who is in which alliance or group. Crowd perception mechanisms, such as ensemble coding, may have even adapted for the important purpose of identifying and distinguishing between social groups. If so, those visual mechanisms may have been applied more broadly to the nonsocial domains in which ensemble coding operates (e.g., average size of circles; Ariely, 2001) and thus be in keeping with strongly social approaches to the evolution of human cognition (e.g., Caporael, 1997; Dunbar & Shultz, 2007; Humphrey, 1976). More likely, perhaps, human crowd perception may be an exaptation or by-product of general visual mechanisms oriented toward defining and distinguishing stimuli, leaving open the possibility that specific elements of crowd perception are socially specific. Ultimately, as scientific understanding of crowd perception expands, it may prove to be critical to how people perceive, learn about, and think about coalitions.

Contextualizing effects. Participants in the emotional segregation condition categorized, on average, 1 more face as biracial than those in the control condition. Although small, the effects observed here are potentially meaningful in their scope. Existing work, for example, shows that categorizing even one to two people as biracial (vs. monoracial) can have meaningful effects on cognition. For example, Young et al. (2013) demonstrated that reading a short biography in which the target was described as biracial was enough to reduce race essentialism among participants compared with reading a short biography in which the target was described as monoracial. The effects of reading a biography about a biracial individual may differ in strength from categorizing someone as biracial based on their appearance. Nonetheless, even a single categorization can be meaningful. Furthermore, the perceptual manipulation on each trial was subtle and may underrepresent the degree to which emotional segregation is experienced in everyday life. Emotionally segregated subgroups existed within a single emotion category (i.e., happy, angry, or fearful) rather than *between* emotion categories and even the control condition included a small degree of emotional segregation. Moreover, total exposure to this manipulation was about 65 s in total, which likely underestimates typical exposure time to emotional segregation (whether via social media, TV/Film, or in person). As such, our design may underestimate emotional segregation effects on biracial categorization especially if naturally encountered interracial groups express emotional segregation *across* emotional category boundaries. Third, we observed significantly stronger essentialism and more biracial categorizations in the emotional segregation group than the control group after just a 1-hr experiment. Prolonged exposure in ecologically valid settings may accumulate to much more than that.

Vision science. As noted above, theories in social psychology might help to inform understanding of the process of ensemble coding as it applies to summarizing faces and bodies. However, the opposite may also be true; work in vision science might help to inform the study of social grouping and stereotyping. Past work in vision science has provided evidence for the existence of efficient mechanisms to extract socially relevant information (e.g., average emotion expressed by a set of faces; Haberman et al., 2009, 2015; Haberman & Whitney, 2007, 2009, 2010; Whitney et al., 2014). Yet, these visual mechanisms of group perception have only infrequently been applied to the psychological study of stereotyping, prejudice, and social groups, especially as potential mechanisms of social cognition. Here, we have demonstrated a mechanism by which people could use intergroup heterogeneity to inform their schemas about social groups. Greater intergroup heterogeneity was associated with stronger group category boundaries. We suggest that the social psychological study of group processes may benefit from similar methodological cross-talk to examine visual mediators and contributors.

Conclusion

People regularly encounter crowds, and the information they extract from such encounters may be particularly informative for shaping their own behavior and beliefs. Here, we present evidence that brief encounters with emotionally segregated crowds can shape people's racial category boundaries. Furthermore, visual mechanisms of ensemble coding accounted, in part, for effects on social cognition. This evidence is some of the first to demonstrate that ensemble coding can shape social cognition. We have detailed how these findings may inform categorization, emotion, and perception theory in both social cognition and vision science and how the cross-talk between these two disciplines may be critical for theoretical advancement. Integrating precise measurement of visual mechanisms with social–cognitive outcomes is uniquely situated to answer compelling questions about not only the adaptive role of summary perception for social life but also the applications of summary perception to intergroup relations in an increasingly diverse society.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15,* 122–131. http://dx.doi.org/10.1016/j.tics.2011.01.003

Antonio, A. L. (1998, April). *Student interaction across race and outcomes in college*. Paper presented at the annual conference of the American Educational Research Association, San Diego, CA.

Antonio, A. L. (2001). The role of interracial interaction in the development of leaderhsip skills and cultural knowledge and understanding. *Research in Higher Education, 42,* 593–617. http://dx.doi.org/10.1023/A:1011054427581

Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: When older children underperform in social categorization. *Developmental Psychology, 44,* 1513–1518. http://dx.doi.org/10.1037/a0012835

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162. http://dx.doi.org/10.1111/1467-9280.00327

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63,* 596–612. http://dx.doi.org/10.1037/0022-3514.63.4.596

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly, 47,* 644–675. http://dx.doi.org/10.2307/3094912

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 401–432). New York, NY: Cambridge University Press.

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development, 68,* 530–543. http://dx.doi.org/10.2307/1131676

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory. *Current Directions in Psychological Science, 16,* 162–166.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH'99,* 187–194. http://dx.doi.org/10.1145/311535.311556

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3,* 14–25. http://dx.doi.org/10.1002/bs.3830030103

Caporael, L. R. (1997). The evolution of truly social cognition: The core configurations model. *Personality and Social Psychology Review, 1,* 276–298. http://dx.doi.org/10.1207/s15327957pspr0104_1

Castano, E., Yzerbyt, V., Bourguignon, D., & Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology, 38,* 315–322. http://dx.doi.org/10.1006/jesp.2001.1512

Chao, M. M., Hong, Y. Y., & Chiu, C. Y. (2013). Essentializing race: Its implications on racial categorization. *Journal of Personality and Social Psychology, 104,* 619–634. http://dx.doi.org/10.1037/a0031332

Chen, J. M., & Hamilton, D. L. (2012). Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology, 48,* 152–164. http://dx.doi.org/10.1016/j.jesp.2011.10.005

Corbett, J. E. (2017). The Whole Warps the Sum of Its Parts. *Psychological Science, 28,* 12–22. http://dx.doi.org/10.1177/0956797616671524

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7,* 173–179. http://dx.doi.org/10.1016/S1364-6613(03)00057-3

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A: Optics,* *Image Science, and Vision, 18,* 1016–1026. http://dx.doi.org/10.1364/JOSAA.18.001016

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108,* 13341–13346. http://dx.doi.org/10.1073/pnas.1104517108

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science, 317,* 1344–1347. http://dx.doi.org/10.1126/science.1145463

Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2015). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science, 18,* 469–483. http://dx.doi.org/10.1111/desc.12228

Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin, 29,* 360–370. http://dx.doi.org/10.1177/0146167202250215

Eerkens, J. W. (1999). Common pool resources, buffer zones, and jointly owned territories: Hunter-g. . . . *Human Ecology, 27,* 297–318. http://dx.doi.org/10.1023/A:1018777311943

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble perception of dynamic emotional groups. *Psychological Science, 28,* 193–203. http://dx.doi.org/10.1177/0956797616678188

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 470,* 451–470.

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science, 27,* 502–517. http://dx.doi.org/10.1177/0956797615627418

Fritz, M. S., & Mackinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18,* 233–239. http://dx.doi.org/10.1111/j.1467-9280.2007.01882.x

Gil-White, F. J. (2001). Are ethnic groups biological "species" to the human brain? Essentialism in our cognition of some social categories. *Current Anthropology, 42,* 515–553. http://dx.doi.org/10.1086/321802

Goodman, A. H. (2000). Why genes don't count (for racial differences in health). *American Journal of Public Health, 90,* 1699–1702. http://dx.doi.org/10.2105/AJPH.90.11.1699

Gould, S. J., & Vrba, E. S. (1982). Exaptation: A missing term in the science of form. *Paleobiology, 8,* 4–15. http://dx.doi.org/10.1017/S0094837300004310

Graves, J. (2001). *The emperor's new clothes: Biological theories of race at the millennium*. New Brunswick, NJ: Rutgers University Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9,* 1–13. http://dx.doi.org/10.1167/9.11.1

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision, 15,* 16. http://dx.doi.org/10.1167/15.4.16

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17,* R751–R753. http://dx.doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 718–734. http://dx.doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72,* 1825–1838. http://dx.doi.org/10.3758/APP.72.7.1825

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hess, R. F., & Holliday, I. E. (1992). The coding of spatial position by the human visual system: Effects of spatial scale and contrast. *Vision Research, 32,* 1085–1097. http://dx.doi.org/10.1016/0042-6989(92)90009-8

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 4296–4301. http://dx.doi.org/10.1073/pnas.1516047113

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision, 15,* 5. http://dx.doi.org/10.1167/15.4.5

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303–317). Cambridge, United Kingdom: Cambridge University Press.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85,* 616–626. http://dx.doi.org/10.1037/0022-3514.85.4.616

Jayakumar, U. M. (2008). Can higher education meet the needs of an increasingly diverse and global society?: Campus diversity and cross-cultural workforce competencies. *Harvard Educational Review, 78,* 615–651. http://dx.doi.org/10.17763/haer.78.4.b60031p350276699

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102,* 116–131. http://dx.doi.org/10.1037/a0025335

Kelly, R. L. (1995). *The foraging spectrum*. Washington, DC: Smithsonian Institution Press.

Kelly, R. L. (2003). Colonization of new land by hunter-gatherers: Expectations and implications based on ethnographic data. In M. Rockman & J. Steele (Eds.), *Colonization of unfamiliar landscapes: The archaeology of adaptation* (pp. 44–58). New York, NY: Routledge.

Kinzler, K. D., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology, 40,* 581–592. http://dx.doi.org/10.1002/ejsp.739

Kubota, J. T., Peiso, J., Marcum, K., & Cloutier, J. (2017). Intergroup contact throughout the lifespan modulates implicit racial biases across perceivers' racial group. *PLoS ONE, 12,* e0180440. http://dx.doi.org/10.1371/journal.pone.0180440

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 98,* 15387–15392. http://dx.doi.org/10.1073/pnas.251541498

Lewontin, R. (1972). The apportionment of human diversity. *Evolutionary Biology, 6,* 381–398.

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology, 78,* 223–246. http://dx.doi.org/10.1037/0022-3514.78.2.223

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., & Weyers, P. (2008). Modulation of facial mimicry by attitudes. *Journal of Experimental Social Psychology, 44,* 1065–1072. http://dx.doi.org/10.1016/j.jesp.2007.10.007

Lloyd, E. P., Kunstman, J. W., Tuscherer, T., & Bernstein, M. J. (2017). The face of suspicion. *Social Psychological & Personality Science, 8,* 953–960. http://dx.doi.org/10.1177/1948550617699251

Long, J. C., & Kittles, R. A. (2009). Human genetic diversity and the nonexistence of biological races. *Human Biology, 81,* 777–798. http://dx.doi.org/10.3378/027.081.0621

Maass, A., Suitner, C., Favaretto, X., & Cignacchi, M. (2009). Groups in space: Stereotypes and the spatial agency bias. *Journal of Experimental Social Psychology, 45,* 496–504. http://dx.doi.org/10.1016/j.jesp.2009.01.004

Magee, J. C., & Tiedens, L. Z. (2006). Emotional ties that bind: The roles of valence and consistency of group emotion in inferences of cohesiveness and common fate. *Personality and Social Psychology Bulletin, 32,* 1703–1715. http://dx.doi.org/10.1177/0146167206292094

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142,* 245–250. http://dx.doi.org/10.1016/j.actpsy.2012.11.002

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 31,* A93–A102. http://dx.doi.org/10.1364/JOSAA.31.000A93

McDonnell, R., Larkin, M., Dobbyn, S., Collins, S., & O'Sullivan, C. (2008). Clone attack! Perception of crowd variety. *ACM Transactions on Graphics, 27,* 1–8. http://dx.doi.org/10.1145/1360612.1360625

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology, 30,* 93–161. http://dx.doi.org/10.1016/S0065-2601(08)60383-4

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38,* 379–387. http://dx.doi.org/10.1037/0003-066X.38.4.379

Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research, 31,* 2075–2083. http://dx.doi.org/10.1016/0042-6989(91)90165-2

National Center for Education Statistics. (2016). Participation in education. In *The condition of education* (pp. 6–11). Washington, DC: Author.

Nei, M., & Roychoudhury, A. K. (1993). Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution, 10,* 927–943.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology, 30,* 125–144. http://dx.doi.org/10.1111/j.2044-8309.1991.tb00930.x

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology, 59,* 173–191. http://dx.doi.org/10.1037/0022-3514.59.2.173

Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development, 81,* 1799–1813. http://dx.doi.org/10.1111/j.1467-8624.2010.01511.x

Pauker, K., Ambady, N., & Freeman, J. B. (2013). The power of identity to motivate face memory in biracial individuals. *Social Cognition, 31,* 780–791. http://dx.doi.org/10.1521/soco.2013.31.6.780

Peery, D., & Bodenhausen, G. V. (2008). Black + white = black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science, 19,* 973–977. http://dx.doi.org/10.1111/j.1467-9280.2008.02185.x

Phillips, L. T., Weisbuch, M., & Ambady, N. (2014). People perception: Social vision of groups and consequences for organizing and interacting. *Research in Organizational Behavior, 34,* 101–127. http://dx.doi.org/10.1016/j.riob.2014.10.001

Pike, G. R., & Kuh, G. D. (2006). Relationships among structural diversity, informal peer interactions and perceptions of the campus environment. *Review of Higher Education: Journal of the Association for the Study of Higher Education, 29,* 425–450. http://dx.doi.org/10.1353/rhe.2006.0037

Plaks, J. E., Malahy, L. W., Sedlins, M., & Shoda, Y. (2012). Folk beliefs about human genetic variation predict discrete versus continuous racial categorization and evaluative bias. *Social Psychological and Personality Science, 3,* 31–39. http://dx.doi.org/10.1177/1948550611408118

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75,* 811–832. http://dx.doi.org/10.1037/0022-3514.75.3.811

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67,* 741–763. http://dx.doi.org/10.1037/0022-3514.67.4.741

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology, 38,* 119–126.

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology, 59,* 244–274. http://dx.doi.org/10.1016/j.cogpsych.2009.05.001

Sanchez, D. T., Young, D. M., & Pauker, K. (2015). Exposure to racial ambiguity influences lay theories of race. *Social Psychological and Personality Science, 6,* 382–390. http://dx.doi.org/10.1177/1948550614562844

Slepian, M. L., Weisbuch, M., Pauker, K., Bastian, B., & Ambady, N. (2014). Fluid movement and fluid social cognition: Bodily movement influences essentialist thought. *Personality and Social Psychology Bulletin, 40,* 111–120. http://dx.doi.org/10.1177/0146167213506467

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62,* 207–218. http://dx.doi.org/10.1037/0022-3514.62.2.207

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 329–337. http://dx.doi.org/10.1037/a0028712

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. *Emotion, 13,* 76–91. http://dx.doi.org/10.1037/a0029193

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science, 25,* 1903–1913. http://dx.doi.org/10.1177/0956797614544510

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science, 1*(Suppl. 1), 173–191. http://dx.doi.org/10.1017/S0021932000023336

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168,* 242–249. http://dx.doi.org/10.1016/j.psychres.2008.05.006

Turner, B. S. (1984). *The body and society: Explorations in social theory*. Oxford, UK: Basil Blackwell.

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision, 15,* 8. http://dx.doi.org/10.1167/15.4.8

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica, 146,* 7–18. http://dx.doi.org/10.1016/j.actpsy.2013.11.012

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138,* 1172–1217. http://dx.doi.org/10.1037/a0029333

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research, 29,* 47–59. http://dx.doi.org/10.1016/0042-6989(89)90173-9

Weber, A. W., White, D., Bazaliiskii, V. I., Goriunova, O. I., Savel'ev, N. A., & Anne Katzenberg, M. (2011). Hunter-gatherer foraging ranges, migrations, and travel in the middle Holocene Baikal region of Siberia: Insights from carbon and nitrogen stable isotope signatures. *Journal of Anthropological Archaeology, 30,* 523–548. http://dx.doi.org/10.1016/j.jaa.2011.06.006

Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology, 95,* 1063–1079. http://dx.doi.org/10.1037/a0011993

Weisbuch, M., Grunberg, R. L., Slepian, M. L., & Ambady, N. (2016). Perceptions of variability in facial emotion influence beliefs about the stability of psychological characteristics. *Emotion, 16,* 957–964. http://dx.doi.org/10.1037/emo0000123

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). London, England: Routledge & Kegan Paul. http://dx.doi.org/10.1037/11496-005

Whitaker, D., McGraw, P. V., Pacey, I., & Barrett, B. T. (1996). Centroid analysis predicts visual localization of first- and second-order stimuli. *Vision Research, 36,* 2957–2970. http://dx.doi.org/10.1016/0042-6989(96)00031-4

Whitney, D., Haberman, J., & Sweeny, T. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Wener & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 695–710). Cambridge, MA: MIT Press.

Wiese, H., Schweinberger, S. R., & Neumann, M. F. (2008). Perceiving age and gender in unfamiliar faces: Brain potential evidence for implicit and explicit person categorization. *Psychophysiology, 45,* 957–969. http://dx.doi.org/10.1111/j.1469-8986.2008.00707.x

Williams, K. E. G., Sng, O., & Neuberg, S. L. (2016). Ecology-driven stereotypes override race stereotypes. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 310–315. http://dx.doi.org/10.1073/pnas.1519401113

Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology, 94,* 1033–1047. http://dx.doi.org/10.1037/0022-3514.94.6.1033

Young, D. M., Sanchez, D. T., & Wilton, L. S. (2013). At the crossroads of race: Racial ambiguity and biracial identification influence psychological essentialist thinking. *Cultural Diversity and Ethnic Minority Psychology, 19,* 461–467. http://dx.doi.org/10.1037/a0032565

Yzerbyt, V., Corneille, O., & Estrada, C. (2001). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review, 5,* 141–155. http://dx.doi.org/10.1207/S15327957PSPR0502

Appendix A

Additional Measures

Inclusion of Other in the Self Scale (adapted for race; Aron, Aron, & Smollan, 1992)
Racial Outgroup Homogeneity (Park & Judd, 1990)
Internal and External Motivation to Respond Without Prejudice Scale (Plant & Devine, 1998)
Intergroup Contact Questions
Social Dominance Orientation Scale (Pratto, Sidanius, Stallworth, & Malle, 1994)
Marlowe-Crowne Social Desirability in Responding Scale (Reynolds, 1982)
Race-Valence Implicit Associations Task (Greenwald, McGhee, & Schwartz, 1998)

(*Appendices continue*)

Appendix B

Colleges and Universities Used in Content Analysis (Study 2)

University	Region	Size	School type	Acceptance rate	Proportion White	Instagram followers
Appalachian State	South	18,026	Public	63%	86%	19,200
Brown	Northeast	6,548	Private	9%	43%	29,700
Central State	Midwest	2,152	Public	32%	1%	2,322
Emerson	Northeast	3,765	Private	49%	72%	4,634
Hampton	South	3,504	Private	29%	2%	7,997
Kenyon	Midwest	1,662	Private	25%	77%	5,102
Mercer	South	8,600	Private	67%	60%	2,500
Notre Dame	Midwest	12,124	Private	21%	74%	65,200
Oklahoma State	South	25,962	Public	75%	72%	23,800
Santa Clara	West	9,015	Private	49%	50%	7,185
St. Cloud State	Midwest	15,461	Public	89%	81%	2,362
Texas A&M	South	58,577	Public	69%	65%	108,000
Towson	Northeast	22,284	Public	59%	62%	6,108
Tulane	South	13,449	Private	27%	73%	9,530
Tuskegee	South	2,588	Private	48%	<1%	6,398
Central Florida	South	60,767	Public	50%	54%	39,400
Chicago	Midwest	14,467	Private	8%	26%	13,400
Michigan	Northeast	46,625	Public	26%	66%	99,300
Nevada - Reno	West	19,934	Public	84%	62%	6,045
Puget Sound	West	2,553	Private	79%	75%	3,776
South Carolina	South	3,972	Public	65%	78%	35,500
Virginia	South	21,238	Both	29%	63%	38,900
California – Los Angeles	West	43,239	Public	17%	31%	65,100
Vanderbilt	South	12,686	Private	12%	59%	21,200
Villanova	Northeast	7,118	Private	49%	76%	5,963
Wayne State	Midwest	3,453	Public	100%	83%	1,551
Westminster	West	2,233	Private	68%	76%	2,499
Willamette	West	2,287	Private	41%	62%	2,064

Received December 20, 2016
Revision received December 18, 2017
Accepted March 11, 2018 ■