Journal of Experimental Social Psychology 124 (2026) 104889

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Ensemble perception in entitativity judgments of natural crowds☆

Sarah Ariel Lamer [a,*], Spencer Dobbs [b], Lindsay Goolsby [c], Timothy D. Sweeny [c], Max Weisbuch [c]

[a] Department of Psychology, University of Tennessee, Knoxville, United States
[b] Department of Psychology, Southeastern Louisiana University, United States
[c] Department of Psychology, University of Denver, United States

ARTICLE INFO

Keywords:
Entitativity
Crowds
Social perception
Vision
Emotion

ABSTRACT

We conducted two studies to examine how rapid ensemble perceptions of emotion shape judgments of entitativity. We hypothesize that when a perceiver encounters a crowd, they rapidly discern the degree of emotional variability in that group via ensemble coding, and this input provides a strong basis for both immediate and deliberative entitativity inferences. We test this hypothesis using images of a representative sample of natural groups collected via Instagram®. Each person in each of ~200 group images was independently coded for facial affect, gender, and race. Participants evaluated facial emotion variability in each group. We hypothesized that people perceive group variability in facial emotion after brief exposure (150 ms), and that these rapid perceptions predict entitativity judgments of groups viewed for longer periods (3 s). We further hypothesized that rapid ensemble perceptions of emotion variability support the rapid formation of entitativity judgments: entitativity judgments made after 150 ms will approximate those made after 3 s. In Study 1, participants saw images for 5 s each, evaluating (a) shared emotion among group members or (b) entitativity. Study 2 employed a 3 (exposure: 150 ms, 750 ms, 3 s) x 2 (rating: shared emotion, entitativity) x 2 (faces visible: yes, no) between-subjects design. All hypotheses were supported, and the effects were largely independent of the race and gender composition of the groups. These data suggest that complex conceptual judgments about small groups can be formed in the initial instants of perception. We discuss the implications for ensemble coding theories and methods, as well as for entitativity theories.

1. Ensemble coding in entitativity judgments of natural crowds

Although people often experience emotions as highly personal phenomena, the experience and expression of emotion tend to be similar among people with a common group identity (e.g., Clark & Finkel, 2005; Clark & Taraban, 1991; van Schaik et al., 2011; Weisbuch & Ambady, 2008). The relationship between group processes and shared emotion is noteworthy for several reasons (e.g., Parkinson, 2020; Smith et al., 2007; van Kleef, 2009), including the inferences perceivers draw about people who express similar emotions. Perceivers often infer shared group membership among people who express similar emotions (Heerdink et al., 2015; Magee & Tiedens, 2006) and tend to think that such shared emotion connotes cooperation, trust, and satisfaction in the group (Homan et al., 2015). Among factors that may explain these phenomena, existing theories seldom consider the visual processes that allow perceivers to translate *perceived* emotions into conceptual (semantic) inferences about group membership and functioning. We argue that an established visual process (*ensemble perception*) can help explain the relationship.

We focus on *entitativity*: evaluations of the degree to which a collection of people represents a social group entity (e.g., the Detroit Lions football team) or are only connected by happenstance (e.g., waiting in line at a bus stop; Lickel et al., 2001). Although the antecedents of entitativity have been examined in-depth (see Hamilton et al., 2015), these antecedents are typically post-perceptual (e.g., the *belief* that group members experience similar outcomes) and thus separate from the sensory and perceptual processes necessary for *seeing* groups. Thus, the processes responsible for translating *perceptions* of group emotion to *conceptions* of people remain arcane.

Building on Campbell's (1958) argument that perceptual organization processes support entitativity evaluations, we argue that processes involved in ensemble perception can support rapid evaluations of entitativity. Perceptual organization processes are fundamental to *ensemble perception*: a process through which perceivers rapidly identify collective

☆ This paper has been recommended for acceptance by Evava Pietri.

* Corresponding author at: University of Tennessee, Knoxville, Austin Peay Building, Knoxville, TN 37996, United States.
E-mail address: slamer@utk.edu (S.A. Lamer).

https://doi.org/10.1016/j.jesp.2026.104889
Received 17 July 2025; Received in revised form 28 January 2026; Accepted 3 February 2026
Available online 10 February 2026

Table 1
Hypotheses.

	Description	Studies
Hypothesis 1	Perceivers should evaluate a group as entitative to the degree that group members display visible similarity in their socially meaningful characteristics (e.g., emotion, race, gender)	Studies 1 & 2
Hypothesis 2	Ensemble perceptions of facial emotion similarity should be predictive of perceivers' entitativity judgments formed after viewing a group for an extended period of time	Studies 1 & 2
Hypothesis 3	Rapid perceptions of group emotion should predict entitativity judgments better than extended perceptions of group emotion	Study 2
Hypothesis 4	Entitativity judgments formed after 150 ms of exposure time should closely correspond to entitativity judgments formed after extended viewing	Study 2

characteristics of a group without closely analyzing each member. Ensemble emotion perceptions emerge in as little as 50 ms of seeing a group and exhibit surprising accuracy (e.g., Haberman & Whitney, 2009). The efficiency of ensemble perception could make it an especially useful mechanism for identifying entitativity in the "real time" of social life, enabling perceivers to rapidly identify whether nearby people constitute a social group or a collection of strangers. We thus examined whether ensemble perception can translate the sight of a group into conceptions of that group. In two experiments, participants evaluated the degree to which people in Instagram images express similar emotions and are entitative. Facial emotion similarity varied by image, and we manipulated how long perceivers viewed each group, whether perceivers rated the group for entitativity or emotion similarity, and whether faces were visible or occluded. We find support for four interrelated hypotheses, collectively suggesting that ensemble perceptions of emotion can translate seeing a natural group into conceptions about that group.

1.1. Perceptions and conceptions of groups

The link between perceptions and conceptions of groups was first described by Campbell (1958), who drew from Wertheimer's "Gestalt" principles of perceptual organization to explain when several people should be regarded as a unique entity (i.e., a social group). For example, Campbell illustrates that a rock is composed of many molecules yet is perceived as a single entity. He noted that principles such as proximity (rock molecules are close to each other), similarity (rock molecules are structurally similar), and shared fate (when one molecule of the rock moves, so do the others) characterize entities. Campbell then applied these principles to social groups.

Modern research on entitativity has extended Campbell's work to lay judgments, emphasizing psychological factors that cause people to evaluate a group as an entity. Although research on entitativity has followed Campbell's emphasis on perceptual organization, such research often uses verbal labels or descriptions as stimuli, focusing on post-perceptual processes. For example, in a widely-cited entitativity article (Lickel et al., 2001), participants were provided labels for a range of collectives (e.g., members of a professional sports team, women, doctors) and rated entitativity (from "not at all a group" to "very much a group") and its antecedents (e.g., similarity, shared goals) for each collective. This and similar approaches have enabled scientists to identify cognitions that drive entitativity evaluations and to examine how entitativity evaluations inform stereotyping and essentialism (Adelman et al., 2019; Lickel et al., 2001; Rydell et al., 2007; Spencer-Rodgers et al., 2007).

Even as Campbell's original essay was based on perceptual organization principles, the role of perception in entitativity evaluations remains poorly understood. This research lacuna is not trivial: encounters with human groups are mediated by the senses and reflect dynamic sensory input. A thorough understanding of entitativity should include an account of the processes that translate seeing multiple individuals into semantic beliefs about their entitativity. A handful of studies using non-linguistic stimuli have begun to examine how visible factors shape entitativity evaluations. For example, Ip et al. (2006) observed that shared movement and physical features predicted entitativity by altering perceptions of common goals and similarity, respectively. However, these authors did not identify the *visual processes* that might explain perceived entitativity. Moreover, the stimulus crowds were nonhuman, yet many visual processes appear to be specific to perceiving humans, such as processes tailored to uniquely human body movements (e.g., "biological motion"; Troje & Chang, 2023) and facial characteristics (e.g., Duchaine & Yovel, 2015). On this basis, the Ip et al. paradigm and other paradigms using nonhuman stimuli (e.g., Lakens, 2010; Lakens & Stel, 2011) are limited in their ability to identify visual processes that give rise to entitativity evaluations of *human* crowds.

Unlike the Ip et al. and Lakens studies, Magee and Tiedens (2006) used human stimuli (faces) and reported study outcomes consistent with our hypotheses (see below). Specifically, they observed that both shared emotion (consistency in emotion expression among the 4 faces) and positive emotion (happy faces) led to evaluations of entitativity. However, like the studies reviewed above, Magee and Tiedens did not examine visual processes that might explain how facial appearance gives rise to judgments of entitativity. Conversely, we examine a visual process that can help explain how visible group cues give rise to entitativity judgments.

1.2. Ensemble perception and entitativity

We argue that ensemble perceptions, and the perceptual organization processes that support them, underlie judgments of a visible group's entitativity. Indeed, perceivers can identify the summary characteristics of a group before they can closely inspect each group member (for reviews, see Alvarez, 2011; Phillips et al., 2014; Whitney et al., 2014; Whitney & Yamanashi Leib, 2018). Dozens of studies have demonstrated that participants can identify the average emotion from a spatially distributed collection of faces, even when the crowd is displayed too briefly to inspect faces sequentially (e.g., Haberman & Whitney, 2009, 2010; Whitney & Yamanashi Leib, 2018). These ensemble percepts are thought to be driven by more general perceptual organization processes that integrate bits of information into coherent wholes (e.g., a face, a car), or group these objects into sets of related things (Palmer, 1999; Peterson & Kimchi, 2013; Wagemans et al., 2012), and those that segment these coherent wholes from each other (Clifford, 2014; Morgan, 1999; Westheimer & Levi, 1987). Crucially, the precision of ensemble percepts depends on visible similarities and differences among crowd members: similarities facilitate perceptual integration, whereas differences facilitate perceptual segmentation, impairing precise ensemble percepts (Brady & Alvarez, 2011; Corbett, 2017; Elias et al., 2017; Elias & Sweeny, 2020; Im & Chong, 2014).

1.3. Ensemble perceptions of emotion

Although ensemble perception applies to virtually any visible characteristic, we focused on ensemble perceptions of facial emotion. Indeed, people base their evaluations of entitativity on inferences of shared behavior and mental states (emotion, goals) to a greater degree

than inferences of shared identity (e.g., social category membership; Lickel et al., 2000), and people who express similar facial emotions exhibit both shared behavior (the expression) and shared mental states (inferred from the expression). Ensemble perceptions of emotion are thus highly relevant to entitativity. Moreover, emotions expressed in the human face are perceived rapidly and consensually (Adolphs, 2002), thus providing an economical substrate from which perceivers may infer shared behavior and mental states.

Departing from most (but not all) prior studies of ensemble perception, we examine perceptions of emotion variability rather than average emotion. Whereas the average emotion of a group may rapidly inform perceivers about group satisfaction, it does not directly convey information about the similarity and cohesiveness of group members and thus does not index shared emotion. Perceived emotional similarity does. Indeed, the relationship between emotion similarity and group membership is ecologically valid: there is considerable evidence that similar emotion expression is more common among members of the same than different groups (e.g., Lamer et al., 2018; van Schaik et al., 2011; Weisbuch & Ambady, 2008; see Parkinson, 2020) and that people express emotion more often to friends and family than strangers (Clark & Finkel, 2005; Clark & Taraban, 1991). People may passively learn about the ecological correlation between group membership and emotion similarity, and—as implied by the ecological models of the Gibsons (Gibson, 1979; Gibson & Pick, 2000) and McArthur and Baron (1983)—infer entitativity from emotion similarity.[1]

1.4. An ensemble perception framework for entitativity judgments of visible groups

We theorized that the ease or precision through which perceivers extract characteristics about a group will shape impressions of group entitativity. Indeed, ensemble percepts are more accurate and precise when crowds share emotion (e.g., Elias et al., 2017). Accordingly, perceivers should evaluate a group as entitative to the degree that group members display visible similarity in emotion, even after controlling for shared social categories (*Hypothesis 1*). We predicted that rapid ensemble perceptions of shared facial emotion would predict perceivers' deliberated entitativity judgments (*Hypothesis 2*), thus accounting for the effects of shared facial emotion similarity on entitativity.

Hypothesis 3 regards how rapid versus extended perceptions of group emotion contribute to entitativity judgments. When people see a group too briefly to permit close analysis of each member, ensemble perception enables high-fidelity representations of the group's average emotion. But, with time to inspect each group member, representations of group emotion are a function of ensemble perception processes *and* localized, deliberate person-by-person analysis. In principle, entitativity judgments could draw from rapid ensemble percepts and/or from face-by-face analysis. We hypothesized, however, that rapid ensemble perceptions of emotion may be the "signal" for entitativity judgments, whereas piecemeal emotion perceptions, though useful for evaluating individuals, may be "noise" that does not contribute to or interferes with entitativity judgments. If so, rapid ensemble perceptions of emotion similarity (150 ms)—as compared to extended perceptions of emotion similarity (3 s)—should be more predictive of entitativity (*Hypothesis 3)*.

Hypothesis 4 suggests that entitativity judgments formed after extended viewing of a crowd are equivalent to judgments formed after viewing a crowd for only a few hundred milliseconds—as fast as ensemble perception occurs. Indeed, rapid ensemble percepts appear to be stored in working memory and shape *memories* for a stimulus crowd (Utochkin & Brady, 2020). Support for our hypothesis would imply that entitativity judgments can form rapidly and do not change with increased exposure to a crowd (for analogous effects with individual faces, see Bar et al., 2006). That is, we have argued broadly that entitativity judgments can form from group information available to the visual system within a few hundred milliseconds of exposure. To the degree that ensemble perceptions drive entitativity judgments for visible groups, those judgments should be formed nearly as rapidly as the ensemble perception occurs. We therefore predicted that entitativity judgments formed after 150 ms of exposure time would be closely correlated with, or even equivalent to entitativity judgments formed after several seconds of exposure time (*Hypothesis 4*).

Collectively, support for these four hypotheses would provide converging evidence that ensemble emotion perception—and the perceptual organization processes underlying them—are one mechanism that can translate the characteristics of the human collectives people see into cognitions about their entitativity. Table 1 lists the 4 hypotheses, which we test in two studies.

1.5. Ensemble perception in judgments of natural group images

Our theoretical framework (Hypotheses 1–4) could be tested with artificial groups of disembodied faces, or even groups of non-human entities (Ip et al., 2006). However, such tests fail to model the complexity of social groups as they are encountered in the world and therefore fail to account for visual and cognitive mechanisms that may play an important role in entitativity judgments of real, visible groups. In the current studies, we used scenes of "natural groups" captured in the types of images frequently perceived on social media (*natural crowds)*. This methodological feature marks the second major contribution of the current article: examining whether ensemble perceptions emerge in natural scenes that are saturated with (task-irrelevant) visual "noise."

Natural crowds include several features that may impede ensemble perception of faces. That is, ensemble perception effects are thought to accrue because perceivers rapidly integrate information from different faces into a group summary (e.g., Ji et al., 2020). Yet such integration may be complicated by the bodies attached to faces and by the "scene" information that appears around faces (e.g., Aviezer et al., 2008). These additional visual cues are absent in most prior work on ensemble perception yet may introduce visual crowding—a disruption of awareness that occurs when nearby objects render a target object less visible (e.g., Whitney & Levi, 2011). Some studies, using disembodied faces, suggest that ensemble perception persists despite crowding (Fischer & Whitney, 2011). However, natural crowd images contain rich, diverse visual information that may interfere with ensemble perception of emotion. As such, the current work uses natural crowd images to enhance scientific understanding of if and how ensemble perception operates on the visible groups people typically encounter. We specifically examine ensemble emotion perception of faces in natural crowds, and importantly, we test whether those ensemble perceptions are based solely on facial information or if body and scene information also contribute. In our view, this is a major contribution to research on ensemble perception of faces specifically and group perception more broadly.

Finally, existing studies in the domains of ensemble perception and entitativity rarely account for covariance among social characteristics. More often, scientists eliminate within-group variance on characteristics that are not of interest to a given study. In ensemble perception or

[1] To be clear, whereas the processes that give rise to emotion similarity in groups are likely to be "emergent", ensemble perceptions of emotion similarity may reflect either selective summarization of a few group members' emotions (Allik et al., 2013) *or* a more emergent process that coarsely represents facial emotion from every member of a group (Baek & Chong, 2020). We do not intend to resolve this debate here, and instead utilize a measure that accommodates both perspectives—by examining the positive-negative dimension of facial emotion rather than discrete facial emotions, we allow for coarse perceptions to emerge in ensemble perception. Indeed, there is evidence that it is the positive-negative dimension of facial emotion that is "perceptually" encoded, independent of linguistic (discrete) categories of emotion (e.g., Lindquist et al., 2014). Finally, measurement of these perceptions was conducted using a scale rather than a binary measure, such that departures from the scale midpoint may be equivalent to perceptions of greater emotion intensity.

entitativity studies on group facial emotion, group members are often the same race, gender, or even personal identity. That approach simplifies visual and cognitive processing and fails to model *covariance* among the characteristics of group members. For example, if people are accustomed to seeing less variance in facial emotion among same-identity group members—as suggested by evidence that people preferentially mimic the facial emotions of ingroup members (e.g., van der Schalk et al., 2011)—their visual systems may be attuned to such covariance, yielding biased estimates of variability in group emotion (i.e., lower estimated variance in the emotions for a same-race vs. a racially-diverse group). Thus, we present data on covariance of facial emotion and phenotypic race and gender in natural group images (Study 1), and we examine ensemble perceptions of shared emotion *controlling for* race and gender characteristics.[2]

1.6. Overview of the current studies

In Study 1, we characterize our sample of real, group images (from Instagram) for variance and covariance in facial emotion, phenotypic race, and phenotypic gender. Participants viewed each crowd for several seconds and estimated either the group's variability in facial emotion or its entitativity, allowing us to test Hypothesis 1 and providing a preliminary test of Hypothesis 2. In Study 2, we manipulated how long participants had to analyze each image (150 ms, 750 ms, or 3 s), whether they evaluated variance in facial emotion or the entitativity of the group, and whether the faces were visible or occluded. Study 2 thus enabled us to test Hypotheses 2–4, to replicate any effects observed in Study 1, and to examine the fidelity of rapid ensemble perceptions of natural group images. Study 2 also enabled us to evaluate the unique contribution of *facial* information to ensemble emotion perceptions and entitativity judgments.

2. Study 1

In Study 1, we examined perceptions of shared emotion and entitativity for natural groups. We tested our prediction that people would judge a group as more entitative to the extent that group members share facial emotion (Hypothesis 1) and whether these effects were independent of race and gender homogeneity. Further, we expected to observe individual differences in how accurately people perceive emotion variability, and predicted that this accuracy would mediate the relationship between actual group emotion and entitativity judgments (Hypothesis 2). Finally, we explored ecological correlations among crowd homogeneity in race, gender, and emotion. Neither this study nor Study 2 was preregistered. In both studies, we report all manipulations, measures, and exclusions and how sample size was determined prior to data analysis.

2.1. Method

2.1.1. Participants

For *Sample 1*, 152 Mechanical Turk workers were recruited in July 2018 to evaluate the entitativity of each crowd. Of these participants, 10 elected not to complete the study, yielding an entitativity sample of 142 participants (52 women, 90 men). Of these, 12 identified as African-American, 107 as European-American, 4 as Asian-American, 7 as Latinx-American, 7 as Native American, and 5 as multiracial. The mean age was 33.53 (SD = 9.94, min = 20, max = 61), with one person excluded from this metric because of a typo in their response.

For *Sample 2,* 197 undergraduate students (160 women, 36 men, 1 no gender report) at a university in the western United States of America were recruited in Fall 2018 to evaluate shared emotion (i.e., group variability in facial emotion) in each crowd. Of these, 5 identified as African-American, 145 as European/European-American, 18 as Asian/Asian-American, 11 as Latinx-American, 2 as Native American, 5 as Middle-Eastern American, and 10 as multiracial (1 participant did not report race/ethnicity). The mean age was 19.15 (SD = 1.36, min = 18, max = 27; 1 participant did not report age).

Data analyses used image (N = 203, see below) rather than participant as the analytic unit, with ratings per image aggregated across raters. We replicated these analyses using cross-classified mixed-effects models: MLM analyses are reported in the Supplemental Materials, as the results of hypothesis testing did not differ. The sample sizes noted above were large to ensure interrater reliability in the aggregated (average) rating across participants. In each study, we aimed for a minimum of 30 coders per condition. When lab funds or participant pool allowed, we continued collecting data until funds were exhausted or the quarter ended. Based on a priori sample size calculations (Hulley et al., 2013), the image sample size provides 80% statistical power to detect r = 0.2. Sensitivity analysis conducted using G*Power (Faul et al., 2007) revealed that with ($|1 - \beta| = 0.8$) and N = 203, the minimum detectable effect size was r = 0.14.

2.1.2. Crowd images

Crowd images collected by Lamer et al. (2018) were used here. Full details on image selection are provided by Lamer et al., but in brief, these images were initially selected so as to be relevant to that study population (college students). Specifically, the Instagram feeds of 25 American colleges and universities were archived 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, donors, and alumni. The sample of universities included a range of geographic locations, school types (private or public), acceptance rates, sizes, and racial diversity (for a list of all schools, see Lamer et al., 2018; Appendix B). Each image was downloaded, provided that it featured at least 3 visible faces (visible faces were defined as those that showed at least half of the face). This process yielded 238 crowd images containing a total of 1243 faces.

Image size varied considerably, as we intended to present images to participants in the sizes that would be naturally encountered on webpages. Thus, image sizes ranged from 260 × 169 to 1080 × 609 pixels, though the majority (n = 155) were 640 × 640. In studies using online samples (Study 1), there were likely to be participant-level differences in the size of the images as presented on participants' computer monitors. For all coding and in all studies, each crowd or individual image was presented against a plain-white screen background.

Actual emotion similarity coding. Each individual head was cropped out of each crowd image and saved as its own image (N = 1243 individual faces). Each of 573 Mechanical Turk workers then rated individually presented faces (M = all individual faces from 14 randomly-

Table 2
Study 1 descriptive statistics.

Variable	*M*	*SD*
Actual Emotion Similarity[1]	2.21	2.60
Race Homogeneity[2]	0.36	0.17
Gender Homogeneity[3]	0.26	0.18
Entitativity[4]	4.21	0.55
Perceived Emotion Similarity[5]	65.47	18.61

Note. [1]Inverse of standard deviation (0.81) among individual faces in emotion for each crowd (individual faces were rated on a scale from 1 to 9). [2]Proportion of White persons vs. persons of color in each crowd, absolute difference from 0.5. [3]Proportion of women vs. men in each crowd, absolute difference from 0.5. [4] Rated on a 1–7 scale. [5]Rated on a scale from 0 to 100.

[2] There was a range of race and gender heterogeneity in the images. However, given that the images sampled for this work were drawn from university Instagram feeds, we anticipated limited age representation. Nonetheless, we coded and reported analyses controlling for age in the Supplemental Materials (see also General Discussion).

selected crowds per MTurk worker; participants were not told that some individuals came from the same crowd). Participants viewed each face (until response) and rated the emotion in the face from 1, *Extremely Negative,* to 9, *Extremely Positive*. The number of raters for each face ranged from 27 to 38, and for each face, we took an average across all ratings. Then, for each original crowd image, we computed the standard deviation in emotion across all faces in the crowd. *Actual emotion similarity* was quantified as the inverse of the standard deviation in actual facial emotion among crowd members so that higher scores indicate greater emotion homogeneity (see Table 2 for means).

Race and gender coding. Each face was independently coded by two trained research assistants. First, two coders identified the perceived racial category of each face (White, Black, Latinx, Middle-Eastern, Native American, East Asian, multiracial). Most faces were coded as White. Moreover, the vast majority of non-White (POC) faces were coded as Black, as Lamer et al. (2018) intentionally oversampled Historically Black Colleges and Universities. Thus, the *n* was extremely small for each *type* of mixed-race visible group (e.g., Black and East Asian members, White and Latinx members) and single-race POC groups (apart from all-Black groups). It is obvious that Black and East Asian monoracial individuals (for example) do not share racial identity but for meaningful statistical analyses on race, each face's race was coded as either White or POC, with inter-rater reliability on these categories reaching 97% agreement (disagreements resolved through discussion). Each visible group was thus coded all-White, all-POC, or mixed White and POC, with mixed groups quantified as the percentage of White members. Notably, because the vast majority of faces were either phenotypically-White ($n = 727$) or phenotypically-Black ($n = 346$), there were only 5 mixed-race groups in the image sample that were not majority White or majority Black. *Race homogeneity* was calculated as the absolute difference between the proportion of White faces in each crowd and 0.5, so that higher values indicated less racial diversity. Scores ranged from 0 to 0.5. A homogenous White crowd would have a score of 0.5 (i.e., |1–0.5|). In contrast, a crowd comprised of 3 Black and 3 White individuals would have a score of 0 (i.e., |0.5–0.5|).

Gender was coded as "woman" or "man". As with race, gender was coded from images and thus reflects *phenotypic* gender rather than gender *identity*. For this reason, and because coders had difficulty identifying "non-binary/gender fluid/genderqueer" phenotypes, we elected to use binary-gender coding. *Gender homogeneity* was calculated by first computing the proportion of faces in each crowd that appeared to be men (as coded by a trained research assistant), and then taking the absolute difference between this proportion and 0.5. Thus, higher values represent less gender diversity, with values ranging from 0 to 0.5.

Image Exclusions. An examination of the 238 images revealed that 35 crowd images were actually images of one to two players in an NCAA sport (usually basketball), with hundreds of mostly-blurry faces (the audience) in the background, only a few of which were rated in the emotion coding pretest. Thus, for these 35 images, the characteristics of all but 1–2 faces were not visible and we therefore excluded these 35 images from all studies reported here, yielding a final sample of 203 images.

2.2. Procedure

Sample 1: Entitativity. After providing informed consent, participants were told that they would be evaluating images for the degree to which the people in the image seemed like a social group. Specifically, they were asked to answer two questions for each image: "Do you think these people are a social group?", with endpoints labeled "much more like a random collection of individuals than a social group" (1) and "much more like a social group than a random collection of individuals" (7); and "Do these people constitute a cohesive social group?", with endpoints "not at all" (1) and "very much" (7). Responses to the two entitativity questions were highly correlated ($r = 0.88$) and thus averaged into a single entitativity score. Each image was presented for 5 s, after which participants made their ratings. After evaluating all crowd images, participants were debriefed, compensated, and dismissed.

Table 3
Ecological correlations.

Variable	Actual Emotion Similarity	Race Homogeneity	Gender Homogeneity
Actual Emotion Similarity			
Race Homogeneity	0.09		
Gender Homogeneity	0.15*	−0.07	

Note. $df = 201$; *$p < .05$, *** $p < .001$.

Sample 2: Perceived emotion similarity. Participants were instructed to evaluate a large number of images for the facial emotion similarity displayed by the group. Few studies have examined judgments of variability in visible groups (see Jeong & Chong, 2021). As opposed to the average emotion in a crowd, the standard deviation or variance of emotion in a crowd are statistical terms that lay perceivers do not always understand. Because we wanted to arrive at the best question format for current and future studies, we explored several different question formats. To the degree that participant responses are highly correlated across question format, convergent validity may be said to exist for perceived emotion similarity. Thus, participants were randomly assigned to one of four rating conditions.

In the *emotion similarity* condition, participants were simply asked "For the crowd you just saw, how much did the faces vary in emotion? The faces display…", with endpoints labeled "Extremely different emotions" and "Extremely similar emotions". The response scale was a slider with possible responses between 0 and 100 (numbers not displayed). As for all four question types, response scale onset immediately followed the offset of the crowd image. Of the four question types, this "similarity" question was intended to be the most intuitive to participants. We thus focused our analyses on how responses to this question were associated with each of the other question formats.

The other three conditions included extensive instructions and practice items, and requested precise and potentially unintuitive responses. We provide descriptions of each in the Supplemental Materials. To simplify data presentation, we present only the subjective emotion similarity condition. This decision was justified by the strong correlations between ratings in this condition and the other three question formats used to assess emotion variability estimates. Specifically, *Pearson's r* ranged from −0.81 to −0.90 (as ratings of similarity increased, ratings of variability/range decreased). We, therefore, regard this intuitive question as possessing convergent validity with more precise questions about variability—even to the point of being statistically indistinguishable. However, when any of the other 3 measures of emotion variability are substituted for the similarity item, the results do not meaningfully alter the results presented below (in terms of statistical significance).

For all conditions, the crowd image was presented for 5 s before the questions appeared on screen (and the crowd disappeared). Each participant rated all crowd images. On completion, participants were debriefed, compensated, and dismissed.

2.3. Results

2.3.1. Data aggregation and descriptive statistics

We analyzed the data using images, rather than participants, as the unit of analysis, allowing us to examine relationships between entitativity (Participant sample 1) and perceived emotion similarity (Participant sample 2). See Table 2 for descriptive statistics of key variables. We also conducted cross-classified mixed-effects analyses on each of the 4 critical hypothesis tests to simultaneously account for image-based and participant-based variance. As reported in the Supplemental Materials, patterns of significance for these MLM tests were identical to those reported below.

2.3.2. Ecological correlations

Correlations among ecological variables were generally small (see Table 3).[3] The correlation between actual emotion similarity and gender homogeneity was small but statistically significant, indicating that groups displaying more similar emotion were also more homogeneous with regard to gender.

2.3.3. Entitativity judgments as a function of group characteristics

We hypothesized (Hypothesis 1) that visible groups would be evaluated as entitative to the degree that they displayed similar facial emotions, even after accounting for race and gender. For completeness, we also report the effects of race and gender homogeneity on entitativity. We conducted a regression with emotion, race, and gender entered together as predictors of entitativity. Consistent with Hypothesis 1, actual emotion similarity [$b = 0.04$, $SE = 0.01$, $t\ (199) = 2.52$, $p = .013$, $\beta = 0.17$] predicted entitativity, *independent* of gender and race homogeneity. Notably, race homogeneity also predicted entitativity [$b = 0.46$, $SE = 0.22$, $t\ (199) = 2.12$, $p = .035$, $\beta = 0.14$) as did gender homogeneity, $b = 0.48$, $t\ (199) = 2.30$, $p = .022$, $\beta = 0.16$. Thus, actual emotion similarity predicted entitativity—over and above homogeneity in race or gender. In general, Hypothesis 1 was supported.

2.3.4. Judgments of crowd variability in emotion and entitativity

One assumption in our framework is that the influence of crowd characteristics on entitativity is based on people accurately perceiving those characteristics. We thus examined participants' accuracy in evaluating emotion similarity in natural crowds. As expected, subjective estimates of facial emotion similarity were correlated with actual similarity, $r\ (201) = 0.45$, $p < .001$. Thus, people effectively extracted ensemble perceptions of crowd variability.

We next tested whether perceived emotion similarity predicted entitativity. They did, $r\,(201) = 0.38$, $p < .001$. These data are consistent with the idea that people draw on their judgments of emotion variability to evaluate a group's entitativity (Hypothesis 1), thereby setting the stage for Study 2, which examines whether such effects emerge under limited viewing time.[4]

Although not central to our four hypotheses (see Table 1), we conducted a mediation analysis. Given that participants exhibited some sensitivity in their ensemble perceptions of emotion similarity, and given that such perceptions predicted entitativity ratings, we examined whether perceived emotion similarity mediated the relationship between actual similarity and perceived entitativity. Mediation analyses are correlational and we assume (consistent with contemporary statistical models) that these analyses alone cannot identify causality for effects of a mediator (perceived emotion similarity) on an outcome (entitativity). Instead, we follow logic elaborated elsewhere for interpreting mediation (Fiedler et al., 2011): such analyses cannot tell us whether perceived emotion variability plays a causal role in the effects of actual emotion similarity on entitativity ratings, but they *can* provide evidence consistent with a given causal theory. In the current study, it seemed likely that, if actual similarity in crowd emotion influenced entitativity judgments, such effects would accrue through ensemble perceptions of emotion similarity. We examined the relevant indirect effect to evaluate the empirical plausibility of this account, with the understanding that we cannot rule out unmeasured mediators as the causal mechanism(s).

In this regard, we observed proof-of-concept: the more varied the emotions in a natural crowd, the greater variability perceivers detected, and the less entitative the group seemed. We used the mediate function in the psych package (Revelle, 2025) to examine the indirect effect of emotion heterogeneity on entitativity through perceived emotion similarity. This indirect effect was significant, $b = 0.04$, $SD = 0.01$, 95% CI [0.02, 0.07], consistent with the view that evaluations of emotion similarity mediated the effect of actual emotion similarity on entitativity judgments.[5]

2.4. Discussion

Several important results emerged in Study 1. Consistent with Hypothesis 1 (and Magee & Tiedens, 2006), actual emotion similarity significantly predicted judgments of crowd entitativity. Moreover, participants demonstrated accuracy in their perceived similarity of crowd emotion. Accuracy statistically mediated effects of actual emotion similarity on entitativity, providing evidence consistent with the view that the effects of actual crowd emotion similarity on entitativity are explained by perceived emotion similarity. Most importantly, however, perceived emotion similarity was influenced by actual emotion similarity and correlated with judgments of crowd entitativity (Hypothesis 1).

Perceived entitativity, of course, owes to more than just emotion homogeneity. And indeed, race homogeneity and, to a lesser degree, gender homogeneity, were also predictive of entitativity. Age heterogeneity may also explain entitativity judgments, but was limited in our stimuli (see Supplemental Materials for analyses with target age). This is important to note given that entitativity coders were recruited from Mturk and may have thus found the images less relevant. Accordingly, we limited Study 2 to a college student sample as our primary interest in the current work is the role that actual similarity plays above and beyond other predictors of entitativity. The effects of actual emotion similarity on entitativity judgments were independent of race, gender, and (albeit limited) age homogeneity. Together, the findings in this

[3] Many images had *no* within-crowd heterogeneity in race or gender. We thus recomputed correlations involving either variable by limiting the sample of crowd images to those with *some* variability in race or gender. For race ($n = 90$), correlations were $r\ (88) = -0.03$ ($p = .751$) for emotion homogeneity and $r\ (88) = 0.02$ ($p = .817$) for gender homogeneity. For gender ($n = 139$), correlations were $r\ (137) = 0.02$ ($p = .796$) for emotion homogeneity and $r\ (137) = -0.02$ ($p = .787$) for race homogeneity.

[4] The samples used to rate entitativity and similarity differed in composition. Specifically, entitativity ratings were provided by an online sample with a broad age range (20–61 years), whereas similarity ratings were collected from an in-lab sample of college students (primarily young adults). Sample sizes also varied; 142 coders rated each image's entitativity, while 51 coders rated each image's similarity in the primary condition. This difference could contribute to differences in the precision of these estimates. We addressed these differences in two ways. First, we reran analyses using a subset of the entitativity coder data. To most closely approximate the college sample, we selected the youngest 51 coders from the online sample ($M = 24.57$, $SD = 2.20$, min = 20, max = 28). The same pattern of significant effects was observed with the full coder sample as the coder subset, regardless of whether analyses were conducted using image-based estimates or MLM. For example, perceived similarity predicted entitativity both in the full coder sample, $b = 0.01$, $SE = 0.002$, $t\ (199.10) = 5.94$, $p < .001$, and in the smaller subset of coders, $b = 0.01$, $SE = 0.002$, $t\ (205.95) = 5.77$, $p < .001$. Furthermore, coder age did not moderate the relationship between shared emotion and perceived entitativity, $b < 0.001$, $SE < 0.001$, $t\ (6{,}988) = -0.07$, $p = .943$ and, although the precision of entitativity estimates was higher in the full sample (i.e., smaller *SD*), the means and variability were very similar across samples (Full: $M = 4.21$, $SD = 0.55$; Subset: $M = 4.12$, $SD = 0.64$). We conclude that differences between the two samples in terms of age or size are unlikely to explain the observed effects in Study 1.

[5] Notably, the inclusion of this indirect effect eliminated the effect of actual emotion similarity on entitativity, $b = 0.01$, $se = 0.02$, $t\ (200) = 0.64$, $p = .521$. This model does not rule out that the mediator and outcome are flipped: that actual emotion similarity influences entitativity directly and has downstream influence on perceived similarity, $b = 0.48$, $SD = 0.19$, 95% CI [0.18, 0.92]. This model is plausible, but we argue it is less logically coherent than a model in which perceived similarity mediates the relationship between actual similarity and entitativity. In line with this argument, the inclusion of the indirect effect through entitativity did not eliminate the effect of actual similarity on perceived similarity, $b = 2.77$, $se = 0.44$, $t\ (200) = 6.35$, $p < .001$.

study motivated Study 2, in which we examined the temporal processing characteristics responsible for these effects.

Several other notable effects emerged. First, there was a small (but significant) correlation between emotion and gender homogeneity, such that crowds with emotion variability also had more gender variability. As noted, however, the effect of actual emotion similarity on entitativity was largely independent of gender homogeneity. No other ecological correlations were significant. Finally, participants' abstract estimates of crowd similarity in facial emotion exhibited convergent validity with more concrete estimates in which participants evaluated the numerical range of facial emotions in each crowd (see Supplemental Materials). In Study 2, we relied solely on the more abstract, intuitive measure to assess rapid ensemble perceptions of facial emotion and entitativity judgments.

3. Study 2

In Study 2, we tested all four hypotheses. Participants completed a similar rating procedure as in Study 1, but with crowd viewing times of 150 ms, 750 ms, or 3 s and with crowd stimuli in which faces were visible or occluded (Fig. 1). This yielded a 3 (viewing time) x 2 (judgment: entitativity or facial emotion) x 2 (faces: visible or occluded) between-subjects design.

Having stimuli with both visible and occluded facial information was important because natural crowds include body and scene information alongside faces. We have hypothesized that summary perceptual information about facial emotion variability is key to deriving entitativity judgments. However, it is possible and perhaps likely that perceivers have difficulty isolating facial emotion from body or scene emotion and may therefore also base their emotion variability ratings on bodies or scenes. Thus, we addressed this limitation by asking another set of participants to evaluate the same visual groups in which all faces were cropped out. By controlling for perceived emotion similarity where faces were not visible, we aim to isolate the contribution of *facial* emotion to perceived entitativity.

First, we expected rapid ensemble perceptions of face emotion similarity—formed after only 150 ms of group viewing time—to be predictive of entitativity ratings even when controlling for non-face information (*Hypothesis 2*). In many laboratory studies, faces are often the only stimuli that vary within a group, so in these designs, any relationship between ensemble perceptions and entitativity judgments *must* be driven by face information. Yet when natural crowds are the stimuli, facial emotions are likely to be conflated with body postures and affective features of the scene. It may be these non-face features driving a relationship between rapid ensemble perception and entitativity (for similar effects with individual faces/bodies/contexts, see Aviezer et al., 2012; Chen & Whitney, 2019, 2020) though we predicted that rapid ensemble perceptions of facial-emotion variability would inform entitativity judgments, even when controlling for estimates generated from group images without facial information.

Second, Hypothesis 3 suggests that longer viewing times introduce "noise" on top of an ensemble perception "signal" because people start analyzing individual faces rather than relying on a quick overall impression. Therefore, perceived facial emotion similarity should predict entitativity more strongly during brief (150 ms) than longer (750 ms or 3 s) exposures. To test this, we regressed entitativity ratings on perceived emotion similarity derived from (a) group images with faces and (b) without faces, at each viewing duration (150 ms, 750 ms, 3 s). We predicted that the influence of facial emotion (images with faces) would weaken with longer viewing times, while the influence of non-facial information (images without faces) would strengthen. This pattern would support Hypothesis 3, suggesting that quick ensemble perceptions of facial emotion drive initial entitativity judgments, whereas extended viewing incorporates broader contextual information from bodies or scenes.

Finally, we expected that entitativity judgments formed after only 150 ms of viewing time would largely approximate those made after 3 s of viewing time (*Hypothesis 4*), suggesting that entitativity judgments rely on visual processes about facial emotion that operate early (vs. later) in vision. Nonetheless, it is likely that entitativity judgments—regardless of viewing time—do not rely solely on facial information but also incorporate information from bodies and visual characteristics of the scene. More generally, and—to our knowledge—for the first time, we examined the degree to which rapid ensemble perceptions of facial emotion are accurate even when groups are presented in natural contexts.

3.1. Method

3.1.1. Participants

Three hundred and eighteen undergraduate students participated in exchange for partial course credit in Spring 2019 (faces visible) and Fall 2021 (faces occluded). Due to a data recording error, 9 participants' demographic data were lost. We include their data in analyses, but only report demographics for the remaining 309 participants. The sample included 222 women, 83 men, and 4 non-binary participants and the following self-reported races: 208 White/European-American, 30 Asian/Pacific-Islander, 11 Black/African-American, 34 Latinx, 20 mixed-race identities, 4 Native American, 1 Middle Eastern, and 1 unreported. The mean age was 19.31 ($SD = 1.73$, min = 17, max = 35), with one additional person excluded from this metric because of a typo in their response. We aimed for 30 participants per condition with a stopping rule at the end of the academic quarter.

3.1.2. Crowd images

We used the same crowd images as in Study 1. Unfortunately, six crowd images were mistakenly excluded from the stimulus set in Study 2 and some participants indicated that they knew people in twelve of the remaining images. Thus, we elected to remove those 12 images to comply with our IRB protocol. Analyses focus on the resulting 185 crowd images. Sensitivity analyses conducted using G*Power (Faul et al., 2007) revealed that with ($|1 - \beta| = 0.8$) and $N = 185$, the minimum detectable effect size for individual predictors was $f^2 = 0.04$. This f^2 converts to a minimum detectable effect size of $\beta = 0.20$.

3.2. Procedure

After completing informed consent, participants were randomly assigned to condition (via Qualtrics©) in a 3 (exposure time) x 2 (rating task) x 2 (face visibility) between-subjects design. Participants assigned to rate entitativity were told that they would be evaluating a large number of crowd images for the degree to which the people in the image seemed like a social group. Given that the two entitativity items in Study 1 were highly correlated ($r = 0.88$), participants in Study 2 were asked only one of the Study 1 items: "To what extent are these people a social group?", with endpoints labeled "much more like a **collection of individuals** than a social group" (1) and "much more like **a group** than a random collection of individuals" (10) (presented to participants with **bolding**). Participants assigned to rate emotion were told that they would rate crowd images for the degree to which the faces in each crowd exhibited similar emotion. Specifically, these participants were asked, "how much did the faces vary in emotion?" with endpoints labeled "Extremely **DIFFERENT** emotions" (1) and "Extremely **SIMILAR** emotions" (10) (presented to participants with bold/underline/caps). Within

Fig. 1. Example Crowd Images with Faces Visible and Occluded. *Note. Images by Henrique Vicente (upper images) and Dan Gaken Images (lower images), used under CC0 license. Although we cannot publish the actual stimuli due to copyright issues, these examples reflect the kinds of images often featured on Instagram news feeds and the editing process used to remove facial emotion.*

each rating condition, participants were randomly assigned to view each image for either 150 ms, 750 ms, or 3 s. On each trial, judgments (either emotion similarity or entitativity) were requested only after the offset of the image. After evaluating all crowd images, participants were debriefed, compensated, and dismissed.

3.3. Results

3.3.1. Analytic strategy and descriptive statistics

As in Study 1, all analyses in Study 2 were conducted at the level of crowd image, with ratings of each crowd averaged across participant judges. As in Study 1, analyses with cross-classified mixed-effects models are reported in the Supplemental Materials: significant results did not differ between image-based and mixed-effects analyses.

Six emotion similarity and six entitativity scores were derived for each image (two scores per exposure time: 150 ms, 750 ms, 3 s – one for when faces were visible and one for when faces were occluded. Our hypothesis tests regarded (1) conditional differences in correlations between judgments and crowd characteristics (e.g., actual emotion similarity) and (2) conditional differences in the correlation *between* perceived emotion similarity and entitativity. Simple comparisons of correlations were computed by applying Fisher's *r*-to-*z* transformation to both correlations and testing the difference between them in a *z*-score distribution. More complex differences among correlations were analyzed via hierarchical linear regression and mediation (Revelle, 2025).

3.3.2. Entitativity as a function of group characteristics (Hypothesis 1)

We first examined whether Hypothesis 1 was supported, as it was in Study 1: did participants evaluate a group as entitative to the degree that group members displayed similar facial emotions? The dependent variable in this analysis was entitativity judgments formed after 3 s of group-viewing time with faces visible (the value closest to the 5 s duration in Study 1). We regressed this variable on actual emotion similarity, controlling for race and gender homogeneity. Replicating Study 1, actual emotion similarity was predictive of entitativity, $b = 0.14$, $SE = 0.05$, $t\,(181) = 2.93$, $p = .004$, $\beta = 0.21$.

In image-based analyses, gender homogeneity also predicted entitativity, but race homogeneity did not [Gender: $b = 1.46$, $SE = 0.70$, $t\,(181) = 2.10$, $p = .038$, $\beta = 0.15$; Race: $b = 1.18$, $SE = 0.72$, $t\,(181) = 1.64$, $p = .103$, $\beta = 0.12$]. These findings largely replicate those of Study 1. Most importantly, however, actual emotion similarity emerged as a significant predictor in both MLM and image-based analyses when accounting for race and gender composition.

For exploratory purposes, we next ran the same regression on entitativity judgments formed after 150 ms and 750 ms viewing times. With 150 ms viewing times, actual facial emotion similarity was associated with increased entitativity, $b = 0.08$, $SE = 0.04$, $t\,(181) = 1.98$, $p = .049$, $\beta = 0.14$. Gender homogeneity also significantly predicted entitativity, $b = 1.50$, $t\,(181) = 2.73$, $p = .007$, $\beta = 0.20$. Race homogeneity did not, $b = 0.80$, $SE = 0.57$, $t\,(181) = 1.40$, $p = .162$, $\beta = 0.10$. With 750 ms viewing times, actual facial emotion similarity was predictive of entitativity, $b = 0.11$, $SE = 0.04$, $t\,(181) = 2.38$, $p = .019$, $\beta = 0.17$. Gender and race homogeneity also significantly predicted entitativity (Gender: $b = 1.34$, $SE = 0.64$, $t\,(181) = 2.10$, $p = .037$, $\beta = 0.15$; Race: $b = 1.40$, $SE = 0.66$, $t\,(181) = 2.12$, $p = .036$, $\beta = 0.15$). In summary, to the extent that members of a group exhibited similar facial emotions, that group was rated as highly entitative—even when the group was only seen for only 150 ms. Additionally, to the extent that groups shared gender or racial categories, those groups also tended to be evaluated as more entitative. In short, we find replicable support for Hypothesis 1.

3.3.3. Entitativity from ensemble perceptions of facial emotion (Hypothesis 2)

Before evaluating Hypotheses 2–3, we tested whether people produced high-fidelity ensemble perceptions of natural crowds. Recall that ensemble perception occurs when people accurately estimate the characteristics of a group seen too briefly to allow for close analysis of all constituents. We examined whether actual emotion similarity predicted perceived emotion similarity, even after controlling for perceived non-face emotion similarity. Indeed, actual facial emotion similarity remained a significant predictor of perceived facial emotion similarity,

$b = 0.05$, $SE = 0.02$, $t\,(182) = 2.55$, $p = .012$, $\beta = 0.12$.[6] However, this effect should be interpreted cautiously given that the β is below the minimum detectable effect for this study. We tentatively conclude that perceivers were accurate in rapidly perceiving within-group variability even in the context of scene and body information. Unlike studies with experimenter-generated stimuli, we find that this effect holds even when perceptions of scene and body information are controlled for. This finding sets the stage for the remaining hypothesis tests.

To test *Hypothesis 2*, we examined whether rapid ensemble perceptions of facial emotion similarity—formed after only 150 ms of group-viewing time—significantly predicted entitativity ratings formed *after 3 s of group-viewing time* while controlling for ensemble perceptions of non-facial emotion similarity. We regressed entitativity (3 s; faces visible) on rapid ensemble perceptions of emotion similarity with faces visible and occluded (150 ms). Even when controlling for perceived emotion similarity drawn from non-face information, rapid ensemble perceptions of *facial* emotion similarity—formed after only 150 ms of group-viewing time—significantly predicted entitativity ratings formed after 3 s of group-viewing time, $b = 0.93$, $SE = 0.14$, $t\,(182) = 6.46$, $p < .001$, $\beta = 0.57$. Conversely, in this model, non-face estimation of emotion similarity did not significantly predict entitativity, $b = 0.12$, $SE = 0.14$, $t\,(182) = 0.85$, $p = .397$, $\beta = 0.07$. These findings provide strong support for Hypothesis 2.

Given support for Hypotheses 1 and 2, we examined the degree to which ensemble perceptions of emotion (150 ms exposure time) mediated the relationship between actual crowd emotion similarity and judgments of entitativity (3 s exposure time). As described in Study 1, statistical mediation analyses were employed to examine proof of concept for our causal theory, rather than using these analyses to argue for causality in the sample data. As theorized, with greater actual emotion similarity in a natural crowd, the more emotion similarity perceivers detected, and the more entitative the group seemed. As in Study 1, we used mediation (Revelle, 2025) to examine the indirect effect of actual emotion similarity on entitativity *through* perceived emotion similarity. This indirect effect was significant, $b = 0.10$, $SD = 0.04$, 95% CI [0.04, 0.20], indicating that ensemble perceptions of emotion similarity are a plausible mediator for the effect of actual emotion similarity on entitativity.[7]

3.3.4. Entitativity: rapid vs. extended perceptions of emotion similarity (Hypothesis 3)

To test *Hypothesis 3*, we examined how deliberative judgments of entitativity (made after 3 s of exposure to the group) were predicted by perceived emotion similarity formed after viewing a group for 150 ms vs. 750 ms vs. 3 s. We expected that entitativity ratings made after 3 s of viewing time would be predicted by perceived emotion similarity formed after 150 ms but that this predictive relationship would be stronger than when participants had more time to perceive emotion similarity (750 ms or 3 s).

Controlling for non-face information, perceived *facial* emotion similarity drawn from 750 ms of group-viewing time significantly predicted entitativity (3 s; faces visible), $b = 0.44$, $SE = 0.11$, $t\,(182) = 4.18$, $p < .001$, $\beta = 0.35$. Note, however, that consistent with Hypothesis 3, this regression weight is considerably smaller than the same weight applied to perceived facial emotion similarity drawn from 150 ms viewing times ($\beta = 0.57$). Perceived *non-facial* emotion similarity drawn from 750 ms of group-viewing time significantly predicted entitativity, $b = 0.55$, $SE = 0.13$, $t\,(182) = 4.32$, $p < .001$, $\beta = 0.36$. Note that this regression weight is considerably *larger* than the same weight applied to 150 ms viewing times ($\beta = 0.07$). Thus, when viewing times for perceived emotion similarity increased from 150 ms to 750 ms, the contribution of rapid ensemble perceptions to entitativity judgments was reduced (as predicted) and the contribution of non-face emotional information increased (as predicted). This pattern of evidence is consistent with Hypothesis 3.

This same pattern of results was observed for perceived emotion similarity formed after 3 s of exposure. Perceived *facial* emotion similarity from 3 s of viewing time significantly predicted entitativity ratings from 3 s of viewing time, $b = 0.29$, $SE = 0.10$, $t\,(182) = 2.81$, $p = .006$, $\beta = 0.21$. Perceived *non-facial* emotion similarity after 3 s of viewing also significantly predicted entitativity ratings after 3 s of viewing, $b = 0.69$, $SE = 0.12$, $t\,(182) = 5.90$, $p < .001$, $\beta = 0.45$. Thus, when exposure time for perceived emotion similarity increased from 150 ms ($\beta = 0.57$) to 750 ms ($\beta = 0.35$) or 3 s ($\beta = 0.21$), the contribution of rapid ensemble perceptions to entitativity was reduced (as predicted by Hypothesis 3) and the contribution of non-face emotional information increased (also as predicted, βs = 0.07, 0.36, 0.45, respectively). These findings (see Table 4) support Hypothesis 3 that entitativity judgments are shaped by *rapid* ensemble perceptions of facial emotion: longer viewing times allow for substantial face-by-face analysis of emotion and, as such, reflect a decaying influence from ensemble perception (formed in the initial moments of perception).

3.3.5. Entitativity as a function of viewing time (Hypothesis 4)

We next tested Hypothesis 4, that entitativity judgments made after a glance at a group will approximate judgments made after a more deliberated perusal of the group. We thus regressed entitativity ratings from extended viewing of a group (3 s, face included) on two predictors: (a) entitativity ratings from briefly-viewed (150 ms) natural crowds with faces visible and (b) entitativity ratings from briefly-viewed (150 ms) natural crowds with faces occluded. In this equation, "a" was a significant predictor ($b = 0.86$, $SE = 0.09$, $t\,(182) = 9.54$, $p < .001$, $\beta = 0.68$), suggesting that rapid visual processing of uniquely facial information shaped downstream judgments of entitativity. But face processing alone cannot explain the rapid formation of entitativity judgments: "b" was also a significant predictor ($b = 0.27$, $SE = 0.08$, $t\,(182) = 3.35$, $p < .001$, $\beta = 0.24$), suggesting that rapid visual processing of non-face information also shaped downstream judgments of entitativity. Combined, the two variables accounted for 80% of the variance (R^2) in entitativity ratings made after 3 s of group viewing time. Thus, entitativity judgments were formed *because* they were based largely—but not in full—on rapid visual perception of faces.

Notably, entitativity judgments made after 750 ms were slightly more predictive of deliberated entitativity judgments (3 s exposure) [$r\,(183) = 0.95$, $p < .001$] than were entitativity judgments made after 150 ms [$r\,(183) = 0.89$, $p < .001$), $z = 3.91$, $p < .001$]. Thus, judgments of entitativity did recruit some information from crowd images that was not included in the 150 ms judgments, suggesting that entitativity judgments were mostly but not *completely* formed after seeing a group for 150 ms. We regard these findings as support for Hypothesis 4. Although this effect is correlational, analogous effects have been regarded as evidence for the time-course of social perception and social judgments: for example, in one set of studies, trustworthiness judged after 39 ms of exposure to a face was roughly equivalent to the same judgments made after extended viewing (1700 ms; Bar et al., 2006). Such results have been replicated with different paradigms (Todorov et al., 2009), and are typically interpreted as evidence for the time-course of social perception.

[6] We conducted follow-up analyses with the average emotion of the crowd included as a moderator, which are detailed in the Supplemental Materials. In brief, the interaction between perceived emotion similarity and image valence was not significant in Study 1, but emerged in Study 2, such that perceived emotion similarity more strongly predicted entitativity for negative (vs. positive) images.

[7] Again, this model does not rule out that the mediator and outcome are flipped, that objective emotion similarity in the images influences entitativity directly and has downstream influences on perceived similarity, $b = 0.06$, $SD = 0.02$, 95% CI [0.04, 0.11].

Table 4
Study 2 beta (β) weights.

Predictor	Exposure Time for Predictor Judgment	Outcome: 3 s Entitativity Ratings	
		Faces Included	Faces Occluded
Judged Emotion Similarity	150 ms	β = 0.57***	β = 0.07
	750 ms	β = 0.35***	β = 0.36***
	3 s	β = 0.21*	β = 0.45***
Judged Entitativity	150 ms	β = 0.68***	β = 0.24***
	750 ms	β = 0.76***	β = 0.21***

Note. β weights from regression equations with two predictors: perceived emotion similarity of crowds with faces visible and occluded. A separate regression equation was constructed for each of the three *predictor* exposure times: 150 ms, 750 ms, and 3 s. *$p < .05$, ***$p < .001$.

3.4. Discussion

In Study 2, we observed broad support for our theoretical framework, isolating face perception by controlling for perception of visual information that was not specific to faces. In general, we found that facial emotion plays a substantial role in how people evaluate natural crowds. First, we replicated effects observed in Study 1; within-group variability in facial emotion negatively predicted entitativity (Hypothesis 1). Additionally, rapid ensemble perceptions of group variability in facial emotion (150 ms viewing time) were accurate, despite the many obstacles to ensemble perception of natural crowds, outlined earlier. Importantly, we also observed evidence consistent with Hypotheses 2–4. In real social scenes—where bodies, objects, and scene information are available to the visual system—rapid (150 ms) ensemble perceptions of *facial* emotion similarity shaped downstream judgments of a group's entitativity (Hypothesis 2), above and beyond non-face context. Conversely, rapid (150 ms) perceptions of *non-facial* emotion information had little if any influence on downstream judgments of entitativity. This effect weakened for perceptions of shared facial emotion formed from more than 150 ms of viewing time, consistent with the view that rapid ensemble perception processes can provide a basis for entitativity judgments; the relationship between perceived facial emotion similarity and entitativity weaken with longer viewing times as ensemble perceptions are degraded by noise when perceivers are given more time to analyze the crowd (Hypothesis 3). In contrast, perceptions of *non-facial* information shaped entitativity judgments through later visual processes. Moreover, entitativity judgments of visual groups seen for only 150 ms approximated entitativity judgments for groups seen for 3 s (Hypothesis 4). Put differently, entitativity judgments made after 150 ms of group exposure accounted for 79% of the variance in entitativity judgments made after 3 s of group exposure, despite being just 5% of the viewing time. Thus, the initial visual percept may be an efficient means for perceivers to generate a nearly "finalized" entitativity judgment.

4. General discussion

Ensemble perception is a perceptual phenomenon in which people can identify the characteristics of a group without closely inspecting its members. We theorized that ensemble perception processes can play a foundational role in evaluating entitativity (i.e., the extent to which a collection of people constitutes a group entity). We focused our efforts on ensemble perceptions of emotion similarity and their role in entitativity judgments of natural crowds. We tested four hypotheses drawn from this theory; results across our two studies collectively supported these hypotheses.

In Study 1, we observed that entitativity was influenced by the degree to which group members exhibited similar emotion, controlling for race and gender homogeneity (supporting Hypothesis 1). Importantly, about half of the variance in entitativity was explained by perceived emotion similarity. Finally, mediation analyses provided preliminary evidence supporting the theory that ensemble perceptions of emotion similarity explain the impact of *actual* emotion similarity on entitativity judgments. In Study 2, we examined the role of timing and face perception in entitativity judgments by quantifying (and then controlling for) the role of non-facial information. The data from the non-face condition is a robust control, as each image in this control included all of the non-face information in an image. Accordingly, any unique effects in the "faces visible" condition can be attributed to face perception in particular, controlling for judgments about anything in an image that was *not* a face, such as body or scene information. Participants exhibited accuracy in their rapid (150 ms) ensemble perceptions of facial emotion similarity, even when controlling for the contribution of non-facial information. Moreover, these rapid ensemble perceptions of facial emotion similarity influenced entitativity judgments (formed after 3 s), again controlling for the impact of non-facial information (Hypothesis 2). We observed strong support for Hypothesis 3: when controlling for the impact of non-face information, rapid ensemble perceptions of facial emotion similarity *better* predicted entitativity judgments than did more extended perceptions of facial emotion similarity. Finally, we found that rapid entitativity judgments are highly predictive of those made after extended viewing, and that perceivers used face and non-face information to form those rapid judgments.

4.1. Implications

We observed initial support for our theory that ensemble perception processes underlie entitativity judgments of natural crowds. These findings have implications for both social psychology and vision science, as detailed below.

Emergent perceptions and conceptions of social entities. People respond to "entities" in their environment and the field of social psychology is largely dedicated to understanding how people respond to *social* entities – both individuals and groups. It is thus critical for scientists to understand *how* people perceive and identify entities and to understand the consequences of doing so. For several centuries, scientists have examined the visual processes involved in identifying an object (DiCarlo et al., 2012), person (Hu et al., 2020), array (Wagemans et al., 2012), or the gist of a scene (Oliva & Torralba, 2006) as an entity. Drawing from this literature, Campbell (1958) introduced the concept of *entitativity* to social psychologists as a means for them to identify a given collection of people as a social group entity. Specifically, drawing from Gestalt approaches to vision, he argued that scientists might regard people as a "group" to the extent that those people display similar behavior, move in the same direction, and so on.

Subsequent studies applied Campbell's approach to lay conceptions of entitativity (Brewer et al., 2004; Lickel et al., 2001), finding that entitativity judgments were shaped by the degree to which people believed a collection of individuals interacted frequently, had shared goals or agency, and so on. Yet in most such studies, participants were given very little to "perceive" apart from the words in a description or label of people. Accordingly, the relationship between perceptions and conceptions of social entities—foreshadowed by Campbell's (1958) approach—remained obscure. The current work aimed to demystify this relationship. Moving beyond prior studies on collections of amoeba-like entities (Ip et al., 2006), stick figures (Lakens, 2010), or disembodied faces (Magee & Tiedens, 2006), we examined the relationship between perceptions of and cognitions about entitativity using information-rich images of natural crowds: we identified a visual process that appears to account for entitativity evaluations of visible groups.

We focused our hypotheses on the visual process of ensemble perception through which people arrive at rapid and precise perceptions of a group. These perceptions so rapid they cannot be reduced to sequential and deliberative processing of all individuals in the group, but are nonetheless highly accurate. Based on ample evidence that experiences and judgments of emotion similarity in a group increase group commitment and judgments of group cohesion, the two studies we

conducted focused on ensemble perceptions of emotion similarity. Evidence from two studies suggested that rapid ensemble perceptions of emotion similarity can strongly shape entitativity judgments, thus pointing to a visual process that can gate the translation of *perceptions* to *conceptions* of groups (Phillips et al., 2014).

Much remains to be learned (see below, "Limitations and Future Directions") about the relationship between what people see when observing a group and their beliefs about how entitative that group is. The current studies provide a basis for those scientific endeavors. We have provided: evidence for a visual mechanism involved in entitativity judgments of natural crowds; evidence that entitativity judgments of natural crowds can be generated from rapid visual processes that operate on coarse visual information gleaned from brief presentations; evidence for how *natural* variation in the properties of visible groups can influence entitativity judgments; and a *methodology* for examining how people evaluate entitativity in natural crowds. Given the role entitativity judgments can have in stereotyping, prejudice, and intergroup biases (Crawford et al., 2002; Dang et al., 2018; Newheiser et al., 2012; Yzerbyt & Estrada, 2003)—all of which are important to understand *because* of their implications for behavior toward the people they encounter in the world—the current work should have considerable import to scientists examining these topics, both theoretically and methodologically.

Emotion perception. It is well-established that people quickly perceive and respond to individual's facial emotions (e.g., Murphy & Zajonc, 1995), influencing a range of responses including how a perceiver negotiates, cooperates, and competes with others (e.g., van Kleef, 2009). People can also rapidly identify the facial emotions of groups, though the consequences of these perceptions of emotion remain poorly understood (but see Lamer et al., 2018). In the current work, we began to address this lacuna: we theorized and observed that ensemble perception is responsible for translating a group's facial emotion into entitativity judgments about that group. Although we limited our investigation to entitativity, it has been argued that entitativity is fundamental to both intergroup processes (e.g., stereotyping and prejudice; Roets & Van Hiel, 2011) and intragroup processes (e.g., group cohesion, group entry-exit; Blanchard et al., 2022). Thus, entitativity judgments are one possible consequence of collective emotion perception, but such judgments are also likely to gate a variety of other cognitions and behaviors. Indeed, as reviewed earlier, there is a close relationship between emotions and group functioning such that emotion similarity tends to be more prevalent within than between groups, and perceivers appear to be sensitive to this relationship. Following Campbell's classic essay on entitativity, we reasoned that the same processes (ensemble perception) that enable people to rapidly identify the emotions of a collective serve as the input for entitativity judgments.

Consistent with our theory, we observed that ensemble perceptions of emotion similarity—formed after only 150 ms of exposure to a social scene—are tightly correlated with entitativity judgments formed after a full 3 s of exposure to scenes that include a wide variety of meaningful social information. This relationship did *not* become stronger with more time to view the facial emotions in a crowd, and in some circumstances actually becomes weaker—these latter findings are especially important in that they illustrate the role of ensemble perception processes, as opposed to more detailed and piecemeal processing of every member in the collective (which is only possible with longer exposure times). Further, we find that entitativity judged after 150 ms of exposure to the scene were roughly equivalent to evaluations formed after a full 3 s. Together, these findings suggest that entitativity judgments formed from information available to perceivers with just a 150 ms glance at a group and critically, such effects were dominated by rapid processing of collective facial emotion variability.

These findings should have considerable import for affective science. Specifically, there exist several influential theories that aim to explain the relationship between group emotions and other group processes (e.g., Mackie et al., 2015; Parkinson, 2020). Although these theories differ in the cognitive and affective processes used to explain the emotion-group process relationship, visual perceptions of groups have received little emphasis in this literature. Yet in the current work we find that rapid ensemble *perceptions* of emotion can explain a great deal of the variance in entitativity judgments about visible groups. Given the fundamental role of entitativity in group processes, existing theories of the relationship between emotion and group processes should yield increased predictive validity by incorporating ensemble perception processes.

Ensemble perception. For several decades, vision scientists have examined ensemble perception (also called "ensemble coding" or "summary statistical perception") of low-level visual features. For example, when you see a "texture"—like the rough surface of a brick wall or the lush expanse of a manicured lawn— you are experiencing the tail end of a process that summarizes the visible edges of local details (e.g., lines and contours) into an emergent and global feature. More recently, vision scientists have also discovered that ensemble perception applies to high-level vision, including ensemble perceptions of groups of faces and bodies (Whitney & Yamanashi Leib, 2018). Although this body of work has nicely delineated how ensemble perception operates on visual input, the standard approach typically features reductionist designs aimed at isolating the mechanism or visible feature of interest (e.g., displaying faces or point-light-walkers on uniform backgrounds). Thus, it remained unclear whether ensemble perceptions have a meaningful impact on how people perceive natural collections of people seen in complex contexts, and social scientists have also been delayed in exploring the consequences of such ensemble perception for social judgment.

In the current work, we found that rapid ensemble perceptions of facial emotion similarity accurately reflected the actual facial emotion similarity displayed in a group. Moreover, we found that such rapid ensemble perceptions have a profound influence on the extent to which perceivers believed a visible group was an entity, as opposed to a collection of individuals. Together, these findings suggest that ensemble perception not only operates on the groups that people see in their chaotic visual environments, but also has an important influence on the impressions people form of those groups. As such, the current work provides a bridge between vision science and social psychology in the study of groups, and supports the importance of ensemble perception to understanding social cognition.

4.2. Limitations and future directions

Readers of this work will have noticed a variety of methodological details that depart from contemporary studies of face perception, of ensemble perception, and of entitativity. Although some of those details represent methodological advances for "leveling up" from reductionist designs, other details should be considered limitations of the current work.

One such limitation is with respect to our measure of ensemble perception, which (unlike some prior studies) did not compare the accuracy of ensemble perceptions with accuracy in perceptions of *individual* group members. Instead, we operationally defined such ensemble perception in terms of how long participants had to view the crowd, in that it is not possible for participants to closely analyze multiple members of a group in 150 ms so any accuracy must be due to ensemble perception processes that summarize across them. Scientists have frequently provided support for this assumption. For example, when crowds were presented rapidly in prior research, participants did not retain information about individuals within the crowd (e.g., Haberman & Whitney, 2009) or ensemble perceptions were generated from only a few crowd members (e.g., Sweeny & Whitney, 2014). Moreover, the fact that participants in our studies were able to accurately summarize the facial emotion similarity of a *natural* crowd—in which faces were connected to bodies in a specific social context with many other objects present—from seeing that group for only 150 ms strongly suggests that ensemble perception processes were operating. Finally, participants

judged properties that do not exist within individuals (i.e., emotional similarity) and only exist at the group level. Combined with recent evidence that ensemble representations of summary information seem to be extracted automatically, even without visual awareness of a set's constituents (Ward et al., 2016), we consider it a safe assumption that ensemble mechanisms were at work in our studies and that we measured them with our designs. Nonetheless, a more thorough test of this assumption would bolster work in this area.

Another limitation regards how emotion was measured. In this set of studies, we used coding that allowed for more reliable estimates of *individual* face affect to serve as a criterion for ensemble perception accuracy. In real social settings, facial expressions often include a mixture of multiple categorical emotions (e.g., Ekman, 1992) that can be more accurately identified as positive or negative (Russell, 2003) than as (for example) disgust versus anger. However, an alternative approach would be to code each face for its categorical emotion (e.g., anger) rather than for affect (negative-positive). A next step will be to determine if categorical-emotion can be summary-perceived in natural crowds and if those categorical judgments have a greater impact on entitativity judgments than do the valence judgments used here. Indeed, there is evidence that it is the positive-negative dimension of facial emotion that is "perceptually" encoded, but that linguistic (discrete) categories of emotion become tied to emotion perception as language develops (see Lindquist et al., 2015). Furthermore, people often perceive continuous stimuli categorically (Fugate, 2013), an approach that enables rapid categorization of otherwise ambiguous stimuli (e.g., facial emotion; Calder et al., 1996; Etcoff & Magee, 1992). Given that the human mind often warps linear, continuous stimuli into more nonlinear, distinguishable categories (Goldstone & Hendrickson, 2010), categorical emotion judgments may better predict entitativity judgments, especially among perceivers with fully developed language capacities.

Another limitation regards how entitativity was measured. In many studies on entitativity, scientists measure components of entitativity (e.g., the specific belief that a given group has shared goals) and observe that each component has unique antecedents and consequences. In the current studies, we simply asked participants to evaluate the extent to which several people seemed like a cohesive group, as opposed to a collection of individuals. Our results thus do not lend support to any specific theory of entitativity, which differ widely in how conceptions of group factors such as shared goals, agency, interactivity, and essentialism are related to entitativity judgments. The measurement limitation in the current work is a limitation of both theory and of measurement. Our theory did not specify the more specific conceptual inferences that might mediate perceptions of natural, visible groups with broad entitativity judgments (though we suspect, "shared goals" or "interactivity," both of which are consistent with the approach of Campbell, 1958). Conversely, by integrating our theory with extant theories, new hypotheses may be formulated and tested, thereby contributing to a more precise scientific understanding. Methodologically, including a single question about entitativity may yield a less reliable estimate than would a more comprehensive set of items that capture all of the factors that play into entitativity judgments. A fair critique of the current work is that entitativity was measured by asking participants how much the individuals in each picture made up a cohesive group. Although this question is closely tied to the *scientific* definition of entitativity, it may not reflect the spontaneous judgments people make when viewing a group. Instead, people may spontaneously think about how well group members get along, hang out together, or have similar experiences. A more comprehensive approach that assesses entitativity with an array of questions, including those that more closely reflect spontaneous cognitions, may yield more precise conclusions than our studies, though this idea awaits testing.

A final limitation regards the natural crowd stimuli. The stimuli were sampled from college Instagram feeds (see Lamer et al., 2018) and have a range of gender and race homogeneity values, enabling us to control for these variables when considering emotion similarity. These stimuli offer substantive advances in theory, method, and application, as natural crowds exhibit variation in facial emotion, identity, scene, and posture. However, the stimuli included limited representation of older adults and children. The images were sampled from college Instagram feeds and thus primarily featured young adults. When age heterogeneity occurred, it typically stemmed from a single older adult being pictured with a group of young students, rendering the current stimuli unsuited to explore age effects. Nonetheless, tests of age warrant further investigation given how target age influences emotion perception. For example, perceivers tend to struggle more to identify the emotions of older than younger adults, an effect that is theorized to emerge due to age-related stereotypes and features of the face that change with age (Fölster et al., 2014). In our studies, the age of the perceiver or target did not interact with the extent to which perceived similarity predicted entitativity (see Supplemental Materials). That is, people did not evaluate groups close to their own age as more (or less) entitative, regardless of emotion similarity. However, the absence of an interaction may be due to limited age representation in the stimuli, which warrants caution in interpreting these results and highlights age homogeneity in groups as an avenue for future work.

5. Concluding remarks

What makes people think that a visible collection of individuals is a group *entity*? We observed evidence that entitativity judgments are driven by rapid ensemble perceptions of facial emotion similarity. These rapid ensemble perceptions accounted for the majority of variance in entitativity judgments drawn from extended inspection of groups, suggesting that rapid visual impressions from the immediate moments of perception can shape downstream entitativity judgments—even when those groups are seen in a rich social scene characteristic of how people encounter groups in their daily lives.

Author note

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Open practices

All data and analysis scripts are available here: https://osf.io/wtbqk/?view_only=277d6c64785d4e95834cf1c90a694955.

CRediT authorship contribution statement

Sarah Ariel Lamer: Writing – review & editing, Validation, Project administration, Formal analysis, Data curation. **Spencer Dobbs:** Writing – review & editing, Project administration, Data curation. **Lindsay Goolsby:** Writing – review & editing, Project administration, Investigation, Data curation. **Timothy D. Sweeny:** Writing – review & editing, Conceptualization. **Max Weisbuch:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used sentence-level editing suggestions from Grammarly and used ChatGPT to revise one paragraph of the manuscript. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2026.104889.

References

Adelman, L., Yogeeswaran, K., & Lickel, B. (2019). They're all the same, sometimes: Prejudicial attitudes toward Muslims influence motivated judgments of entitativity and collective responsibility for an individual's actions. *Journal of Experimental Social Psychology, 80*(November), 31–38. https://doi.org/10.1016/j.jesp.2018.10.002

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 21–62. https://doi.org/10.1177/1534582302001001003

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83*, 25–39. https://doi.org/10.1016/j.visres.2013.02.018

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., … Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science, 19*(7), 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science (American Association for the Advancement of Science), 338*(6111), 1225–1229. https://doi.org/10.1126/science.1224313

Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics, 82*, 63–79. https://doi.org/10.3758/s13414-019-01827-z

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269–278. https://doi.org/10.1037/1528-3542.6.2.269

Blanchard, A. L., McBride, A. G., & Ernst, B. A. (2022). How are we similar? Group level entitativity in work and social groups. *Small Group Research, 54*(3), 369–395. https://doi.org/10.1177/10464964221117483

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brewer, M. B., Hong, Y.-Y., & Li, Q. (2004). Dynamic entitativity: Perceiving groups as actors. In V. Yzerbyt, C. M. Judd, & O. Corneille (Eds.), *The psychology of group perception: Perceived variability, entitativity, and essentialism* (pp. 25–38). New York, NY: Psychology Press.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*(2), 81–118. https://doi.org/10.1080/713756735

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3*(1), 14–25.

Chen, Z., & Whitney, D. (2019). Tracking the affective state of unseen persons. *Proceedings of the National Academy of Sciences - PNAS, 116*(15), 7559–7564. https://doi.org/10.1073/pnas.1812250116

Chen, Z., & Whitney, D. (2020). Inferential emotion tracking (IET) reveals the critical role of context in emotion recognition. *Emotion (Washington, D. C.)*. https://doi.org/10.1037/emo0000934

Clark, M. S., & Finkel, E. J. (2005). Willingness to express emotion: The impact of relationship type, communal orientation, and their interaction. *Personal Relationships, 12*(2), 169–180. https://doi.org/10.1111/j.1350-4126.2005.00109.x

Clark, M. S., & Taraban, C. (1991). Reactions to and willingness to express emotion in communal and exchange relationships. *Journal of Experimental Social Psychology, 27*(4), 324–336.

Clifford, C. (2014). The tilt illusion: Phenomenology and functional implications. *Vision Research (Oxford), 104*, 3–11. https://doi.org/10.1016/j.visres.2014.06.009

Corbett, J. E. (2017). The whole warps the sum of its parts: Gestalt-defined-group mean size biases memory for individual objects. *Psychological Science, 28*(1), 12–22. https://doi.org/10.1177/0956797616671524

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology, 83*(5), 1076–1094. https://doi.org/10.1037/0022-3514.83.5.1076

Dang, J., Liu, L., Ren, D., & Gu, Z. (2018). "Groupy" allies are more beneficial while "groupy" enemies are more harmful. *Social Psychological and Personality Science, 9*(8), 925–934. https://doi.org/10.1177/1948550617729409

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science, 1*(1), 393–416.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3–4), 169–200.

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble perception of dynamic emotional groups. *Psychological Science, 28*(2), 193–203. https://doi.org/10.1177/0956797616678188

Elias, E., & Sweeny, T. D. (2020). Integration and segmentation conflict during ensemble coding of shape. *Journal of Experimental Psychology. Human Perception and Performance, 46*(6), 593–609. https://doi.org/10.1037/xhp0000733

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*(3), 227–240.

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*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, 47*(6), 1231–1236. https://doi.org/10.1016/j.jesp.2011.05.007

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106*(3), 1389–1398. https://doi.org/10.1152/jn.00904.2010

Fölster, M., Hess, U., & Werheid, K. (2014). Facial age affects emotional expression decoding. *Frontiers in Psychology, 5*, 30.

Fugate, J. M. (2013). Categorical perception for emotional faces. *Emotion Review, 5*(1), 84–89.

Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. USA: Oxford University Press.

Gibson, J. J. (1979). The ecological approach to visual perception. In *The Ecological Approach to Visual Perception*. Houghton Mifflin. https://doi.org/10.4324/9781315740218.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69–78.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718–734. https://doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception & Psychophysics, 72*(7), 1825–1838. https://doi.org/10.3758/APP.72.7.1825

Hamilton, D. L., Chen, J. M., Ko, D. M., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology, 109*(4), 569–588. https://doi.org/10.1037/pspa0000034

Heerdink, M. W., van Kleef, G. A., Homan, A. C., & Fischer, A. H. (2015). Emotional reactions to deviance in groups: The relation between number of angry group members and conformity. *Frontiers in Psychology, 6*, 858. https://doi.org/10.3389/fpsyg.2015.00858

Homan, A. C., Van Kleef, G. A., & Sanchez-Burks, J. (2015). Team members' emotional displays as indicators of team functioning. *Cognition and Emotion, 30*(1), 134–149. https://doi.org/10.1080/02699931.2015.1039494

Hu, Y., Baragchizadeh, A., & O'Toole, A. J. (2020). Integrating faces and bodies: Psychological and neural perspectives on whole person perception. *Neuroscience and Biobehavioral Reviews, 112*, 472–486. https://doi.org/10.1016/j.neubiorev.2020.02.021

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). *Designing clinical research: An epidemiologic approach* (4th ed., p. 79). Lippincott Williams & Wilkins. Appendix 6C.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception, 43*(7), 663–676. https://doi.org/10.1068/p7719

Ip, W. M., Chiu, C. Y., & Wan, C. (2006). Birds of a feather and birds flocking together: Physical versus behavioral cues may lead to trait-versus goal-based group perception. *Journal of Personality and Social Psychology, 90*(3), 368.

Jeong, J., & Chong, S. C. (2021). Perceived variability reflects the reliability of individual items. *Vision Research, 183*, 91–105.

Ji, L., Pourtois, G., & Sweeny, T. D. (2020). Averaging multiple facial expressions through subsampling. *Visual Cognition, 28*(1), 41–58. https://doi.org/10.1080/13506285.2020.1717706

van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18*(5), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Lakens, D. (2010). Movement synchrony and perceived entitativity. *Journal of Experimental Social Psychology, 46*(5), 701–708. https://doi.org/10.1016/j.jesp.2010.03.015

Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition, 29*(1), 1–14. https://doi.org/10.1521/soco.2011.29.1.1

Lamer, S. A., Sweeny, T. D., Dyer, M. L., & Weisbuch, M. (2018). Rapid visual perception of interracial crowds: Racial category learning from emotional segregation. *Journal of Experimental Psychology: General, 147*, 683–701. https://doi.org/10.1037/xge0000443.

Lickel, B., Hamilton, D. L., & Sherman, S. J. (2001). Elements of a lay theory of groups: Types of groups, relational styles, and the perception of group entitativity. *Personality and Social Psychology Review, 5*(2), 129–140. https://doi.org/10.1207/S15327957PSPR0502_4

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(2), 223–246. https://doi.org/10.1037/0022-3514.78.2.223
Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*(2), 375.
Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology, 6*, Article 121301. https://doi.org/10.3389/fpsyg.2015.00444
Mackie, D. M., Maitner, A. T., & Smith, E. R. (2015). Intergroup emotions theory. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (2nd ed., pp. 149–174). Taylor & Francis. https://doi.org/10.4324/9780203361993-12.
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*(12), 1703–1715. https://doi.org/10.1177/0146167206292094
McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*(3), 215–238. https://doi.org/10.1037/0033-295X.90.3.215
Morgan, M. J. (1999). The Poggendorff illusion: A bias in the estimation of the orientation of virtual lines by second-stage filters. *Vision Research (Oxford), 39*(14), 2361–2380. https://doi.org/10.1016/S0042-6989(98)00243-0
Murphy, S. T., & Zajonc, R. B. (1995). Additivity of nonconscious affect: Combined effects of priming and exposure. *Journal of Personality and Social Psychology, 69*(4), 589–602. https://doi.org/10.1037/0022-3514.69.4.589
Newheiser, A. K., Sawaoka, T., & Dovidio, J. F. (2012). Why do we punish groups? High entitativity promotes moral suspicion. *Journal of Experimental Social Psychology, 48*(4), 931–936. https://doi.org/10.1016/j.jesp.2012.02.013
Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2
Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press.
Parkinson, B. (2020). Intragroup emotion convergence: Beyond contagion and social appraisal. *Personality and Social Psychology Review, 24*(2), 121–140. https://doi.org/10.1177/1088868319882596
Peterson, M. A., & Kimchi, R. (2013). Perceptual organization in vision. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 9–31). Oxford University Press.
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.
Revelle, W. (2025). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. R package version 2.5.6 https://CRAN.R-project.org/package=psych.
Roets, A., & Van Hiel, A. (2011). The role of need for closure in essentialist entitativity beliefs and prejudice: An epistemic needs approach to racial categorization. *British Journal of Social Psychology, 50*(1), 52–73. https://doi.org/10.1348/014466610X491567
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172.
Rydell, R. J., Hugenberg, K., Ray, D., & MacKie, D. M. (2007). Implicit theories about groups and stereotyping: The role of group entitativity. *Personality and Social Psychology Bulletin, 33*(4), 549–558. https://doi.org/10.1177/0146167206296956
van Schaik, J. E., Endenburg, L., & Hunnius, S. (2011). Goal-directed gaze shifts during the observation of everyday intentional actions. *Frontiers in Psychology, 2*, 264. https://doi.org/10.3389/fpsyg.2011.00264
van der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion, 11*(2), 286. https://doi.org/10.1037/a0022592
Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). Can emotions be truly group level? Evidence regarding four conceptual criteria. *Journal of Personality and Social Psychology, 93*(3), 431.
Spencer-Rodgers, J., Hamilton, D. L., & Sherman, S. J. (2007). The central role of entitativity in stereotypes of social categories and task groups. *Journal of Personality and Social Psychology, 92*(3), 369–388. https://doi.org/10.1037/0022-3514.92.3.369
Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science, 25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510
Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813
Troje, N. F., & Chang, D. H. (2023). Life detection from biological motion. *Current Directions in Psychological Science, 32*(1), 26–32.
Utochkin, I. S., & Brady, T. F. (2020). Individual representations in visual working memory inherit ensemble properties. *Journal of Experimental Psychology: Human Perception and Performance, 46*(5), 458.
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*(6), 1172–1217. https://doi.org/10.1037/a0029333
Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition, 152*, 78–86. https://doi.org/10.1016/j.cognition.2016.01.010
Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology, 95*(5), 1063–1079. https://doi.org/10.1037/a0011993
Westheimer, G., & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vision Research, 27*(8), 1361–1368. https://doi.org/10.1016/0042-6989(87)90212-4
Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: Multiple levels of summary statistical perception. In *The new visual neurosciences* (pp. 695–710).
Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005
Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232
Yzerbyt, C. O., & Estrada, C. (2003). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review, 5*(2), 141–155. https://doi.org/10.1207/S15327957PSPR0502_5