



OPEN

Sex ratios in vocal ensembles affect perceptions of threat and belonging

Kelsey L. Neuenswander¹✉, Brianna M. Goodale², Gregory A. Bryant¹ & Kerri L. Johnson^{1,3}

People often interact with groups (i.e., ensembles) during social interactions. Given that group-level information is important in navigating social environments, we expect perceptual sensitivity to aspects of groups that are relevant for personal threat as well as social belonging. Most ensemble perception research has focused on visual ensembles, with little research looking at auditory or vocal ensembles. Across four studies, we present evidence that (i) perceivers accurately extract the sex composition of a group from voices alone, (ii) judgments of threat increase concomitantly with the number of men, and (iii) listeners' sense of belonging depends on the number of same-sex others in the group. This work advances our understanding of social cognition, interpersonal communication, and ensemble coding to include auditory information, and reveals people's ability to extract relevant social information from brief exposures to vocalizing groups.

Keywords Ensemble coding, Group perception, Auditory perception, Social cognition

Social cognition research has recently expanded to investigate *people* perception in addition to *person* perception^{1,2}—groups of people have perceptible qualities that are separable from the characteristics of the individuals that make them up. Much of human social interaction and communication involves groups of people. For example, full-time college students often spend up to 20 h per week in a classroom. The average U.S. citizen spends over 50,000 h with coworkers throughout their lifetime³. Many individuals are involved in teams, clubs, and other social groups. Outside of formal group interactions, people perceive groups of strangers at restaurants, live shows, and sporting events. Thus, it is important to understand how groups (i.e., ensembles) are perceived.

Most research investigating ensemble perception has focused on the visual system's ability to extract summary statistical information from a group, often in a brief glance⁴. Perceivers can accurately extract summary statistics of ensembles from low-level features such as hue^{5,6}, brightness⁷, orientation^{8–10}, spatial position^{11–13}, and motion^{14,15}, as well as mid-level features such as size^{16,17}. Perceivers can also extract high-level features from ensembles including emotion¹⁸, family resemblance^{19–22}, walk motion²³, eye-gaze direction²⁴, and social category membership^{25–28}. Importantly, the social category membership of a group influences evaluations. Groups with more men are perceived as more threatening²⁵ and more likely to harbor sexist norms²⁶, whereas perceivers feel greater belonging in groups with more same-sex others²⁶. Threat and belonging are particularly important social judgments for marginalized group members and can have critical impacts on welfare and safety. These judgments also influence adaptive behavior. When a group is perceived as threatening, individuals often avoid it^{29–31}. When individuals feel that they fit or belong within a group, they tend to approach it^{32–34}. Overall, observers efficiently get the gist of groups from their visual properties which has implications for group evaluations and subsequent social action.

Auditory or vocal ensembles are relatively understudied compared to visual ensembles. However, vocal ensembles are an important component of overall ensemble perception and present an adaptive problem for many species, including humans. Most generally, visual information is not always available. For example, in many environments, especially at night, groups can be heard but not seen, requiring the ability to extract information about them from only voices and other sounds. We should expect people to be attuned to auditory ensembles and be able to draw rapid inferences regarding socially relevant features such as whether the groups constitute possible adversaries or allies. Moreover, when visual information is available, it is sometimes obscured. In large group scenarios such as in the examples provide earlier (i.e., restaurants, live shows, and sporting events), we

¹Department of Communication, University of California, Los Angeles, 2225 Rolfe Hall, Los Angeles, CA 90095, USA. ²Julius Clinical, Zeist, The Netherlands. ³Department of Psychology, University of California, Los Angeles, USA. ✉email: kneuenswander@ucla.edu

often hear group members before we are close enough to see them. In academic and professional settings, the use of telework on platforms such as Zoom, Microsoft Teams, and Webex has grown drastically in response to the Covid-19 pandemic^{35,36}. Visual information is not always shared on these platforms—to combat “Zoom fatigue,” researchers are suggesting people keep their cameras off³⁷. Overall, individuals must rely on auditory information when visual information is nonexistent or suboptimal, so we should expect an evolved auditory perceptual sensitivity to group-level information.

Despite vocal ensembles being relatively understudied, ensemble perception is expected to operate similarly across sensory modalities. Listeners can extract summary statistics of a group from low-level auditory features such as pure tones³⁸ and sound textures³⁹, as well as high-level features such as social category membership. For instance, perceivers use sexually dimorphic aspects of the voice (e.g., pitch and timbre) to accurately judge the ratio of men to women in a group after listening to very brief (1500 ms) audio recordings of 5 or 10 simultaneous voices^{40,41}. However, no research has investigated whether summary statistics of auditory information have downstream consequences for social judgments of threat or belonging. Given that sex composition information affects social evaluative judgments in visual ensembles, and people can extract sex ratios from auditory ensembles, we should expect an effect of sex composition in vocal ensembles on social evaluations of vocalizing groups. The following studies address this by testing if listeners can extract the sex composition of a group from voices alone (Studies 1–3) and whether this influences perceivers’ feelings of threat (Study 2) and belonging (Study 3) toward the group.

In this paper, we differentiate between sex and gender. We refer to voice sex to reflect sexual dimorphism in voice properties, including fundamental frequency (f_0 ; the acoustic correlate of perceptual pitch) and formants. However, because all voice recordings are from cisgender individuals, we also use the terms “men” and “women” to discuss group composition. For masculinity and femininity judgments, we refer to gender due to the full spectrum this encompasses. Participants only provided their gender identity and thus we refer to participant gender rather than sex.

Study 1a

The primary aim of Study 1a was to demonstrate that listeners can accurately extract the sex ratio of a group from voices alone. This study extends the existing research investigating the perception of vocal sex ratios in five distinct ways^{40,41}. First, vocal ensembles that consisted of twelve people rather than five or ten people were used to test whether listeners accurately extract summary statistics from larger groups. Second, sequential rather than simultaneous recordings of ensembles were used to emulate turn-taking scenarios in academic or professional settings (e.g., classroom, boardroom, panel). The decision to use sequential stimuli aligns with previous studies investigating auditory ensemble coding^{38,42,43}. Third, the duration of our ensembles differed due to our sequential approach. While the overall length is longer than previously tested vocal ensembles (3000 ms vs. 1500 ms), the length of each individual voice recording within the ensemble is shorter (250 ms vs 1500 ms). In the visual literature, observers accurately extracted the sex ratio of a group after a mere glimpse^{25,26}. Similarly, we tested if people could extract summary statistics of a group after hearing a brief snippet of their voices. Fourth, semantic content was controlled by using audio samples that used the same utterance (“hi”). Fifth, an additional dependent measure was collected to determine if accuracy estimates are cross modal; that is, whether the sex ratio of the group presented auditorily influenced the estimated visual appearance of the average group member. It is possible that low-level perceptual features of the voice activate mental representations of social categories that transcend sensory domains.

Participants listened to 50 vocal ensembles that varied in the ratio of men to women (0:12, 3:9, 6:6, 9:3, 12:0) and estimated the sex ratio and average facial appearance of the group. We predicted that our results would replicate and extend findings reported previously^{40,41}. Specifically, we made the following predictions: (i) as the actual ratio of men to women in a group increases, the perceived ratio of men to women in a group will also increase, and (ii) as the actual ratio of men to women in a group increases, the perceived average group member will become more facially masculine.

Method

All studies were approved by the UCLA Institutional Review Board and were performed in accordance with relevant guidelines and regulations. Data and analysis code are publicly available on OSF (<https://osf.io/28gvd/>).

Participants

An a priori power analysis was run to determine the recommended sample size for a CNC within-subjects linear effects mixed model with actual sex ratio of the group as a fixed effect and participants and targets as random intercepts. The letters N and C indicate whether the random factors in each pair are nested or crossed, respectively⁴⁴. The first letter indicates that participants are crossed with actual sex ratio, the second letter indicates that targets are nested within actual sex ratio, and the third letter indicates that participants and targets are themselves crossed.

A priori power analyses for linear mixed models are complex given they require knowing numerous parameters such as the number of level-one groups, the estimated effect size, variance of random effects, covariance of random effects, regression coefficients, and the design effect^{45–49}. To address these complexities, we were conservative in our sample size estimates. Our analysis determined that a sample size of 64 participants was needed to detect an effect size of 0.50 with 80% power and an alpha level of 0.05⁴⁶. We exceeded this target number and recruited 96 individuals (Gender: 59% women, 39% men, 2% genderqueer; Race: 72% White, 12% Black, 9% Asian, 7% biracial/other; Age: $M = 38.77$ years, $SD = 14.63$ years, $\min = 19$, $\max = 71$) through an online participant pool, Prolific (<https://www.prolific.co/>), who were paid \$3.00 for their participation in a 20-min study. We

initially included participant gender as a potential moderator in all studies. However, participant gender did not significantly moderate any results in Studies 1 and 2 and was therefore dropped from the analyses.

Stimuli

Audio clips were sampled from voice recordings used in prior research (e.g.⁵⁰), and also unpublished work. The voice recordings consisted of 104 individuals (52 cisgender men, 52 cisgender women) saying “Hi, I’m a student at UCLA”. Voices were recorded in a quiet room on a digital (16 bit, 44.1 kHz) recorder (Marantz PMD-660 or M-Audio MicroTrack 24/96) with a cardioid condenser microphone (AKG C535 EB) 15–20 cm away from the mouth.

Audio clips were trimmed to 250 ms in length and only included the word “hi.” A custom Python script was created to generate ensembles, such that voices were randomly selected within trial, but with replacement between trials, from the available bank of voice stimuli. Ensembles consisted of twelve voices that played sequentially. The order of the voices was randomized within each ensemble and varied in the ratio of men to women (0:12, 3:9, 6:6, 9:3, 12:0). A set of 10 unique ensembles was generated for each of the 5 possible sex ratios, resulting in a total of 50 ensembles that were presented to each participant.

Procedure

After completing an informed consent form, participants verified that they were wearing headphones and in a quiet environment. The study protocol consisted of two blocks presented in counterbalanced order. Participants completed 50 trials in each block, for a total of 100 trials. On each trial, participants listened to an ensemble and provided judgments. In one block, participants were asked to estimate the ratio of men to women in the ensemble using a stick figure scale (Fig. 1). In the other block, participants were asked to estimate what the average group member looked like on a gender morph continuum from very masculine to very feminine (Fig. 2). After completing these two blocks, participants provided us with demographic information and were debriefed.

Results

The R packages “lme4” and “lmerTest” were used to create hierarchical linear models that accounted for within-subject variation and nesting within participants^{51,52}. For all studies, we created models using the CNC design described above⁴⁴ with actual sex ratio as a fixed effect and participants and targets as random intercepts. Significance was determined using traditional cutoff values of $\alpha = 0.05$, and we conducted thorough checks for assumptions in our models, including normality assessments, to ensure there were no severe violations.

First, we tested whether participants were accurate in their perceptions of ensembles by regressing the estimated number of men in the group onto the actual sex ratio. Indeed, as the ratio of men to women increased, so did participants’ numeric estimates of men in an ensemble, $B = 0.81$, $SE = 0.04$, $t(129) = 20.54$, $p < 0.001$. Next, we tested whether the actual number of men in each ensemble affected how masculine or feminine the average group member was perceived to be. As the number of men in the group increased, participants estimated that the average group member was more facially masculine, $B = 0.88$, $SE = 0.03$, $t(108) = 33.20$, $p < 0.001$.

Discussion

The results of Study 1 supported both of our hypotheses. As the actual ratio of men to women in a group increased, so did participants’ estimated sex ratio. Furthermore, as the actual number of men to women in a group

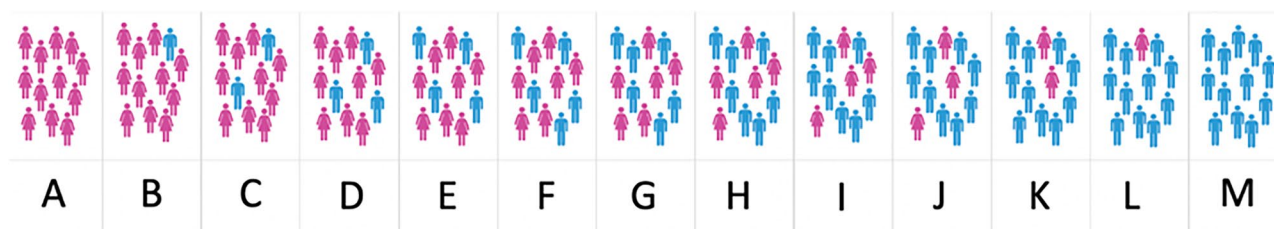


Figure 1. Sex ratio scale. Note: Which picture (labeled by letter) best represents the ratio of men to women you just heard?

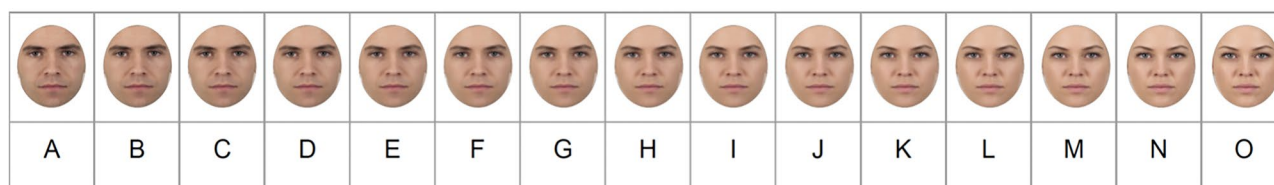


Figure 2. Gender morph scale. Note: Which picture (labeled by letter) best represents the average group member?

increased, so did the perceived facial masculinity of the group. In sum, listeners were calibrated to the group's sex ratio after briefly hearing their voices and this had cross-modal influence on predicted group appearance.

Given the sequential nature of our stimuli, it is possible that extracting the sex ratio from the group was the result of participants counting the number of men and women in the group. This is unlikely given the short duration of voices within the presented ensemble (250 ms). Nonetheless, to address this concern, we ran an additional study in which participants estimated the sex ratio while simultaneously undergoing a cognitive load.

Study 1b

Study 1b replicated the accuracy block of Study 1a. However, participants were told that the study was designed to test the influence of auditory information on their memory for different shapes. Instead of listening solely to vocal ensembles, participants were tasked with memorizing the position of four different shapes on their computer screen while listening to a vocal ensemble. If extracting the sex ratio of a vocal ensemble is an efficient process, as we expect, then participants should be able to accurately estimate the sex ratio of a group even with the presence of a cognitive load.

Method

Participants

130 individuals were recruited through Prolific (<https://www.prolific.co/>) and participated in this 10-min study in exchange for \$1.50. Three individuals were excluded from analyses for providing identical judgments across all trials, and 14 individuals were excluded for reporting audio issues. This yielded a final sample size of 113 (Gender: 52% women, 47% men, 1% genderqueer; Race: 67% White, 18% Black, 6% Asian, 9% biracial/other; Age: $M = 40.8$ years, $SD = 15.0$, $\min = 19$, $\max = 79$).

Ensemble stimuli

The ensemble stimuli were the same stimuli used in Study 1a.

Shape stimuli

We selected four shapes to use in the cognitive load memory task: square, circle, triangle, and diamond. Each shape was outlined in black, had a transparent center, and measured 1 inch in height.

Procedure

After completing an informed consent form, participants verified that they were wearing headphones and in a quiet environment. Participants were instructed that the study was investigating the effect of audio information on their memory for shapes. The instructions specified that they would hear a group of voices while they were briefly presented with four shapes simultaneously in varying corners on the screen (top left, top right, bottom left, bottom right). Participants were told that it was important for them to try to remember the various positions of the shapes on the screen for a subsequent memory test.

On each trial, a randomly selected ensemble was played that varied in sex ratio ranging from all men to all women (0:12, 3:9, 6:6, 9:3, 12:0). While the ensemble played, the location of the four shapes was randomly shown on the screen (top left, top right, bottom left, bottom right). Immediately after hearing the vocal ensemble, participants were asked "where was the [circle/square/diamond/triangle] located on the screen?" followed by four multiple choice options: top left, top right, bottom left, bottom right. After they provided their answer, they were then asked to estimate the ratio of men to women in the ensemble using the same visual scale from Study 1a (Fig. 1). The study protocol consisted of 50 total trials. At the end of the study, participants provided demographic information and were debriefed.

Results

First, we tested whether participants were accurate in efficiently extracting the sex ratio of a group by regressing the estimated number of men in the group onto the actual sex ratio. As the ratio of men to women increased, so did participants' numeric estimates of men in an ensemble, $B = 0.75$, $SE = 0.03$, $t(138) = 24.48$, $p < 0.001$. The effect size was slightly smaller than Study 1a (0.75 versus 0.81, respectively) which was unsurprising given the addition of a cognitive load (Table 1). However, participants were still notably efficient and accurate at estimating the sex ratios of groups even when that was not their primary goal. This result suggests that extracting information

	Perceived sex ratio				Perceived facial gender			
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Study 1a	0.81	0.04	20.54	<0.001	0.88	0.03	33.20	<0.001
Study 1b	0.75	0.03	24.48	<0.001	–	–	–	–
Study 2	0.84	0.04	23.03	<0.001	0.85	0.03	32.50	<0.001
Study 3	0.80	0.04	19.93	<0.001	0.88	0.03	33.80	<0.001

Table 1. Comparison of statistical effects for Studies 1–3. Effects of actual sex ratio on perceived sex ratio and perceived facial gender.

about a group's sex ratio from their voices is a relatively effortless process, like extracting sex ratio information from visual ensembles^{25,26}.

Discussion

The results from Study 1b replicated the findings from Study 1a with a cognitive load. Even when participants' primary motivation was to memorize basic shapes, they were still attuned to the relative ratio of men and women within a group, suggesting that extracting group sex information can be accomplished efficiently and with little awareness. In fact, although the effect size was smaller in Study 1b with the cognitive load, it was not drastically different from Study 1a. Because of this, the cognitive load task was dropped from future studies.

Thus far, Studies 1a and 1b established that listeners rapidly and accurately represent the relative sex ratio of an ensemble. However, these studies did not test downstream consequences of accuracy on judgments of the group. Evidence suggests that perceivers draw evaluative inferences about a group based on its sex composition. For instance, male-dominated groups are judged as more threatening²⁵. Multiple factors account for this. First, men have historically held historic social, economic, and political power imbalances. According to social identity^{53–55} and intergroup threat^{56,57} theories, groups with more power are often seen as more threatening, especially by lower-power groups. Second, men are more likely than women to pose a threat, as reflected in violent crime statistics worldwide^{58,59}. Third, physically strong men are more militant⁶⁰ and masculine voices are a reliable indicator of men's physical strength (i.e., threat potential)⁶¹. Thus, it is evolutionarily adaptive to ascertain the number of men in a group quickly and accurately. However, accurate detection of a group's composition may inadvertently undermine the accuracy of threat detection when observers use these percepts as a proxy for threat. Specifically, perceptions of a group's sex composition might elevate or mitigate perceived threat when men are common or rare, respectively.

Another important social judgment that might be influenced by the sex ratio of vocal ensembles is a sense of fit or belonging. Humans are fundamentally motivated to achieve a sense of belonging⁶². In the visual literature, social belonging depends on the number of same-gender others in the group²⁶. That is, men feel more belonging in groups with more men, and women feel more belonging in groups of more women. Given that voices communicate information that perceivers are attuned to, and perceivers are motivated to determine whether they fit within a group, it is likely that perceptions of belonging are tethered to the number of similar others (i.e., same-gender) in the group.

To test the influence of auditory ensemble coding on evaluative judgments, the following studies assessed whether the sex composition of a group influenced perceptions of threat (Study 2) and belonging (Study 3).

Study 2

Study 2 aimed to replicate and extend the findings from Study 1 with a specific focus on understanding whether group sex composition influenced evaluative judgments of threat. Participants listened to ensembles and estimated the sex ratio of the group, the perceived average facial appearance of the group, and the perceived threat of the group. Given that threat judgments appear heavily tethered to the number of men within an ensemble²⁵, we predicted that as the actual ratio of men to women in a group increases, perceived threat will also increase.

Method

Participants

Ninety-two participants from Prolific completed this 30-min study in exchange for \$4.50. Twenty-one participants were excluded from analyses due to identical responses on every scale item. This yielded a final sample of 71 participants (Gender: 58% women, 42% men; Race: 70% White, 13% Black, 8% Asian, 8% biracial/other; Age: $M = 38.9$ years, $SD = 14.7$ years, $min = 20$, $max = 77$).

Stimuli and procedure

The ensemble stimuli were the same as in Study 1. Participants completed a consent form before verifying that they were wearing headphones in a quiet environment. Participants were exposed to 3 different counterbalanced blocks. Two of these blocks were identical to Study 1 and collected estimates of sex ratio and average face gender. The third block randomly presented 50 ensembles to participants and asked them how threatening they perceived the group to be on a scale from 1 (*not at all threatening*) to 7 (*extremely threatening*). Lastly, participants provided demographic information and were debriefed.

Results

As expected, we replicated the accuracy results found in Study 1. As the ratio of men to women increased, so did participants' numeric estimates of men in an ensemble, $B = 0.84$, $SE = 0.04$, $t(105) = 23.03$, $p < 0.001$. We also replicated the average group member appearance findings from Studies 1a and b. As the number of men in the group increased, participants perceived the average group member to be more masculine, $B = 0.85$, $SE = 0.03$, $t(86) = 32.50$, $p < 0.001$.

Importantly, we investigated whether the sex ratio of the group affected perceptions of threat. Regressing the overall threat judgment score onto the actual sex ratio, we found that as the ratio of men to women increased, threat judgments also increased, $B = 0.11$, $SE = 0.02$, $t(81) = 6.68$, $p < 0.001$. As expected, these results suggest that threat judgments are sensitive to group sex composition.

Discussion

The general findings from Study 1 were replicated in Study 2 (Table 1). Listeners were accurate at judging the sex ratio of a group from their voices alone, and groups with more men were estimated to appear more facially masculine on average. Importantly, the sex composition of a group had implications for perceived threat. As the number of men in the group increased, so did threat judgments.

Study 3

The goal of Study 3 was to investigate whether group sex composition influenced social belonging. Participants listened to ensembles and estimated the sex ratio of the group, the perceived average facial appearance of the group, and their perception of fit within the group. Because feelings of belonging in a group are related to the number of people who hold similar identities to you²⁶, we predicted that reported belonging would increase the number of same-gender others increased.

Method

Participants

84 participants from Prolific completed this 30-min study in exchange for \$4.50. Three participants were excluded from analyses due to identical responses on each response item. One participant was excluded because they did not report demographic information and we investigated participant gender as a potential moderator. This yielded a final sample of 80 participants (Gender: 59% women, 41% men; Race: 75% White, 8% Black, 6% Asian, 11% biracial/other; Age: $M = 38.9$ years, $SD = 12.8$ years, min = 18, max = 73).

Stimuli

The ensemble stimuli were created in the same way as Studies 1 and 2.

Procedure

Participants completed a consent form and verified that they were wearing headphones in a quiet environment. Like Study 2, they were randomly presented 3 counterbalanced blocks. Each block played 50 ensembles in randomized order, with 10 ensembles of each sex ratio. One block asked participants about the perceived sex ratio of the ensemble. Another block asked participants about the perceived average group member. The third block asked participants how much they felt they would fit or belong in the group on a scale from 1 (*not at all*) to 7 (*extremely*). Lastly, participants provided demographic information and were debriefed.

Results

We replicated the general accuracy results found in Studies 1 and 2. As the ratio of men to women in a group increased, so did participants' numeric estimates of men in an ensemble, $B = 0.80$, $SE = 0.04$, $t(111) = 19.93$, $p < 0.001$. Furthermore, as the ratio of men to women increased, participants perceived the average group member to be more masculine, $B = 0.88$, $SE = 0.03$, $t(95) = 33.80$, $p < 0.001$.

Next, we regressed perceived fit and belonging onto actual sex ratio. Perceived fit and belonging decreased as the group's ratio of men to women increased, $B = -0.08$, $SE = 0.03$, $t(87) = -2.68$, $p = 0.009$. Importantly, this trend was different depending on participant gender, $B = 0.40$, $SE = 0.04$, $t(77) = 10.02$, $p < 0.001$. Tests of simple slopes revealed that men felt that they belonged more when the ratio of men to women increased, $B = 0.15$, $SE = 0.03$, $t(85) = 4.90$, $p < 0.001$, whereas women felt that they fit less, $B = -0.25$, $SE = 0.03$, $t(88) = -9.26$, $p < 0.001$. We also tested how men and women perceivers' feelings of fit and belonging varied when the actual number of men in the group was centered at three (a minority), six (equal representation), and nine (a majority). When groups consisted of three men and nine women, women perceivers ($M = 4.91$, $SD = 1.32$) reported significantly higher feelings of fit and belonging than men perceivers ($M = 4.16$, $SD = 1.29$), $B = -0.74$, $SE = 0.19$, $t(77) = -3.83$, $p < 0.001$. In groups with equal numbers of men and women, there was no significant between women perceivers ($M = 4.40$, $SD = 1.45$) and men perceivers ($M = 4.57$, $SD = 1.16$), $B = 0.17$, $SE = 0.21$, $t(78) = 0.83$, $p = 0.409$. In groups with nine men and three women, women perceivers ($M = 3.84$, $SD = 1.55$) felt significantly less belonging than men ($M = 4.76$, $SD = 1.26$), $B = 2.67$, $SE = 0.32$, $t(78) = 8.43$, $p < 0.001$ (Fig. 3).

Discussion

The findings from Study 3 supported our hypothesis that as the number of men in a group increased, feelings of fit and belonging decreased. However, as we predicted, this trend differed significantly by participant gender. Men and women reported more belonging in groups with more same-gender members. The accuracy findings from the previous studies were replicated (Table 1).

General discussion

Based on previous research on visual ensembles, we expected that listeners would be able to accurately judge the sex composition of auditory ensembles and that these categorizations would affect social judgments of groups. In particular, we expected that increased ratios of men in a group would be associated with perceptions of threat, and that feelings of belonging would be linked to perceivers' gender, such that men would feel more belonging in groups with more men, and women would feel more belonging in groups with more women. All of these predictions were confirmed across our four studies. Listeners accurately judged the sex composition of a group from voices alone and used perceptual features to form mental representations of the average group member that embodied gendered features. Effect sizes for accuracy across studies are reported in Table 1 for comparison. Further, we found that the sex composition of a group had downstream consequences for evaluative inferences.

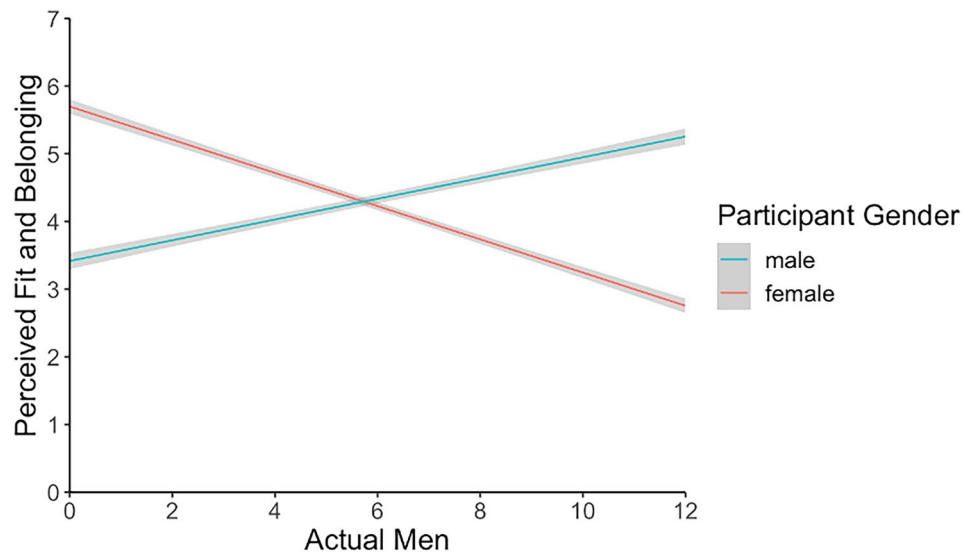


Figure 3. Study 3 perceived fit and belonging by actual sex ratio and participant gender. Note: Error bars represent 95% confidence intervals.

As the number of men in the group increased, so did perceived threat (Study 2). This suggests that the ratio of men to women in a group affects judgments in a manner that is consistent with social stereotypes that characterize men as threatening. Meanwhile, perceived social belonging increased with the number of same-gender others in the group (Study 3).

These findings expand knowledge in multiple domains including ensemble coding, voice perception, interpersonal communication, and social cognition. Ensemble coding is largely focused on narrow aspects of visual perception but the mechanisms underlying ensemble coding extend to social stimuli as well. Our research complements and extends these findings to show that ensemble coding operates similarly in the auditory modality. Furthermore, we extend findings by Neuhoff and colleagues^{40,41} by demonstrating downstream consequences of extracting sex information from a vocal ensemble. Social groups generate a variety of auditory signals and cues that reveal aspects of their composition, intent, and experience⁶³. We should expect a fine-tuned ability to extract sex composition from both the visual and auditory modalities in the service of important social judgments such as how much a group might either constitute a threat or possible alliance. These are the first studies to test the downstream consequences of rapidly extracting summary statistical information from vocal ensembles. Here, we show that people can make social judgments of threat and alliance rather effortlessly, relying on accurate information gleaned from just brief exposures to auditory information alone. This suggests that social inferences of voices are shaped by a combination of low-level (e.g., pitch) and high-level (e.g., stereotype) information.

Furthermore, these findings have important implications for understanding why groups with historically unbalanced gender representation have resisted change. Our studies suggest that regardless of the specific context, women exposed to auditory cues indicating male-dominated groups may experience a diminished sense of fit or belonging. This judgment might subsequently influence one's approach or avoidance behavior toward the group. Groups that are historically dominated by men (e.g., STEM fields) might deter women from applying due to lack of perceived fit and increased threat. This phenomenon can also occur for men, who might feel decreased fit despite perceived threat being relatively low in fields that are historically dominated by women such as nursing, education, or psychology^{64,65}. Increasing gender representation across fields may positively influence equal participation of men and women.

Future directions

The experimental methods used in this paper allow us to draw strong conclusions about the causal mechanisms underlying threat and belonging judgments by using artificially created ensembles. Future research should examine if these findings hold in real-world vocal ensembles or in ensembles with overlapping or disorganized voices that simulate various group vocalization scenarios. For example, investigating how other social identities conveyed through the voice (e.g., accent, age) influence judgments of threat and belonging would be informative. To enhance ecological validity, future studies could also manipulate semantic content reflective of various environments (e.g., STEM) and collect behavioral measures along with self-reported measures of threat and belonging.

The findings reported here provide a strong foundation for testing the relative contributions of low- and high-level information on evaluative judgments that parallel existing research in the visual literature⁶⁶. That is, research could systematically test whether judgments are affected more strongly by low-level perceptual features of the voice (e.g., pitch and timbre) or by the high-level stereotypes held toward the categories of men and women by manipulating the relative pitch within men and women's voices to be typical (high-pitched women, low-pitched men) or atypical (low-pitched women, high-pitched men). If judgments are affected only by category information communicated through the voice, then perceived threat and fit should only vary as a function of the sex

ratio of the group. However, if judgments are also sensitive to within-sex perceptual variations in the voice, then perceived threat should be exaggerated in groups with more masculinized voices and attenuated in groups with more feminized voices, regardless of actual sex.

It is important to note that our studies build on binary assumptions of sex and gender, and thus future research should test (i) if perceivers discern group composition beyond this binary framework, and (ii) how nonbinary or genderqueer individuals evaluate groups with varying sex ratios. Social cognitive research demonstrates that people automatically categorize others into binary sex and gender groups with minimal control, and form impressions based on activated category stereotypes^{67,68}. However, as societal recognition of nonbinary and genderqueer identities grows, mental representations may expand beyond the binary. This could lead to more nuanced numerosity judgments of group composition beyond males and females or men and women. Additionally, while our findings demonstrate that men and women feel a stronger sense of belonging in groups with similar others, we could not assess this for nonbinary or genderqueer participants due to insufficient power. Future research should explore this area to enhance our understanding of the interplay of identity and group perception.

In most communication environments, perceivers receive both visual and auditory information simultaneously. However, very little work has investigated how information from multiple sensory domains combine to influence social categorizations and judgments. Although some work exists looking at multimodal perception of individuals^{69–74}, extending this research to group settings would contribute to more ecologically valid theories of ensemble perception.

Conclusion

We expected perceptual sensitivity to auditory information in assessing groups given that visual information is often not available in intergroup interactions. In four studies, we demonstrated that individuals are highly attuned to the sex ratios of groups after briefly hearing their voices, and that information about group composition can be extracted efficiently and implicitly. Furthermore, the composition of groups influenced perceptions of group threat and belonging such that threat increased concomitantly with the number of men in the group and belonging was tethered to the number of same-gender others in the group.

Data availability

The data, code, and materials that support the findings of this paper are openly available on OSF (<https://osf.io/28gvd/>).

Received: 3 January 2024; Accepted: 20 June 2024

Published online: 25 June 2024

References

- Alt, N. P. & Phillips, L. T. Person perception, meet people perception: Exploring the social vision of groups. *Perspect. Psychol. Sci.* **17**(3), 768–787. <https://doi.org/10.1177/17456916211017858> (2022).
- Phillips, L. T., Weisbuch, M. & Ambady, N. People perception: Social vision of groups and consequences for organizing and interacting. *Res. Organ. Behav.* **34**, 101–127. <https://doi.org/10.1016/j.riob.2014.10.001> (2014).
- U.S. Bureau of Labor Statistics (2019). American time use survey (ATUS).
- Whitney, D. & Yamanashi Leib, A. Ensemble perception. *Annu. Rev. Psychol.* **69**, 105–129. <https://doi.org/10.1146/annurev-psych-010416-044232> (2018).
- Demeyere, N., Rzeskiewicz, A., Humphreys, K. A. & Humphreys, G. W. Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia* **46**(11), 2861–2864. <https://doi.org/10.1016/j.neuropsychologia.2008.05.014> (2008).
- Webster, J., Kay, P. & Webster, M. A. Perceiving the average hue of color arrays. *J. Opt. Soc. Am. A* **31**(4), A283–A292. <https://doi.org/10.1364/josaa.31.00a283> (2014).
- Bauer, B. Does Steven's power law for brightness extend to perceptual brightness averaging?. *Psychol. Rec.* **59**, 171–186. <https://doi.org/10.1007/BF03395657> (2009).
- Dakin, S. C. & Watt, R. J. The computation of orientation statistics from visual texture. *Vis. Res.* **37**(22), 3181–3192. [https://doi.org/10.1016/S0042-6989\(97\)00133-8](https://doi.org/10.1016/S0042-6989(97)00133-8) (1997).
- Miller, A. L. & Sheldon, R. Magnitude estimation of average length and average inclination. *J. Exp. Psychol.* **81**(1), 16–21. <https://doi.org/10.1037/h0027430> (1969).
- Parke, L., Lund, J., Angelucci, A., Solomon, J. A. & Morgan, M. Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* **4**(7), 739–744. <https://doi.org/10.1038/89532> (2001).
- Alvarez, G. A. & Oliva, A. Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *PNAS* **5**(18), 7345–7350. <https://doi.org/10.1073/pnas.0808981106> (2009).
- Melcher, D. & Kowler, E. Shapes, surfaces and saccades. *Vis. Res.* **39**, 2929–2946. [https://doi.org/10.1016/S0042-6989\(99\)00029-2](https://doi.org/10.1016/S0042-6989(99)00029-2) (1999).
- Vishwanath, D. & Kowler, E. Localization of shapes: Eye movements and perception compared. *Vis. Res.* **43**(15), 1637–1653. [https://doi.org/10.1016/S0042-6989\(03\)00168-8](https://doi.org/10.1016/S0042-6989(03)00168-8) (2003).
- Watamaniuk, S. J., Sekuler, R. & Williams, D. W. Direction perception in complex dynamic displays: The integration of direction information. *Vis. Res.* **29**(1), 47–59. [https://doi.org/10.1016/0042-6989\(89\)90173-9](https://doi.org/10.1016/0042-6989(89)90173-9) (1989).
- Watamaniuk, S. J. & McKee, S. Simultaneous encoding of direction at a local and global scale. *Percept. Psychophys.* **60**(2), 191–200. <https://doi.org/10.3758/bf03206028> (1998).
- Ariely, D. Seeing sets: Representation by statistical properties. *Psychol. Sci.* **12**(2), 157–162. <https://doi.org/10.1111/1467-9280.00327> (2001).
- Chong, S. C. & Treisman, A. Representation of statistical properties. *Vis. Res.* **43**(4), 393–404. [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5) (2003).
- Haberman, J. & Whitney, D. Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* <https://doi.org/10.1016/j.cub.2007.06.039> (2007).
- Bai, Y., Leib, A. Y., Puri, A. M., Whitney, D. & Peng, K. Gender differences in crowd perception. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2015.01300> (2015).
- De Fockert, J. & Wolfenstein, C. Rapid extraction of mean identity from sets of faces. *Q. J. Exp. Psychol.* **62**(9), 1716–1722. <https://doi.org/10.1080/17470210902811249> (2009).

21. Leib, A. Y. *et al.* Crowd perception in prosopagnosia. *Neuropsychologia* **50**(7), 1698–1707. <https://doi.org/10.1016/j.neuropsychologia.2012.03.026> (2012).
22. Neumann, M. F., Schweinberger, S. R. & Burton, A. M. Viewers extract mean and individual identity from sets of famous faces. *Cognition* **128**(1), 56–63. <https://doi.org/10.1016/j.cognition.2013.03.006> (2013).
23. Sweeny, T. D., Whitney, D. & Haroz, S. Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *J. Exp. Psychol.* **39**(2), 329–337. <https://doi.org/10.1037/a0028712> (2013).
24. Sweeny, T. D. & Whitney, D. Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychol. Sci.* **25**(10), 1903–1913. <https://doi.org/10.1177/0956797614544510> (2014).
25. Alt, N. P., Goodale, B., Lick, D. J. & Johnson, K. L. Threat in the company of men: Ensemble perception and threat evaluations of groups varying in sex ratio. *Soc. Psychol. Pers. Sci.* **10**(2), 152–159. <https://doi.org/10.1177/1948550617731498> (2019).
26. Goodale, B. M., Alt, N. P., Lick, D. J. & Johnson, K. L. Groups at a glance: Perceivers infer social belonging in a group based on perceptual summaries of sex ratio. *J. Exp. Psychol. Gen.* **147**(11), 1660–1676. <https://doi.org/10.1037/xge0000450> (2018).
27. Haberman, J. & Whitney, D. Seeing the mean: Ensemble coding for sets of faces. *J. Exp. Psychol. Hum. Percept. Perform.* **35**(3), 718–734. <https://doi.org/10.1037/a0013899> (2009).
28. Thornton, I. M., Srismith, D., Oxner, M., & Hayward, W. G. (2014). Estimating the racial composition of groups of faces: An ensemble other-race effect. In *TSPSC2014* 76–78.
29. Cheryan, S., Meltzoff, A. N. & Kim, S. Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Comput. Educ.* **57**(2), 1825–1835. <https://doi.org/10.1016/j.compedu.2011.02.004> (2011).
30. Emerson, K. T. U. & Murphy, M. C. Identity threat at work: How social identity threat and situational cues contribute to racial and ethnic disparities in the workplace. *Cult. Divers. Ethnic Minor. Psychol.* **20**(4), 508–520. <https://doi.org/10.1037/a0035403> (2014).
31. Murphy, M. C., Steele, C. M. & Gross, J. J. Signaling threat. *Psychol. Sci.* **18**(10), 879–885. <https://doi.org/10.1111/j.1467-9280.2007.01995.x> (2007).
32. Cheryan, S., Plaut, V. C., Davies, P. G. & Steele, C. M. Ambient belonging: How stereotypical cues impact gender participation in computer science. *J. Pers. Soc. Psychol.* **97**(6), 1045–1060. <https://doi.org/10.1037/a0016239> (2009).
33. Good, C., Rattan, A. & Dweck, C. S. Why do women opt out? Sense of belonging and women's representation in mathematics. *J. Pers. Soc. Psychol.* **102**(4), 700–717. <https://doi.org/10.1037/a0026659> (2012).
34. Stout, J. G. & Dasgupta, N. When *he* doesn't mean *you*: Gender-exclusive language as ostracism. *Pers. Soc. Psychol. Bull.* **37**(6), 757–769. <https://doi.org/10.1177/0146167211406434> (2011).
35. Pew Research Center. How the Coronavirus outbreak has—and hasn't—changed the way Americans work. <https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirus-outbreak-has-and-hasnt-changed-the-way-americans-work/> (2020).
36. Pew Research Center. Covid-19 pandemic continues to reshape work in America. <https://www.pewresearch.org/social-trends/2022/02/16/covid-19-pandemic-continues-to-reshape-work-in-america/> (2022).
37. Fosslien, L. & Duffy, M. W. How to combat zoom fatigue. *Harvard Business Review*, 1–6. <https://hbr.org/2020/04/how-to-combat-zoom-fatigue> (2020).
38. Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A. & Whitney, D. Humans use summary statistics to perceive auditory sequences. *Psychol. Sci.* **24**(8), 1389–1397. <https://doi.org/10.1177/0956797612473759> (2013).
39. McDermott, J. H., Schemitsch, M. & Simoncelli, E. P. Summary statistics in auditory perception. *Nat. Neurosci.* **16**(4), 493–498. <https://doi.org/10.1038/nn.3347> (2013).
40. Neuhoff, J. G. The perception of operational sex ratios by voice. *Nature* <https://doi.org/10.1038/s41598-017-18182-4> (2017).
41. Neuhoff, J. & G., & Sikich, T. N., Hearing sex at the cocktail party: Biased sex ratios influence vocal attractiveness. *Audit. Percept. Cogn.* **1**(1–2), 97–111. <https://doi.org/10.1080/25742442.2018.1518949> (2018).
42. Piazza, E., Sweeny, T., Wessel, D. & Whitney, D. Ensemble coding in audition. *i-Perception* **2**(8), 889–889. <https://doi.org/10.1068/ic889> (2011).
43. Ongchoco, J., Knowlton, T. & Papafragou, A. Language shifts the representation of sounds in time: From auditory individuals to auditory ensembles. *Proceedings of the Annual Meeting of the Cognitive Science Society* **45**, 1572–1578 (2023).
44. Judd, C. M., Westfall, J. & Kenny, D. A. Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annu. Rev. Psychol.* **68**, 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702> (2017).
45. Arend, M. G. & Schäfer, T. Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychol. Methods* <https://doi.org/10.1037/met0000195.supp> (2019).
46. Lora, J. Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assess. Educ.* **6**(1), 1–11. <https://doi.org/10.1186/s40536-018-0061-2> (2018).
47. Aguinis, H., Gottfredson, R. K. & Culpepper, S. A. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *J. Manag.* **39**(6), 1490–1528. <https://doi.org/10.1177/0149206313478188> (2013).
48. Snijders, T. A. B. Power and sample size in multilevel linear models. In *Encyclopedia of Statistics in Behavioral Science* Vol. 3 (eds Everitt, B. S. & Howell, D. C.) 15701573 (Wiley, 2005). <https://doi.org/10.1002/0470013192.bsa492>.
49. Westfall, J., Kenny, D. A. & Judd, C. M. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.* **143**(5), 2020–2045. <https://doi.org/10.1037/xge0000014> (2014).
50. Bryant, G. A. & Haselton, M. G. Vocal cues of ovulation in human females. *Biol. Lett.* **5**(1), 12–15. <https://doi.org/10.1098/rsbl.2008.0507> (2009).
51. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**(1), 1–48. <https://doi.org/10.18637/jss.v067.i01> (2015).
52. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in linear mixed effects models. *J. Stat. Softw.* **82**(13), 1–26. <https://doi.org/10.18637/jss.v082.i13> (2017).
53. Tajfel, H. & Turner, J. C. An integrative theory of intergroup conflict. In *The Social Psychology of Intergroup Relations* (eds Austin, W. G. & Worchel, S.) 33–47 (Brooks/Cole, 1979).
54. Crocker, J. & Major, B. Social stigma and self-esteem: The self-protective properties of stigma. *Psychol. Rev.* **96**(4), 608–630. <https://doi.org/10.1037/0033-295X.96.4.608> (1989).
55. Branscombe, N. R., Schmitt, M. T. & Harvey, R. D. Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *J. Pers. Soc. Psychol.* **77**(1), 135–149. <https://doi.org/10.1037/0022-3514.77.1.135> (1999).
56. Stephan, W. G. & Stephan, C. W. An integrated threat theory of prejudice. In *Reducing Prejudice and Discrimination* (ed. Oskamp, S.) 23–45 (Lawrence Erlbaum Associates, 2000).
57. Stephan, W. G., Ybarra, O. & Morrison, K. R. Intergroup threat theory. In *Handbook of Prejudice, Stereotyping, and Discrimination* (ed. Nelson, T. D.) 43–59 (Psychology Press, 2009).
58. Collier, R. & Walgrave, L. Masculinities and crime. *Crim. Justice Matters* **34**(1), 21–24. <https://doi.org/10.1080/09627259808552829> (1998).
59. U.S. Department of Justice. Crime in the United States, 2021. <https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/crime-trend> (2021).
60. Sell, A. *et al.* Physically strong men are more militant: A test across four countries. *Evol. Hum. Behav.* **38**(3), 334–340. <https://doi.org/10.1016/j.evolhumbehav.2016.11.002> (2017).
61. Puts, D. A., Apicella, C. L. & Cárdenas, R. A. Masculine voices signal men's threat potential in forager and industrial societies. *Proc. R. Soc. B Biol. Sci.* **279**(1728), 601–609. <https://doi.org/10.1098/rspb.2011.0829> (2011).

62. Baumeister, R. F. & Leary, M. R. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* **117**(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497> (1995).
63. Bryant, G. A. & Bainbridge, C. M. Signals and cues of social groups. *Behav. Brain Sci.* **45**, 1–64. <https://doi.org/10.1017/S0140525X21001461> (2022).
64. Forsman, J. A. & Barth, J. M. The effect of occupational gender stereotypes on men's interest in female-dominated occupations. *Sex Roles* **76**, 460–472. <https://doi.org/10.1007/s11199-016-0673-3> (2017).
65. Shen-Miller, D. & Smiler, A. P. Men in female-dominated vocations: A rationale for academic study and introduction to the special issue. *Sex Roles* **72**(7–8), 269–276. <https://doi.org/10.1007/s11199-015-0471-3> (2015).
66. Freeman, J. B., Ambady, N., Rule, N. O. & Johnson, K. L. Will a category cue attract you? Motor output reveals dynamic competition across person construal. *J. Exp. Psychol. Gen.* **137**(4), 673–690. <https://doi.org/10.1037/a0013875> (2008).
67. Brewer, M. B. A dual process model of impression formation. In *Advances in Social Cognition* Vol. 1 (eds Srull, T. K. & Wyer, R. S.) 1–36 (Lawrence Erlbaum Associates, 1988).
68. Fiske, S. T. & Neuberg, S. L. A continuum model of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Adv. Exp. Soc. Psychol.* **23**, 1–74. [https://doi.org/10.1016/S0065-2601\(08\)60317-2](https://doi.org/10.1016/S0065-2601(08)60317-2) (1990).
69. Campanella, S. & Belin, P. Integrating face and voice in person perception. *Trends Cogn. Sci.* **11**(12), 535–543. <https://doi.org/10.1016/j.tics.2007.10.001> (2007).
70. Davies-Thompson, J. *et al.* Hierarchical brain network for face and voice integration of emotion expression. *Cereb. Cortex* **29**(9), 3590–3605. <https://doi.org/10.1093/cercor/bhy240> (2019).
71. de Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M. & Vroomen, J. The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neurosci. Lett.* **260**(2), 133–136. [https://doi.org/10.1016/S0304-3940\(98\)00963-X](https://doi.org/10.1016/S0304-3940(98)00963-X) (1999).
72. de Gelder, B. & Vroomen, J. The perception of emotions by ear and by eye. *Cogn. Emot.* **14**(3), 289–311. <https://doi.org/10.1080/026999300378824> (2000).
73. Freeman, J. B. & Ambady, N. When two become one: Temporally dynamic integration of the face and voice. *J. Exp. Soc. Psychol.* **47**(1), 259–263. <https://doi.org/10.1016/j.jesp.2010.08.018> (2011).
74. Stevenage, S. V. & Neil, G. J. Hearing faces and seeing voices: The integration and interaction of face and voice processing. *Psychol. Belg.* **54**(3), 266–281. <https://doi.org/10.5334/pb.ar> (2014).

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: KN, KJ, and BG; data collection: KN; data analysis and interpretation of results: KN, KJ, and GB; draft manuscript preparation: KN, KJ, and GB. All authors reviewed the results and approved the final version of the manuscript.

Funding

This research was supported by a National Defense Science and Engineering Graduate (NDSEG) Fellowship (Neuenschwander) and a grant from the National Science Foundation (BCS2017245) (Johnson).

Competing interests

KN's work has been funded by a National Defense Science and Engineering Graduate (NDSEG) Fellowship. KJ's work has been funded by the National Science Foundation (NSF). BG and GB declare no potential conflict of interest.

Additional information

Correspondence and requests for materials should be addressed to K.L.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024