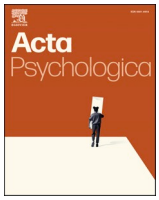
Acta Psychologica 260 (2025) 105664



Contents lists available at ScienceDirect

Acta Psychologica

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





When seeing follows believing: The role of social likelihood in demographic estimation

Yoonseo Heo [a], Sang Chul Chong [a,b,*]

[a] Graduate Program in Cognitive Science, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul, 03722, Republic of Korea
[b] Department of Psychology, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul, 03722, Republic of Korea

ARTICLE INFO

Keywords:
Demographic estimation
Minority overestimation
Majority underestimation
Rational strategy
Intersectionality

ABSTRACT

This study investigates how individuals estimate the size of demographic groups and whether these estimations are guided by contextual expectations rooted in social associations. Across three experiments (N = 180), we explore the hypothesis that people employ a rational compensation strategy when estimating group prevalence, adjusting their judgments toward likely values depending on contextual cues. In Experiment 1, participants estimated the proportions of gender and political affiliation from crowds of South Korean politicians. The results replicated the overestimation of minorities and underestimation of majorities but also demonstrated that estimation errors were modulated by contextual alignment—that is, participants' estimates were more refined when group characteristics aligned with stereotypical associations (e.g., women with progressive politics). Experiment 2 extended these findings by examining categorical judgments of majority status, wherein participants demonstrated greater sensitivity to proportion when gender and political affiliation conformed to expected associations. Experiment 3 confirmed that such likelihood-based refinements stem from semantic associations: participants estimated differing gender ratios in hypothetical political groupings depending on which party was described as being the majority. In Experiments 1 and 2, individual differences in gender–party stereotype endorsement did not significantly correlate with estimation errors. Together, the findings illustrate that group size estimation is not purely perceptual but rather influenced by a context-sensitive rational strategy that incorporates associative knowledge. Thus, this work provides a nuanced account of demographic misperception and offers implications for understanding how people form impressions of diversity in political and social contexts.

Faces are complex sources of information, from which a multitude of information can be effortlessly extracted in split seconds (Young & Bruce, 2013). The range of information includes gender, race, age, emotion, dominance, and trustworthiness (Dobs et al., 2019). Further, these extracted pieces of information interact with each other. For example, people more quickly and accurately recognize angry expressions in male faces, while happy expressions are more quickly recognized in female faces (Becker et al., 2007). Neutral male faces are often miscategorized as angry (Becker et al., 2007), possibly because neutral expressions in male faces objectively resemble angry expressions more than those of female faces (Zebrowitz et al., 2010). Thus, in a way, the perception of gender *guides* the perception of emotion such that observers arrive at the most likely, if occasionally biased, perceptual judgment given associations between these dimensions.

Research on people perception also indicates that observers are adept at extracting information from groups of faces. For example, they can accurately report the gender and racial diversity (white and black) of a group of faces (Phillips et al., 2018). They can also estimate the mean (Haberman & Whitney, 2007, 2009; Sun & Chong, 2020) and variance (Haberman et al., 2015) of facial expressions, as well as the mean attractiveness (Luo & Zhou, 2018) and variance in group dominance (Phillips et al., 2018). As is the case for the perception of individual faces, the perception of a group of faces is also influenced by other dimensions. For example, while people are adept at judging the diversity of a crowd (Phillips et al., 2018), these judgments are influenced by the variance of another dimension; people tend to perceive more gender diversity in groups that are more racially diverse, and vice versa (Daniels et al., 2017; Mijalli et al., 2023).

Another aspect of people perception involves estimating demographic composition, for example, tracking the proportions of different groups. While observers can accurately track the ratio of sex/gender (Alt et al., 2019, 2024; Goodale et al., 2018) and enumerate the

* Corresponding author at: Department of Psychology, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul, 03722, Republic of Korea.
E-mail address: scchong@yonsei.ac.kr (S.C. Chong).

https://doi.org/10.1016/j.actpsy.2025.105664
Received 19 May 2025; Received in revised form 7 September 2025; Accepted 25 September 2025
Available online 29 September 2025
0001-6918/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

number of black and white faces (Yang & Dunham, 2019), they also exhibit systemic biases in their judgments; notably, they tend to overestimate the prevalence of numerical minorities and underestimate that of numerical majorities (Kardosh et al., 2022; Khaw et al., 2021). However, these biases can be largely attributed to a rational strategy whereby people who are not perfectly informed about demographic information adjust their estimates toward a likely value (Landy et al., 2018). Because, as discussed above, the group properties that people perceive interact with one another, it is possible that proportion judgments—another property that can be effortlessly extracted—are similarly influenced by other information about proportions.

Furthermore, we know that not all information conveys equally strong signals and that some kinds of information may more effectively guide perception than other kinds. As a rational strategy to reduce error, information that is more relevant to the task can more effectively direct perception toward the more likely outcome. One such moderator may be the strength of associations between different dimensions. An important insight from the intersectional perception of emotion of gender, whereby the perception of emotion is biased by gender (as discussed above), is that directionality plays an important role: displays of anger are more easily and accurately identified in male faces, while displays of happiness are more easily identified in female faces (Alt et al., 2024). This means that anger provides a stronger perceptual signal for male faces while happiness provides a weaker signal, and vice versa for female faces. Thus, contexts that align with these statistical regularities may better guide perceptions and refine judgments. In demographic estimation, such regularities may similarly modulate signal strength across dimensions. When numerical proportions align with likely associations, judgments may be refined toward the likely estimate; conversely, when dimensions are misaligned with these associations, the signal weakens, rendering less precise judgments through regression toward the mean. Importantly, though this refinement need not reflect objective realities, just as neutral male faces may be misidentified as angry (Becker et al., 2007), such judgments may align with what is expected from regularities rather than veridical accuracy.

We, thus, arrive at the prediction that, in demographic judgments, participants can extract the proportions of multiple dimensions, and the proportion of a nontarget dimension will refine judgments toward the likely estimate, especially when it aligns with the associative expectations. We directly test the corollaries of this prediction in three experiments. In cases where the proportion of the nontarget dimension aligns with these associations, the magnitude of the typical estimation errors often observed in demographic judgments—i.e., overestimation of numerical minorities and underestimation of numerical majorities—is reduced (Experiment 1). Specifically, when the proportion of the nontarget dimension aligns with expectations (e.g., estimating the proportion of females in the minority People's Power Party), it should act as a stronger guiding signal, leading individuals to refine their judgments to enumerate the target dimension toward the actual value rather than regressing toward the mean. Similarly, when the target group to be estimated can also be categorized as another group that is aligned with the direction of association (e.g., when judging whether females comprise the majority in the Democratic Party), participant responses are more refined and clustered around the actual value (Experiment 2). Finally, the refinement need not necessarily reflect objective realities; rather, it may lean toward values that align with commonly held associations (e.g., that female politicians are affiliated with the Democratic Party) or "likely" demographic expectations (Experiment 3).

We chose gender/sex (men/male and women/female) and political party (Democratic Party, which is a South Korean progressive party, and the People's Power Party, which is a South Korean conservative party) as our study dimensions. We selected these dimensions because partisanship has strong gender associations: in South Korea, the more a political party is perceived to be conservative, the more it is perceived as masculine, and the more a political party is perceived to be progressive, the more it is perceived as feminine (Ji, 2018). Further, as these two dimensions have no overlapping visual features, examining these dimensions enabled us to test the effect of alignment of nontarget dimensions without visual confounds that could lead to miscategorization.

In Experiments 1 and 2, we adapted the experimental paradigm used by Kardosh et al. (2022) and assessed and compared error trends in estimating and comparing the size of different demographic groups. In both experiments, people were briefly presented with images of Korean politicians cropped to their busts and displaying party affiliation through their clothing colour. In Experiment 1, participants estimated the percentage of a given subgroup (men, women, Democratic Party politicians, People's Power Party politicians, people wearing glasses, and people wearing masks) to the best of their ability. The minority/majority status of each group was manipulated to examine whether estimation error trends differed based on the alignment of nontarget proportions with the likely associations. In Experiment 2, participants judged whether the presented image contained a greater proportion of male/female politicians than a reference mean (50 %). This binary design allows a relatively bias-free measurement of estimation error patterns across different proportions. Here, the images represented either a likely association (e.g., female Democratic Party and male People's Power Party politicians) or an unlikely one (e.g., male Democratic Party and female People's Power Party politicians). In Experiment 3, participants inferred the percentage of men/women in a hypothetical random 1000-person sample, where the majority of the sample was affiliated with either the Democratic Party or the People's Power Party.

1. Experiment 1

Experiment 1 examined whether errors in demographic estimation reflect differences in the perceived likelihood of the context in which the estimation occurs. One explanation for the systematic overestimation of numerical minorities and underestimation of majorities is that individuals adopt a rational compensation strategy to account for uncertainty: that is, they shift their estimates toward a more "likely" value, typically around the center of the range (Landy et al., 2018). From another perspective, it is expected that estimators aiming to minimize their error adjust the value they hedge toward depending on the "likeliness" of the estimate in different contexts. Consider a scenario where one estimates the proportion of women among Korean politicians. Given that female representation in the Korean National Assembly is quite low—61 out of 300 as of January 2025 (IPU Parline, 2025)—a rational estimator would consider it unlikely that a given sample contains many women. However, this likelihood may change with additional context. If the politicians in the sample are mostly affiliated with the Democratic Party, the presence of women becomes *less* unlikely (albeit slightly), as nearly half of female politicians are affiliated with this party (29/61), compared to only one out of three in the People's Power Party (21/61) (National Assembly of the Republic of Korea, 2024). A rational estimator should, therefore, factor in this context when making demographic judgments.

Given these points, we adapt the paradigm by Kardosh et al. (2022), in which participants estimated the prevalence of a single target group while its proportion was systematically manipulated to be either a minority or the majority. They found consistent overestimation of numerical minorities and underestimation of numerical majorities. Experiment 1 adapted this paradigm by independently and simultaneously manipulating the proportions of two dimensions. Crucially, the proportion of one dimension served as a context for estimating the percentage of another.

Throughout this article, the group being estimated (e.g., women) is referred to as the *target group*, the dimension being estimated (e.g., gender) is referred to as the *target dimension*, the group that is not being estimated (e.g., Democratic Party supporters) is referred to as the *nontarget group*, and the dimension of that group (e.g., political party) is the *nontarget dimension*. Importantly, these groups do not necessarily overlap. For example, female politicians and Democratic Party

politicians might both be in the minority in a given display, but they are distinct dimensions: some Democratic Party politicians may not be women, and some female politicians may not be members of the Democratic Party. Participants reported explicit numerical estimates (e.g., "30 %"), which allowed the precise quantification of errors across conditions. In this way, we examined the patterns of estimation errors to determine whether participants used a rational compensation strategy to reduce error in demographic perception tasks.

We hypothesized that participants would engage in a rational estimation strategy to minimize their error, which would paradoxically lead to systematic error during a demographic estimation task. Our key predictions were twofold. First, we predicted that participants would overestimate minorities and underestimate majorities in general, regardless of the specific target group, consistent with previous research (Kardosh et al., 2022). When briefly presented with a visual display of politicians, it is expected that rational estimators shift their estimates toward the center of the range (Landy et al., 2018). This would result in the systematic overestimation of the minority and the underestimation of the majority as a compensation for the uncertainty that pulls responses away from extreme values (i.e., that are less likely). Second, we predicted that the degree of this adjustment would vary depending on the likelihood of the given context. For example, we expected that people compensate more strongly in contexts where the presence of the target group is less likely. A rational estimator is expected to place less weight on their estimates when their estimates are "unlikely" or are made under an "uncertain" context. This compensation for uncertainty will entail a greater shift away from the veridical proportion value. For example, estimating the percentage of female politicians in a crowd that is predominantly People's Power Party-affiliated should elicit greater compensation compared to estimating female politicians in a crowd that is predominantly Democratic Party-affiliated, as the former is a less plausible scenario.

2. Methods

All experiments, measures, manipulations, and data/participant exclusions are reported in the manuscript.

2.1.1. Participants

We recruited 105 participants aged 19 years or older who were native Korean speakers and reported either normal or corrected-to-normal vision. For their participation, participants were offered either course credit or monetary compensation of 6000 Korean Won. To determine the sample size, we accounted for publication bias and uncertainty following Anderson et al. (2017). We aimed for an effect size of 0.85, consistent with Experiment 5 of Kardosh et al. (2022), which recruited 100 participants to examine the moderating role of social knowledge in minority salience. In their Experiment 5, black minorities were significantly more overestimated than white minorities. Using the same effect size and a level of assurance of 0.80 to adjust for bias and uncertainty, we determined that 45 participants were needed per group. As this study seeks to compare estimation errors depending on the likelihood of the contexts in which estimations were made, we established that 90 participants total were required.

Out of the 105 initial participants, 15 were excluded from the analysis due to a technical error in which the experiment code failed to accurately track the group composition of the images shown in each trial. This resulted in a final sample of 90 participants, with 45 participants in each group (56 women, 34 men, $M_{age} = 21.74$, $SD_{age} = 1.85$).

2.1.2. Stimuli

The experiment was designed and executed using PsychoPy 2022.2.5 (Peirce et al., 2019). Stimuli were presented on an LCD monitor with a resolution of 1920 × 1080 and a refresh rate of 360 Hz. Participants were seated 60 cm from the monitor with their heads positioned on a head–chin rest in a dark room.

The image stimuli, obtained from Google Images, comprised 100 photographs of real politicians, who wore colors indicative of their party affiliation: Democratic Party politicians wore blue, and People's Power Party politicians wore red. In each trial, 100 images were randomly selected from a pool of 95 images of female Democratic Party politicians, 94 images of male Democratic Party politicians, 88 images of female People's Power Party politicians, and 90 images of male People's Power Party politicians to create a 10 × 10 matrix (21.36° × 21.36°). Either gender or political party affiliation could independently be in the minority, which resulted in four possible combinations: female minority and Democratic Party minority, female minority and People's Power Party minority, male minority and Democratic Party minority, and male minority and People's Power Party minority. The proportions were manipulated independently, and a minority group in each dimension was defined by a percentage randomly selected from a uniform distribution ranging from 15 % to 45 %, maintaining an average of 30 % across the experiment. Additionally, participants were randomly assigned to one of the two groups: a group with men in the minority (male minority group) or a group with women in the minority (female minority group). In both groups, Democratic Party politicians were in the minority for half the trials, and People's Power Party politicians were in the minority for the other half.

Participants' conceptual associations between political party affiliation and gender stereotypes were assessed on a 9-point Likert scale, which was specifically adapted from the questionnaire on gender stereotype endorsement (GSE) (Cundiff & Vescio, 2016; Oh et al., 2020) to probe whether a given gender-stereotypic word was more indicative of an average Democratic Party politician or an average People's Power Party politician. Twenty gender-stereotypic Korean words (Lee & Bang, 2003) were presented in a random order (1: Explains Democratic Party politicians the best – 9: Explains People's Power Party politicians the best), as shown in Table 1.

2.1.3. Procedure

Both groups completed identical tasks. Participants estimated the percentage of one of six groups—women, men, Democratic Party politicians, People's Power Party politicians, people wearing masks (filler), and people wearing glasses (filler)—to the best of their ability from a briefly presented picture of 100 politicians (Fig. 1, immediate estimation). They were not informed about the nontarget dimension. Each trial began with a 2000 ms display of a 10 × 10 matrix of pictures of politicians, with proportions of gender and political party affiliation varying across trials. Immediately after the presentation, participants entered their estimate for the probed target group in a blank text box under the question asking for the percentage of the group within the matrix. Participants pressed the space key to save their response and proceeded to the next trial. Trials were conducted in a self-paced manner, with no time limits or feedback provided. The sequence of trials was random and varied across participants. The task consisted of 120 trials: 2 minority

Table 1
Twenty Gender Stereotypical Korean Words Translated to English.

Stereotypically male words	Stereotypically female words
Wearing the pants	Cruel
Arrogant	Two-faced
Smart	Jealous
Insensitive	Beautiful
Relaxed	Decorated
Serving in the military	Devoted
Making money	Neurotic
Dependable	Earnest
Selfish	Lovely
Wolf-like	Tearful

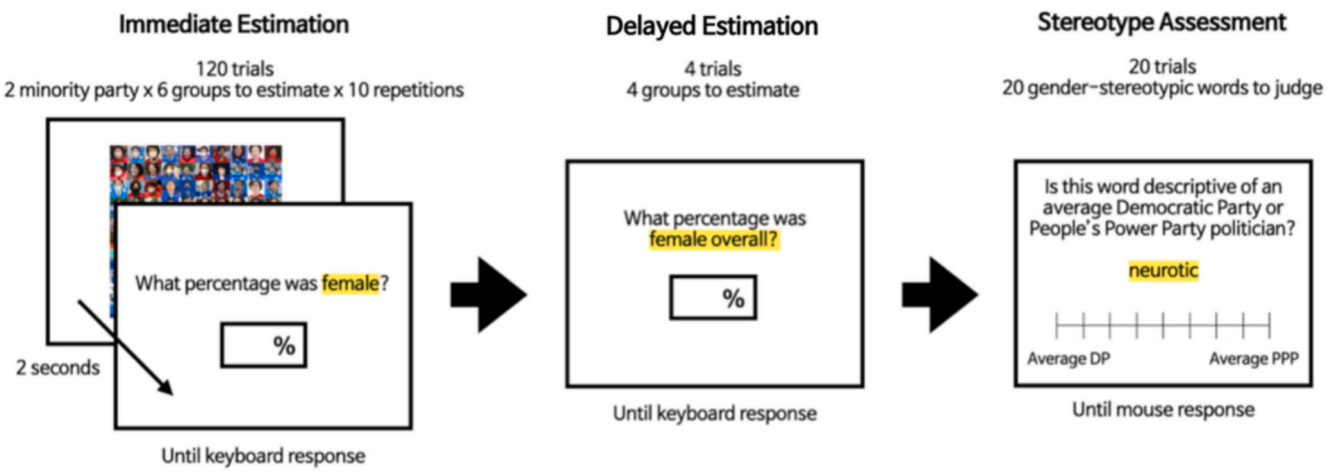


Fig. 1. Experiment 1 Procedure Overview

Note. All participants, regardless of group assignment, completed the tasks in identical order, from immediate estimation to delayed estimation to stereotype assessment. Immediate Estimation: In each trial, images of Korean politicians affiliated with either the Democratic Party or the People's Power Party were presented for two seconds. Participants immediately estimated the proportion of a given group from the image that had just been presented: male, female, Democratic Party politicians, People's Party politicians, those wearing glasses, and those wearing a mask. Delayed Estimation: Participants estimated the overall percentage of each of the four groups—male, female, Democratic Party politicians, and People's Party politicians—throughout the presentation. Stereotype Assessment: In each trial, people indicated whether a given word connoting a gender stereotype is more representative of an average Democratic Party politician or an average People's Power Party politician using a 9-point Likert scale. Images are not drawn to scale. Text highlighted in yellow indicates changeable contents depending on estimation questions and was not present in the actual study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

party conditions (either Democratic Party politicians or People's Power Party politicians in the minority) × 6 target groups × 10 repetitions.

After completing this task, participants were asked to estimate the overall percentage of each of the four groups across all trials: women, men, Democratic Party politicians, and People's Power Party politicians (Fig. 1, delayed estimation). By design, the correct answer for Democratic Party politicians and People's Power Party politicians was always 50 %; the correct answer for women and men depended on the group assignment, where the correct answer for the minority gender was always 30 % and 70 % for the majority gender. As in the immediate estimation task, participants were not timed and did not receive feedback for their responses. Following this task, participants completed a surprise task (Fig. 1, stereotype assessment) that probed their associations between gender and political party affiliations. Twenty Korean words associated with gender stereotypes (Table 1) were presented in a random order, and participants rated whether each word described an average Democratic Party politician or an average People's Power Party politician better using a 9-point Likert scale. This task was also untimed. Finally, participants provided demographic information, including age, gender, political orientation, and party support. After completing all tasks, participants were debriefed and released.

2.1.4. Analysis

The raw data and analyses code of Experiment 1 were reported in the Open Science Framework (https://osf.io/d4gxp/). The estimation errors were defined as the difference between the participant responses and the actual proportions of each group, calculated by subtracting the actual proportion from the response. Positive and negative estimation errors indicated overestimation and underestimation, respectively. These estimation errors were analyzed across conditions for the majority gender, minority gender, Democratic Party politicians, and People's Power Party politicians. Trials estimating the proportions of politicians wearing masks or glasses were excluded, as they were filler trials. Responses exceeding 100 % were treated as typos and were excluded from the analysis, which resulted in the removal of nine trials. Additionally, responses containing special characters or letters were excluded, which resulted in the removal of three trials. In total, twelve out of 7200 trials were excluded from the analysis.

Next, analyses of covariance (ANCOVAs) were conducted to evaluate trends in estimation errors for both the immediate and delayed estimation tasks. For the immediate estimation task, we examined whether the estimation errors for the four target groups varied depending on the minority political party and the overall gender ratio, controlling for participants' gender, age, political orientation, and party support. For the delayed estimation task, we assessed whether the estimation errors for the four groups varied depending on the conditions.

To investigate whether people who demonstrated stronger gender–party stereotypical thinking exhibited stronger compensation for unlikely contexts, individual GSE scores were correlated with the degree of compensation, i.e., the degree to which people provided higher estimates of female politicians when Democratic Party politicians were in the minority and the degree to which people provided higher estimates of male politicians when People's Power Party politicians were in the minority.

Individual GSE scores were calculated based on participants' responses to the surprise stereotype assessment task. The questions assessed participants' beliefs about whether stereotypically male or female descriptors were more characteristic of an average Democratic Party politician or a People's Power Party politician using a 9-point Likert scale (1 = the strongest resemblance to a Democratic Party politician, 9 = the strongest resemblance to a People's Power Party politician). To examine the conceptual associations between the Democratic Party and women and the People's Power Party and men, responses to stereotypically female descriptors were reverse-coded. This ensured that a low Likert response for stereotypically female descriptors aligned with a stronger association between the Democratic Party and women. For example, a response of 1 to a stereotypically female word was equivalent to a response of 9 to a stereotypically male descriptor. The twenty coded responses were summed up to produce an individual GSE score ranging from 20 to 180. Higher GSE scores indicated a stronger conceptual association between the Democratic Party and women and the People's Power Party and men, while lower scores indicated the opposite association. A score of 100 reflected no association in either direction. To investigate whether differences in responses were related to individual levels of stereotype endorsement regarding gender and political affiliation, we correlated individual GSE scores with the differences in

estimation error between nontarget proportion conditions at each level of the target group.

2.2. Results and discussion

2.2.1. Immediate estimation

Fig. 2 presents the results of the immediate estimation task. We conducted a three-way repeated-measures ANCOVA between the target group, minority gender, and minority party, controlling for participant gender, age, political orientation, and party support. A significant three-way interaction emerged ($F(3, 694) = 3.09$, $p = .03$, $\eta_p^2 = 0.01$), and the interaction was further probed with simple effects analyses.

When estimating the percentage of the minority gender, the interaction between minority gender and minority party was significant ($F(1, 166) = 5.22$, $p = .02$, $\eta_p^2 = 0.03$). When Democratic Party politicians were in the minority, people estimated female politicians to be more prevalent than male politicians ($F(1, 78) = 16.24$, $p < .001$, $\eta_p^2 = 0.17$). People estimated female politicians to be more prevalent when Democratic Party politicians were in the minority, compared to when People's Power Party politicians were in the minority ($F(1, 78) = 7.59$, $p = .01$, $\eta_p^2 = 0.09$). When People's Power Party politicians were in the minority, estimation errors for estimating female politicians and male politicians did not differ significantly ($F(1, 78) = 1.92$, $p = .17$, $\eta_p^2 = 0.02$). Estimates for male politicians did not differ depending on which party was in the minority ($F(1, 79) = 0.30$, $p = .59$, $\eta_p^2 = 0.00$).

When estimating the percentage of the majority gender, people underestimated the prevalence of male politicians significantly more than they did female politicians ($F(1, 166) = 6.27$, $p = .01$, $\eta_p^2 = 0.04$). However, which party was in the minority did not significantly change the pattern of estimation error ($F(1, 166) = 2.54$, $p = .11$, $\eta_p^2 = 0.01$). When estimating the percentage of the minority party, people overestimated the prevalence of Democratic Party politicians compared to People's Power Party politicians ($F(1, 166) = 7.25$, $p = .01$, $\eta_p^2 = 0.04$). Which gender was in the minority did not significantly change the pattern of estimation error ($F(1, 166) = 0.37$, $p = .55$, $\eta_p^2 = 0.00$). When estimating the percentage of the majority party, people underestimated the prevalence of People's Power Party politicians significantly more than they did for Democratic Party politicians ($F(1, 166) = 25.77$, $p < .001$, $\eta_p^2 = 0.13$). Which gender was in the minority did not significantly change the pattern of estimation error ($F(1, 166) = 2.04$, $p = .15$, $\eta_p^2 = 0.01$).

2.2.2. Delayed estimation

Fig. 3 presents the results of the delayed estimation task. For patterns of gender estimation error, we conducted two-way repeated-measures ANCOVA between estimated gender and group, controlling for participant age, gender, and political and voting preferences. Participants' error was larger when the estimated gender was in the minority rather

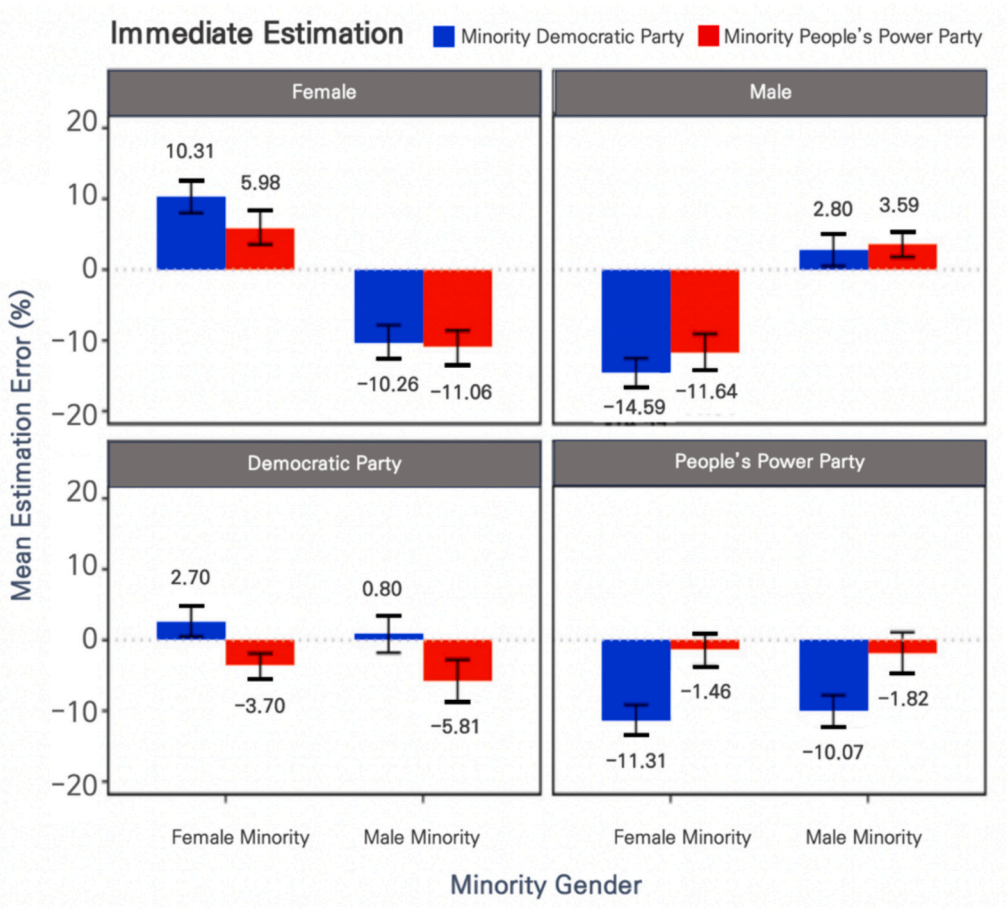


Fig. 2. Patterns of Immediate Estimation Error by Target Group

Note. Patterns of mean estimation error depending on which party was in the minority for estimating male, female, Democratic Party, and People's Power Party politicians. Each subplot refers to the estimated group. Within each subplot, pairs of two bars on the left and right depict responses from the female minority and male minority groups, respectively. Bar colour denotes which political party was in the minority, with blue denoting minority Democratic Party and red denoting minority People's Power Party. Error bars indicate 95 % confidence intervals and numbers in the graph indicate mean estimation errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

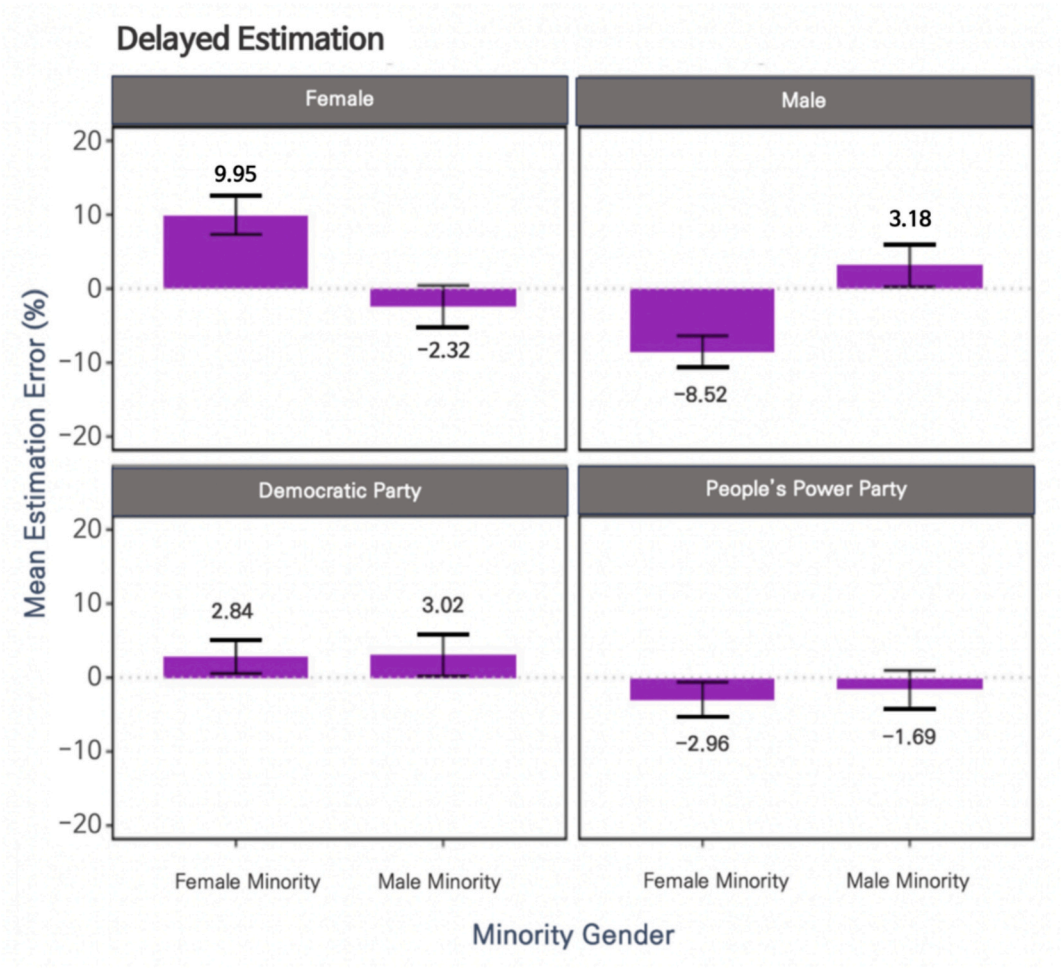


Fig. 3. Patterns of Delayed Estimation Error by Estimated Group
Note. Patterns of delayed estimation error for minority gender, majority gender, Democratic Party politicians, and People's Power Party politicians. Each subplot refers to the estimated group. Within each subplot, bars on the left and right depict responses from the female and male minority groups, respectively. Error bars indicate 95 % confidence intervals.

than majority ($F(1, 163) = 21.92$, $p < .001$, $\eta_p^2 = 0.12$) both for the male minority ($F(1, 78) = 6.42$, $p = .01$, $\eta_p^2 = 0.08$) and female minority ($F(1, 76) = 101.89$, $p < .001$, $\eta_p^2 = 0.57$) groups. The simple effect of estimated gender was further decomposed with Scheffe post hoc tests: participants assigned to the female minority group, compared to participants assigned to the male minority group, overestimated minority gender more (9.95 % vs. 3.18 %; $t(163) = -3.22$, $p = .002$) and underestimated majority gender more (−8.52 % vs. -2.32 %; $t(163) = 3.19$, $p = .002$). Likewise, participants' error differed according to group assignment—e.g., which gender was in the minority—when estimating minority gender ($F(1, 163) = 10.38$, $p = .002$) and majority gender ($F(1, 163) = 10.18$, $p = .002$). Scheffe post hoc tests revealed that people provided higher estimates for minority gender than they did for majority gender in both the male minority group (3.18 % vs. −2.32 %; $t(163) = 2.83$, $p = .01$) and female minority group (9.95 % vs. −8.52 %; $t(163) = 9.43$, $p < .001$).

By contrast, in case of party estimation, two-way ANCOVAs between estimated party and group (controlling for participant age, gender, and political and voting preferences) revealed that the two-way interaction between estimated party and group was not significant ($F(1, 166) = 0.17$, $p = .68$, $\eta_p^2 = 0.001$). Participants assigned to the male minority group and female minority group did not differ in their patterns of delayed estimation error for party estimation ($F(1, 166) = 0.27$, $p = .60$, $\eta_p^2 = 0.002$). As the main effect of estimated party was significant ($F(1, 166) = 15.58$, $p < .001$, $\eta_p^2 = 0.09$), a Scheffe post hoc test was conducted to probe the effect. Democratic Party politicians (2.93 %) were estimated to be higher in proportion than People's Power Party politicians (−2.33 %), despite the fact that the party ratio was maintained at 50:50, in both male minority (3.02 vs. −1.69 %; $t(166) = 2.50$, $p = .01$), and female minority (2.84 vs. −2.96 %; $t(166) = 3.08$, $p = .002$) groups.

2.2.3. GSE

A one-sample *t*-test revealed that the mean GSE score ($M = 103.80$, $SEM = 0.78$) among participants was significantly different from 100 ($t(89) = 4.86$, $p < .001$). A value larger than 100 would indicate a stereotypical association between gender and politics, suggesting an overall trend of associating femininity with the Democratic Party and masculinity with the People's Power Party. To investigate whether people who demonstrated stronger gender–party stereotypes exhibited stronger reported compensation, individual GSE scores were correlated with the degree to which participants provided higher estimates of female politicians when Democratic Party politicians were in the minority and higher estimates of male politicians when People's Power Party politicians were in the minority. However, none of the computed Pearson's correlation coefficients reached significance. In the case of estimating the proportion of females, the difference in estimation error between the minority Democratic Party and the minority People's Power Party conditions was not correlated with GSE. Both when female politicians were in the minority ($r(43) = -0.01$, $p = .94$) and when they were in the majority ($r(43) = -0.10$, $p = .53$), the correlation between

GSE and the degree to which participants provided higher estimates of female politicians when Democratic Party politicians were in the minority was not significant. Likewise, in the case of estimating the proportion of males, the difference in estimation error between the minority People's Power Party and the minority Democratic Party conditions was not correlated with GSE. Again, both when male politicians were in the minority ($r(43) = -0.14$, $p = .36$) and when they were in the majority ($r(43) = -0.09$, $p = .56$), the correlation between GSE and the degree to which participants provided higher estimates of male politicians when People's Power Party politicians were in the minority was not significant. Overall, it seems that the conceptual association between gender and party affiliation, at least as measured in this study, is unrelated to the degree of compensation in bias in demographic number estimation.

In sum, we find that in an estimation task where proportions of target and nontarget dimensions are independently manipulated, estimators make systematic errors in both immediate and delayed judgments, consistently overestimating numeric minorities and underestimating numeric majorities. It is noteworthy that delayed estimation errors were either lower or comparable to immediate estimation errors despite the considerable delay in estimation time. We also find that while this qualitative pattern remains stable across conditions, the magnitude of these estimation errors varies depending on the context. For example, female politicians are generally perceived as more prevalent than male politicians, especially so when participants make their estimates in contexts where the display is predominantly composed of People's Power Party politicians. While seemingly counterintuitive, this pattern of results is consistent with our central hypothesis: under unlikely contexts, participants increasingly shift their responses toward the likely value as a strategy to compensate for uncertainty. Returning to the example from the introduction to this experiment, spotting female politicians in a crowd of politicians is relatively unlikely, especially so when the crowd predominantly comprises People's Power Party politicians. As a result, rational estimators hedge their estimates more strongly toward the expected value, leading to greater shifts in these conditions. The robust overestimation of minorities and underestimation of majorities reflects this compensation strategy, and the extent of this compensation is flexibly adjusted based on the contextual likelihood even when estimating the same target group. This strategy persists while controlling for demographic factors such as participant age, gender, and voting preferences and is unrelated to participants' endorsement of gender–political party stereotypes.

3. Experiment 2

In Experiment 1, participants exhibited a consistent pattern of overestimating minority groups and underestimating majority groups, with the magnitude of these estimation errors varying depending on the context, consistent with a rational compensation strategy. That is, participants compensated for uncertainty in demographic estimation by shifting their estimates toward a "likely" value, and did so to an even greater extent in contexts that were "unlikely." This response adjustment was not fixed but rather flexibly modulated based on the nontarget proportion, even given the brief exposure time of just two seconds, which suggests that participants systematically incorporated likelihood when estimating demographic compositions.

Building on these findings, Experiment 2 examines a related question: when do people perceive that a group constitutes the majority (i.e., over half of the total)? The intergroup relations literature suggests that group size is a general feature that signals political power and that the larger the groups become, the less they are tolerated by majority groups (Blumer, 1958; Cikara et al., 2022; Quillian, 1995; Schlueter & Scheepers, 2010). This question is also relevant in our context of political representation, where majority status confers significant implications; theories of congressional organization suggest that holding a majority position carries substantial procedural advantages (Aldrich & Rohde, 1997; Cox & Magar, 1999; Cox & McCubbins, 1993). Thus, understanding the perceptual threshold for majority status can provide insights into how people intuitively judge political power dynamics.

In this context, Experiment 2 investigates the perceptual thresholds at which participants categorize different groups as comprising the majority. To do so, we presented visual stimuli with systematically varied proportions (ranging from 10 % to 90 %) of male and female politicians and examined how participants determined whether the target group constituted the majority. Each display paired target groups (e.g., women) with nontarget groups that were either relatively more likely (e.g., Democratic Party politicians) or relatively less likely (e.g., People's Power Party politicians). Participants judged whether the proportion of the target group exceeded an implicit standard (i.e., 50 %). By spanning a broad range of proportions, this design enabled us to examine how proportional variation interacts with likelihood and influences participants' majority judgments. We anticipated that the probability with which participants indicated that a group comprised the majority would depend on the likelihood of spotting the target group in a given context.

3.1. Methods

3.1.1. Participants

We recruited 39 participants (19 women, 20 men, $M_{age} = 22.10$) aged 19 years or older who were all native Korean speakers with normal or corrected-to-normal vision from the university pool in exchange for course credit. As in Experiment 1, we based our sample size on Anderson et al. (2017). We aimed for a similar statistic to that obtained in Experiment 1, which was nearly identical in design to Experiment 2. Based on an F value of 16.24, obtained from testing the effect of the nontarget proportions, we arrived at a minimum required sample size of 38 participants.

3.1.2. Stimuli

We used the same images and word stimuli as in Experiment 1, and the experiment was designed and implemented using PsychoPy 2023.2.3 (Peirce et al., 2019). All stimuli were presented on an LCD monitor with a resolution of 2560 × 1440 and a refresh rate of 144 Hz. Participants were seated 60 cm from the monitor with their heads positioned on a head–chin rest in a dark room.

There were two main differences in how the stimuli were presented in Experiment 2 compared to Experiment 1. First, unlike Experiment 1, each target group in Experiment 2—i.e., female and male politicians—was represented in five equidistant intervals ranging from 10 % to 90 %: 10 %, 30 %, 50 %, 70 %, and 90 %. Second, in Experiment 2, the trials were divided into two conditions based on likelihood. In "likely" trials, female politicians were always affiliated with the Democratic Party, and male politicians were always affiliated with the People's Power Party. In "unlikely" trials, female politicians were always affiliated with the People's Power Party, and male politicians were always affiliated with the Democratic Party. This design differed from that in Experiment 1, where the two dimensions were independently manipulated.

3.1.3. Procedure

We compared estimation trends across the likelihood conditions using the method of single stimuli (Morgan et al., 2000). This method comprises two sessions. In the first session, participants engage in a standard two-interval forced-choice task, where the display in the first interval serves as a standard for comparison for the second interval display. In the second session, participants use the implicit mean of all the standard stimuli from the first session as the reference.

In our study, participants completed 42 blocks, with 21 blocks designated each for estimating the proportion of either men or women. The first session comprised a single block, while the second session included 20 blocks. In the first block, participants performed a two-interval forced-choice task, judging whether the second image contained a higher percentage of a specific gender group than the

proportion shown in the first image. Each trial consecutively displayed two 10 × 10 matrices for 2000 ms each. Importantly, the display in the first interval always showed a 50:50 sex ratio. Participants pressed the right arrow key if they perceived the second image to contain a larger proportion of the given gender and the left arrow key if not. They then proceeded to the next trial upon pressing the key.

In the subsequent 20 blocks, participants evaluated whether the presented image contained a higher percentage of the specified group than the mean proportion of the group shown in the first intervals during the first block. In these trials, only the image to be judged was presented (single interval), without the 50:50 display preceding it. Trials were self-paced, and no feedback was provided. The sequence of trials was random and rerandomized within each block, and it varied across participants. A total of 400 trials were conducted (2 groups to be estimated × 5 different proportions × 2 likelihood conditions × 20 repetitions), with separate sets of 21 blocks for estimating female and male politicians. After completing the task, participants proceeded to the same surprise stereotype assessment task as in Experiment 1. Participants were debriefed and released after providing demographic information, including participant age and gender.

3.1.4. Analysis

The raw data and analyses code of Experiment 2 were reported in the Open Science Framework (https://osf.io/d4gxp/). We fit a generalized linear mixed model to examine the pattern of estimation errors regarding the percentage of the target gender group, depending on whether the target group could also be characterized as representing another group that was more versus less likely. The model predicted the likelihood of participants indicating that the test image contained a higher percentage of the gender in question than the implicit mean of the standard images from the first interval, referred to as the "greater than standard" response for brevity. The model included the target gender (male vs. female), likelihood (likely vs. unlikely), percentages of the target gender (10, 30, 50, 70, and 90 %), all possible interactions, and a random effect for participant ID.

The GSE scores were obtained in the same manner as in Experiment 2. Following the approach in Experiment 1, we examined whether a stronger endorsement of gender–party stereotypes was associated with greater likelihood-based compensation. To this end, we correlated individual GSE scores with three measures of compensation with point-biserial correlation. Specifically, we correlated GSE scores with (a) the extent to which participants were more likely to give a "greater than standard" response at 50 % target gender proportions when estimating the prevalence of women compared to men, (b) the extent to which they were more likely to give a "greater than standard" response at 50 % target gender proportions when estimating female People's Power Party politicians compared to female Democratic Party politicians, and (c) the extent to which they were more likely to do so when estimating male Democratic Party politicians compared to male People's Power Party politicians.

3.2. Results and discussion

3.2.1. Trends in estimation

The likelihood of perceiving the test image as containing a higher percentage of the target gender than the standard (i.e., the implicit mean of all standard images in the first interval, consistently set at 50 %) is shown in Fig. 4. The generalized linear mixed model predicted the likelihood of a "greater than standard" response based on the target gender, its percentages (mean-centered), the likelihood of the display, all possible interactions, and participant ID as a random intercept.

Participants were generally more likely to indicate that female politicians ($M = 0.45$, $SE = 0.02$, 95 % CI [0.41, 0.48]) were more prevalent than the standard image—that is, in the majority—than male politicians ($M = 0.40$, $SE = 0.02$, 95 % CI [0.36, 0.44]), $OR = 0.83$, $SE = 0.04$, $z = -3.62$, $p < .001$). When participants assessed whether female politicians were more prevalent than in the standard image, they were sensitive to the percentage of the target group in the display ($b = 0.10$, $SE = 0.003$, $z = 33.56$, $p < .001$), but this sensitivity did not differ by display likelihood ($b = 0.003$, $SE = 0.004$, $z = 0.61$, $p = .55$). Participants were more likely to indicate that the display contained more female politicians than the standard (i.e., 50 %) when the test display was likely ($b = -0.31$, $SE = 0.07$, $z = -4.24$, $p < .001$). In other words, participants perceived a greater number of female politicians when the female politicians were affiliated with the progressive Democratic Party than with the conservative People's Power Party.

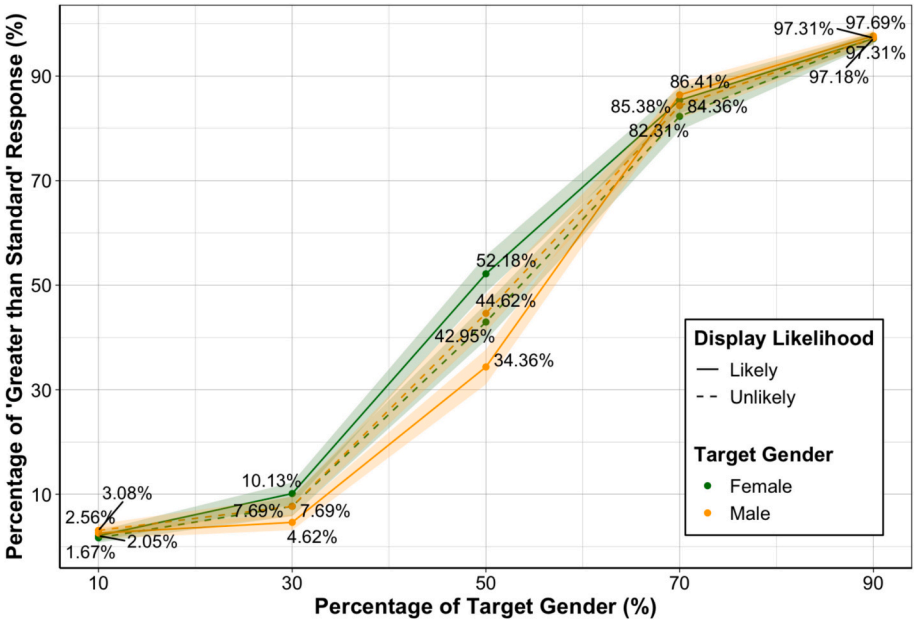


Fig. 4. Percentage of "Greater than Standard" Response per Target Gender, its Percentages, and Display Likelihood
Note. Green and yellow lines indicate patterns of response for female and male politicians, respectively. Solid and dotted lines depict responses for "likely" and "unlikely" trials, respectively. Ribbons indicate 95 % confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

When participants assessed the prevalence of male politicians, their sensitivity to the percentage of the target group in the display was moderated by likelihood ($b = -3.10$, $SE = 0.005$, $z = -3.10$, $p = .002$). Simple slopes analysis revealed that this relationship was steeper when the test display was likely ($b = 0.12$, $SE = 0.00$, $z = 32.54$, $p < .001$) than when it was unlikely ($b = 0.10$, $SE = 0.00$, $z = 33.55$, $p < .001$). In other words, participants were more likely to increase their estimates of male politicians when the display was "likely," indicating greater sensitivity to the canonical gender proportions in the display.

3.2.2. GSE

A one-sample *t*-test revealed that the mean GSE score ($M = 105.00$, $SEM = 1.47$) among the participants was significantly different from 100 ($t(38) = 3.39$, $p = .002$). To investigate whether stronger endorsement of gender–party stereotypes was associated with greater sensitivity to contextual likelihood and subsequent likelihood-based compensation, we conducted a series of point-biserial correlations between GSE scores and the three aforementioned measures of likelihood effects. We, thus, tested whether stereotype endorsement influenced the degree to which estimators incorporated contextual likelihood into their responses, as reflected in the percentage in which they indicated that a given group constituted the majority. The GSE scores were not significantly correlated with any of the likelihood-based measures, namely, (a) the likelihood of overestimating the prevalence of women compared to men at 50 % target gender proportions ($r(37) = 0.13$, $p = .42$), (b) the likelihood of overestimating women when they were associated with the People's Power Party compared to the Democratic Party ($r(37) = 0.13$, $p = .42$), or (c) the likelihood of overestimating men when they were associated with the Democratic Party compared to the People's Power Party ($r(37) = 0.14$, $p = .41$). Although all correlations were positive, none of the relationships reached statistical significance.

Taken together, the results reveal that when participants estimate whether a target group comprises the majority in a perceptual task where target groups are paired with nontarget groups and presented in systematically varied proportions, their responses differ across target groups and depending on the likelihood of the paired nontarget group. More specifically, female politicians were more likely to be judged as comprising the majority when affiliated with the Democratic Party, and participants exhibited greater sensitivity in judging male politicians as the majority when affiliated with the People's Power Party.

Fig. 4 indicates that participants displayed substantial sensitivity to the canonical proportions of the target groups across conditions. In other words, the degree of rational compensation as reported in Experiment 1 was not large enough to deter participants from accurately reporting the majority status of the target group. We speculate that one reason for this result may be differences in the task demands between Experiments 1 and 2. In Experiment 1, participants provided explicit numerical estimates, whereas in Experiment 2, they made categorical judgments. This implies that within our experimental context, where target group proportions ranged from 10 % to 90 %, initial perceptual judgments were relatively accurate, and the bulk of the estimation error may have been introduced by the formation of explicit numerical estimates.

4. Experiment 3

Experiments 1 and 2 demonstrated that participants systematically incorporated contextual likelihood when estimating the prevalence of a given group. Specifically, participants assessed whether the context was likely to contain members of a particular group and, when deemed unlikely, compensated for this uncertainty by shifting their estimates toward a more probable value (i.e., 50 %). Both experiments assumed that participants perceived women as more likely to be a Democratic Party politician/among majority Democratic Party politicians and men as more likely to be a People's Power Party politician/among majority People's Power Party politicians, potentially reflecting accurate yet uncertain knowledge about gender representation in the Korean parliament. Accordingly, in Experiment 3, we directly probed lay beliefs about likely gender representation.

A key factor that may shape this knowledge (and, therefore, perceptions of contextual likelihood) is semantic association. Research on associative judgments consistently demonstrates that humans often rely on co-occurrence-based statistical regularities when making judgments, which enables fast and efficient processing but is susceptible to errors (Bhatia, 2017; Tversky & Kahneman, 1983). A classic example of such error-prone judgment is illustrated by Kahneman and Tversky (1973), wherein participants systemically ignored base rates (e.g., the proportion of engineers and lawyers) when estimating the probability that "Jack" was an engineer/lawyer, given a description characteristic of a typical engineer.

Building on this insight, Experiment 3 examined whether estimates of demographic composition are influenced by information about the proportion of another dimension. Similar to Ponce de Leon et al. (2022), participants first read about a hypothetical survey in which respondents were described as either 25 % Democratic Party supporters and 75 % People's Power Party supporters or 25 % People's Power Party supporters and 75 % Democratic Party supporters. Participants then estimated the gender composition of the hypothetical survey respondents.

We expected that the participants in Experiment 3 would provide responses that deviated from the base rate of gender (i.e., 50 %). We also anticipated that their estimates of the percentage of men/women would vary depending on whether Democratic Party supporters or People's Power Party supporters were in the minority, reflecting gendered associations. Such results, where the estimates of the target group vary depending on the proportion of the nontarget dimension, would provide evidence of how associations—that is, prior beliefs—influence the perception of likely proportions, thereby helping to elucidate how people incorporate likelihood when estimating the demographic structure of society.

4.1. Methods

4.1.1. Participants

We recruited 52 participants (36 women, 15 men, 1 other; $M_{age} = 22$ years) from the university participant pool in exchange for course credit or monetary compensation of 3000 Won. The mean age was calculated for only 51 participants, as one participant responded with "20s" rather than providing a specific age.

Using a sequential Bayes factor design analysis (Schönbrodt & Wagenmakers, 2018), we planned to collect a minimum sample of 12 participants per group (i.e., 12 participants × 2 nontarget proportions × 2 target groups, which returns 48 participants). Data collection was to continue until strong evidence was obtained for or against the main effect of nontarget proportion on target group estimation. Following Anderson et al. (2017), we accounted for publication bias and uncertainty and set an assurance level of 0.80 and statistical power of 0.85. Our decision to collect a minimum sample size of 12 per group was based on the results from our Experiment 2, which examined visual perception of the proportions of multiple social categories.

All participants passed attention and memory checks. One participant's response was excluded from the analysis because it exceeded 100. This exclusion criterion was not preregistered. Thus, all analyses reported pertain to responses from the remaining 51 participants. As we obtained strong evidence for the effect of nontarget proportion information on estimating both women ($BF_{10} = 173.89$) and men ($BF_{10} = 439.34$), we stopped further data collection.

4.1.2. Stimuli

Participants were presented with a scenario describing a hypothetical survey of 1000 randomly sampled Korean citizens. The scenarios manipulated the proportion of the nontarget group (political party supporters). Specifically, participants read that Democratic Party supporters comprised 25 % of the group and People's Power Party

supporters 75 %, or vice versa (25 % People's Power Party supporters and 75 % Democratic Party supporters).

4.1.3. Procedure

After providing informed consent for the experiment as approved by the university's Institutional Review Board, participants were randomly assigned to one of two scenarios that manipulated the proportion of political party supporters (nontarget dimension) among 1000 hypothetical survey respondents. Participants then proceeded to answer filler questions about the respondents. Next, participants were instructed to imagine the likely gender makeup of the respondents, specifically reporting the likely percentage of either women or men (target group) among the respondents. Participants completed attention and memory check questions and provided demographic information. The attention check asked the participants to solve a simple arithmetic problem (17 + 2), and the memory check asked them to recall whether the hypothetical survey respondents were categorized by political party, gender, or age in the scenario.

4.1.4. Analysis

The design and analysis plan of Experiment 3 were registered before data acquisition in the Open Science Framework (https://osf.io/d4gxp/), and this site also contains the data and analyses code. To test our hypotheses, we conducted two Welch's independent-sample *t*-tests comparing gender composition estimates (target dimension) between the two scenario conditions (varying proportions of the nontarget dimension). We also examined whether these estimates deviated significantly from 50 %, the base rate of gender proportion expected given random sampling.

5. Results and discussion

Estimation responses exceeding 100 % were treated as typos and excluded from our analysis as, by definition, percentages cannot exceed 100 %. This resulted in the removal of one participant, leaving 51 participants in the total sample (12 participants in nontarget proportion = Democratic Party supporter minority and target group = male condition, 13 participants in nontarget proportion = People's Power Party supporter minority and target group = male condition, 14 participants in nontarget proportion = Democratic Party supporter minority and target group = male condition, 12 participants in nontarget proportion = People's Power Party supporter minority and target group = male condition).

Participants guessed that the 1000 survey respondents would comprise more women when the majority of respondents were Democratic Party supporters ($M = 57.92$, $SD = 8.14$) rather than People's Power Party supporters ($M = 41.33$, $SD = 9.81$, $t(21.47) = 4.58$, $p < .001$, $d = 1.85$). Conversely, they estimated that there would be a higher proportion of men when the majority of the respondents were People's Power Party supporters ($M = 66.93$, $SD = 7.68$) rather than Democratic Party supporters ($M = 49.50$, $SD = 10.06$, $t(20.43) = 4.90$, $p < .001$, $d = 1.97$).

Participants' responses deviated from 50 % when estimating the percentage of women when the majority of respondents were Democratic Party supporters ($t(12) = 3.51$, 95 % CI [53.00, 62.84], $p = .004$) and when the majority were People's Power Party supporters ($t(11) = -3.06$, 95 % CI [35.10, 47.57], $p = .01$). However, while participants' responses deviated from 50 % when estimating the percentage of men when the majority of respondents were People's Power Party supporters ($t(13) = 8.25$, 95 % CI [62.49, 71.36], $p < .001$), they did not when the majority were Democratic Party supporters ($t(11) = -0.17$, 95 % CI [43.11, 55.89], $p = .87$).

Thus, Experiment 3 demonstrates that in the hypothetical estimation task, gender (target) estimates were influenced by the proportion of political party supporters (nontarget). Specifically, participants estimated a higher proportion of women when the group was described as predominantly Democratic Party supporters and a higher proportion of men when the group was predominantly People's Power Party supporters. This pattern aligns with previous research suggesting that perceived political progressiveness is associated with perceived femininity, whereas perceived conservativeness is associated with perceived masculinity (Ji, 2018). Participants' responses generally deviated from the statistically rational base rate of 50 %, except when estimating the percentage of men in the Democratic Party majority scenario. This pattern is consistent with the assumptions of Experiments 1 and 2 and bolsters our account of the rational compensation strategy involved in visual demographic estimation.

6. General discussion

Following three experiments, we propose a rational compensation strategy for demographic estimation in which individuals increasingly adjust their estimates toward expected values as a function of the likelihood of the context in which the estimation is made. When making explicit judgments about the proportions of certain groups in a crowd, participants systematically overestimated minorities and underestimated majorities, and the amount of adjustment differed depending on the likelihood of nontarget groups (Experiment 1). When making categorical judgments about whether a group comprises the majority, participants demonstrated a good sensitivity to canonical proportions, and they incorporated contextual likelihoods (Experiment 2). Then, in Experiment 3, perceptions of likelihood were grounded in lay beliefs about gender diversity within specific contexts (e.g., political party proportions in parliament), which shaped how participants integrated contextual information into their judgments. These findings collectively suggest that even when estimating groups of identical size, participants produce different numerical estimates depending on the context, likely in an attempt to compensate for the uncertainty associated with estimating the size of a group in a likely (vs. unlikely) context. In other words, rather than relying solely on their perception, individuals incorporate expectations about how prevalent a group *ought* to be in a given setting.

First and foremost, we replicate the pattern reported in the literature of minority overestimation and majority underestimation in perceptual estimation tasks and attest to the robustness of this phenomenon (Landy et al., 2018; Lee et al., 2019). A similar pattern has also been observed in frequency judgments of lethal events, where low frequencies tend to be overestimated and high frequencies are underestimated (Hertwig et al., 2005; Lichtenstein et al., 1978). While previous accounts of the perceptual mechanism underlying demographic perception suggest that estimates are simply pulled toward prior expectations about group size (Guay et al., 2025), our results indicate a more nuanced and context-sensitive process. Specifically, while participants do incorporate prior expectations, they seem to also use contextual information to evaluate whether and how much to compensate for their estimation rather than passively adjusting toward a likely value. If estimates were hedged purely toward prior expectations, then, all else being equal, the group whose prevalence is more contextually expected among two minority groups of equal size should be overestimated to a greater degree, reflecting a higher prior expectation for that group. However, our findings suggest that rather than solely relying directly on these static prior beliefs in their estimation process, participants actively assessed the likelihood of their estimate within the surrounding context—specifically, by considering the proportion of nontarget groups—and use this information to determine the degree of compensation for their estimation. As a result, in less likely contexts, we find that participants shift their estimates more dramatically, which suggests that estimation errors emerge from an active process of contextual inference.

This clarification is critical because, in reality, demographic estimation rarely occurs in isolation. Indeed, individuals belong to multiple overlapping social categories, each carrying different associations. We find that these associations serve as contextual cues that shape the

process of estimating a target group. In this way, our study contributes to domain-general accounts of demographic estimation by elucidating *how* context shapes the hedging process. It also offers a possible explanation as to why the degree of overestimation or underestimation may differ across different groups (Landy et al., 2018) and why the estimation of socially marginalized groups, such as Black people, is prone to larger estimation errors compared to dominant groups, such as White people (Kardosh et al., 2022).

While the pattern of estimation errors in Experiment 1 aligns with our account of likelihood-based compensation, in Experiment 2, participants also exhibited a good sensitivity to canonical proportions. That being said, participants' majority judgment trends varied depending on the context, which suggests the presence of likelihood-based compensation; for instance, participants more often judged that the display contained more female politicians overall. However, visual inspection of Fig. 4 suggests that even if likelihood-based adjustment did occur, it was not strong enough to render majority judgments systematically inaccurate. One possibility is that participants generated estimates for target and test images in the same manner as in Experiment 1 but, because the task was relatively easier and free of bias in Experiment 2, the magnitude of estimation errors was relatively small and insufficient to bias majority judgments.

Specifically, the two tasks in Experiments 1 and 2 engaged different stages of the likelihood-based estimation process. In Experiment 1, participants generated explicit numerical estimates, for which they were required to translate percepts into precise quantitative judgments, whereas in Experiment 2, participants made categorical judgments that did not require numerical transformation. In both experiments, the initial percepts may have been relatively accurate (but prone to some error), which could explain participants' sensitivity to canonical proportions in Experiment 2, where they reached a decision without forming an explicit estimate. However, in Experiment 1, additional errors may have arisen due to participants' adjustment of their uncertain but relatively accurate raw percepts following consideration of the context. Thus, the bulk of the estimation errors may have been primarily introduced when forming explicit numerical estimates, as in Experiment 1, whereas in Experiment 2, where relative judgments sufficed, these additional adjustments may not have occurred to the same extent. Because the reference point of numerical estimates was provided from the beginning in Experiment 2, participants were able to reduce the errors in their relative judgments (Izard & Dehaene, 2008). This account seems reasonable given that people are indeed experts at extracting a multitude of information ranging from social categories like sex, age, and race to transient properties such as emotion, motivation, and intention and traits like dominance and trustworthiness from faces (Dobs et al., 2019; Young & Bruce, 2013). Regardless, while likelihood-based adjustments may fairly influence explicit numerical estimation, they appear less impactful in relative, categorical judgments about majority status.

The findings from this study also elucidate how minority overestimation and majority underestimation may be corrected to foster a more accurate perception of the social world. Although the estimation errors reported here are relatively small compared to those documented in the literature (though this may be due to the nature of the visual estimation task we employed), they consistently occur in the same direction and become especially pronounced in unlikely contexts. People remain prone to systematic misperceptions, regardless of explicit biases, which can unintentionally hinder efforts to promote diversity—even assuming the absence of overt prejudice. The lack of correlation between stereotype endorsement and the degree of estimation error suggests that, even if gender bias were to be mitigated, it would play a minimal role in reducing error. However, we do find in Experiment 1 that while participants adjusted their estimates in unlikely contexts, this adjustment-based estimation error diminished when forming delayed estimates, where participants generated an aggregated estimate based on prior assessments. This is surprising because the delayed estimation incorporates an additional memory component to the immediate estimation. One possible explanation is that through repeated exposure to a given group in a given context, even when their presence is believed to be unlikely given lay associations, participants accumulate more exposure and information, harnessing the wisdom of the inner crowd (Herzog & Hertwig, 2014). As a result, participants' estimation errors are canceled out over repeated immediate judgments, and they ultimately arrive at more accurate final responses. If participants improve their assessments based on increased exposure—that is, more encounters with minority group members—then one way to address this "innumeracy" is to make such encounters less unlikely.

Consistent with the proposed rational compensation strategy, estimation errors can be further reduced if social influence is leveraged appropriately. For instance, individuals may benefit more from others' input in decentralized communication networks than in centralized ones, where dominant individuals are more likely to increase group error (Becker et al., 2017). Moreover, exposure to others' overestimation can help observers correct their own underestimation (Jayles et al., 2021). Finally, the utility of the wisdom of crowds can be enhanced by employing more effective aggregation methods for individual estimates (Kao et al., 2018).

Overall, our findings indicate that estimation errors in demographic judgments may reflect a rational attempt to incorporate context when estimating proportions. While our results replicate the general pattern of relative frequency misjudgments, characterized by overestimating small quantities and underestimating large ones, they also provide a novel perspective by demonstrating how contextual factors shape these compensatory adjustments. That is, in an attempt to compensate for uncertainty in estimating proportions within unlikely contexts, participants systematically adjust their estimates, leading to the consistent overestimation of minorities and the underestimation of majorities. Furthermore, the degree of adjustment depends on the likelihood of the context, which reinforces the idea that the demographic estimation process is not purely perceptual but rather actively influenced by contextual inferences. Our findings suggest that improving people's perceptions of social reality may require the very thing Diversity, Equity, and Inclusion policies aim to promote—greater diversity, especially in contexts where minority representation is particularly unlikely.

7. Open practices

Experiments 1 and 2 were not preregistered. The design and analysis plan of Experiment 3 were registered before data acquisition in the Open Science Framework (https://osf.io/d4gxp/). This site also contains the data and analyses codes of all three experiments.

CRediT authorship contribution statement

Yoonseo Heo: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sang Chul Chong:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Ethics approval

The study was approved by the Institutional Review Board (IRB) of the Department of Psychology at Yonsei University with IRB number 7001988-202409-HR-2223-05.

Funding

This work was supported by the National Research Foundation (NRF) of Korea grant funded by the Korea government (MSIT) (RS-2022-NR070542).

Declaration of competing interest

The authors declare that they have no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

All data are available at Open Science Framework (https://osf.io/d4gxp/)

References

Aldrich, J. H., & Rohde, D. W. (1997). The transition to republican rule in the house: Implications for theories of congressional politics. *Political Science Quarterly, 112*(4), 541–567. https://doi.org/10.2307/2657691

Alt, N. P., Dayley, P. N., Faulkner, A. G., & Johnson, K. L. (2024). Gender/sex categories and gendered cues in people perception: The influence of gender/sex ratio and gendered appearance on group judgments. *Social Cognition, 42*(2), 112–132. https://doi.org/10.1521/soco.2024.42.2.112

Alt, N. P., Goodale, B., Lick, D. J., & Johnson, K. L. (2019). Threat in the company of men: Ensemble perception and threat evaluations of groups varying in sex ratio. *Social Psychological and Personality Science, 10*(2), 152–159. https://doi.org/10.1177/1948550617731498

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science, 28*(11), 1547–1562. https://doi.org/10.1177/0956797617723724

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences, 114*(26), E5070–E5076. https://doi.org/10.1073/pnas.1615978114

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review, 124*(1), 1–20. https://doi.org/10.1037/rev0000047

Blumer, H. (1958). Race prejudice as a sense of group position. *The Pacific Sociological Review, 1*(1), 3–7. https://doi.org/10.2307/1388607

Cikara, M., Fouka, V., & Tabellini, M. (2022). Hate crime towards minoritized groups increases as they increase in sized-based rank. *Nature Human Behaviour, 6*(11), 1537–1544. https://doi.org/10.1038/s41562-022-01416-5

Cox, G. W., & Magar, E. (1999). How much is majority status in the US congress worth? *American Political Science Review, 93*(2), 299–309. https://doi.org/10.2307/2585397

Cox, G. W., & McCubbins, M. D. (1993). *Legislative leviathan: Party government in the house*. University of California Press.

Cundiff, J. L., & Vescio, T. K. (2016). Gender stereotypes influence how people explain gender disparities in the workplace. *Sex Roles, 75*(3), 126–138. https://doi.org/10.1007/s11199-016-0593-2

Daniels, D. P., Neale, M. A., & Greer, L. L. (2017). Spillover bias in diversity judgment. *Organizational Behavior and Human Decision Processes, 139*, 92–105. https://doi.org/10.1016/j.obhdp.2016.12.005

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature communications, 1*(10). https://doi.org/10.1038/s41467-019-09239-1. article 1258.

Goodale, B. M., Alt, N. P., Lick, D. J., & Johnson, K. L. (2018). Groups at a glance: Perceivers infer social belonging in a group based on perceptual summaries of sex ratio. *Journal of Experimental Psychology: General, 147*(11), 1660–1676. https://doi.org/10.1037/xge0000450

Guay, B., Marghetis, T., Wong, C., & Landy, D. (2025). Quirks of cognition explain why we dramatically overestimate the size of minority groups. *Proceedings of the National Academy of Sciences, 122(14), Article e2413064122*. https://doi.org/10.1073/pnas.2413064122

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision, 15(4), Article 16*. https://doi.org/10.1167/15.4.16

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

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

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 621–642. https://doi.org/10.1037/0278-7393.31.4.621

Herzog, S. M., & Hertwig, R. (2014). Harnessing the wisdom of the inner crowd. *Trends in Cognitive Sciences, 18*(10), 504–506. https://doi.org/10.1016/j.tics.2014.06.009

Inter-Parliamentary Union. (2025, January). Monthly ranking of women in national parliaments. Retrieved January 27, 2025, from https://data.ipu.org/women-ranking/?date_year=2025&date_month=01.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*(3), 1221–1247. https://https://doi.org/10.1016/j.cognition.2007.06.004.

Jayles, B., Sire, C., & Kurvers, R. H. (2021). Crowd control: Reducing individual estimation bias by sharing biased social information. *PLoS Computational Biology, 17*(11), Article e1009590. https://doi.org/10.1371/journal.pcbi.1009590

Ji, C. (2018). The influence of gender stereotypes on the perception of women in politics. *master's thesis, Sookmyung Women's University*.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251. https://doi.org/10.1037/h0034747

Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., … Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of the Royal Society Interface, 15*(141), 20180130. https://doi.org/10.1098/rsif.2018.0130

Kardosh, R., Sklar, A. Y., Goldstein, A., Pertzov, Y., & Hassin, R. R. (2022). Minority salience and the overestimation of individuals from minority groups in perception and memory. *Proceedings of the National Academy of Sciences, 119(12), Article e2116884119*. https://doi.org/10.1073/pnas.2116884119

Khaw, M. W., Kranton, R., & Huettel, S. (2021). Oversampling of minority categories drives misperceptions of group compositions. *Cognition, 214, Article 104756*. https://doi.org/10.1016/j.cognition.2021.104756

Landy, D., Guay, B., & Marghetis, T. (2018). Bias and ignorance in demographic perception. *Psychonomic Bulletin & Review, 25*, 1606–1618. https://doi.org/10.3758/s13423-017-1360-2

Lee, E., Karimi, F., Wagner, C., Jo, H. H., Strohmaier, M., & Galesic, M. (2019). Homophily and minority-group size explain perception biases in social networks. *Nature Human Behaviour, 3*(10), 1078–1087. https://doi.org/10.1038/s41562-019-0677-4

Lee, J., & Bang, H. (2003). Implicit representations of gender stereotypes. *Korean Journal of Cognitive Science, 14*(2), 37–48.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 551. https://doi.org/10.1037/0278-7393.4.6.551

Luo, A. X., & Zhou, G. (2018). Ensemble perception of facial attractiveness. *Journal of vision, 8*(18). https://doi.org/10.1167/18.8.7. article 7.

Mijalli, Y., Price, P. C., & Navarro, S. P. (2023). Spillover bias in social and nonsocial judgments of diversity and variability. *Psychonomic Bulletin & Review, 30*(5), 1829–1839. https://doi.org/10.3758/s13423-023-02259-5

Morgan, M. J., Watamaniuk, S. N. J., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research, 40*(17), 235–244. https://doi.org/10.1016/S0042-6989(99)00236-6

National Assembly of the Republic of Korea. (2024). 제*22*대 국회의원 검색 *[Search for the 22nd National Assembly members]*. 열린국회정보 [Open Assembly Information]. Retrieved February 10, 2025, from https://open.assembly.go.kr/portal/assm/search/memberSchPage.do.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General, 149*(2), 323–342. https://doi.org/10.3758/s13423-023-02259-5

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., … Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology, 114*(5), 766–785. https://doi.org/10.1037/pspi0000120

Ponce de Leon, R., Rifkin, J. R., & Larrick, R. P. (2022). "They're everywhere!": Symbolically threatening groups seem more pervasive than nonthreatening groups. *Psychological Science, 33*(6), 957–970. https://doi.org/10.1177/09567976211060009

Quillian, L. (1995). Prejudice as a response to perceived group threat: Population composition and anti-immigrant and racial prejudice in Europe. *American Sociological Review, 60*(4), 586–611. https://doi.org/10.2307/2096296

Schlueter, E., & Scheepers, P. (2010). The relationship between outgroup size and anti-outgroup attitudes: A theoretical synthesis and empirical test of group threat-and intergroup contact theory. *Social Science Research, 39*(2), 285–295. https://doi.org/10.1016/j.ssresearch.2009.07.006

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Sun, J., & Chong, S. C. (2020). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology: General, 149*(3), 550–563. https://doi.org/10.1037/xge0000667

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Yang, X., & Dunham, Y. (2019). Hard to disrupt: Categorization and enumeration by gender and race from mixed displays. *Journal of Experimental Social Psychology, 85, Article 103893*. https://doi.org/10.1016/j.jesp.2019.103893

Young, A., & Bruce, V. (2013). Face perception. *Psychology Press.*. https://doi.org/10.4324/9781003279426

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology, 98*(2), 175–189. https://doi.org/10.1037/a0017990