Quirks of cognition explain why we dramatically overestimate the size of minority groups

Brian Guay[a,1] (iD), Tyler Marghetis[b] (iD), Cara Wong[c] (iD), and David Landy[d,e]

Affiliations are included on p. 11.

Americans dramatically overestimate the size of African American, Latino, Muslim, Asian, Jewish, immigrant, and LGBTQ populations, leading to concerns about downstream racial attitudes and policy preferences. Such errors are common whenever the public is asked to estimate proportions relevant to political issues, from refugee crises and polarization to climate change and COVID-19. Researchers across the social sciences interpret these errors as evidence of widespread misinformation that is topic-specific and potentially harmful. Here, we show that researchers and journalists have misinterpreted the origins and meaning of these misestimates by overlooking systematic distortions introduced by the domain-general psychological processes involved in estimating proportions under uncertainty. In general, people systematically rescale estimates of proportions toward more central prior expectations, resulting in the consistent overestimation of smaller groups and underestimation of larger groups. We formalize this process and show that it explains much of the systematic error in estimates of demographic groups ($N = 100{,}170$ estimates from 22 countries). This domain-general account far outperforms longstanding group-specific explanations (e.g., biases toward specific groups). We find, moreover, that people make the same errors when estimating the size of racial, nonracial, and entirely nonpolitical groups, such as the proportion of Americans who have a valid passport or own a washing machine. Our results call for researchers, journalists, and pundits alike to reconsider how to interpret misperceptions about the demographic structure of society.

misperceptions | political psychology | cognition | race

Misperceptions about the size of demographic groups in society, particularly racial and ethnic minority groups, are among the most cited instances of citizen ignorance. Americans dramatically overestimate the size of the African American, Latino, Muslim, Asian, Jewish, and LGBTQ populations (1–6), and people around the world overestimate the size of their country's foreign-born population (7–9). On average, Americans estimated that immigrants made up 33% of the U.S. population in 2022, while the actual number was 15% (10). Past research has interpreted these errors as concerning evidence of political ignorance (7, 9, 11). When perceptions of group size serve as cognitive shortcuts in political decision-making, *mis*perceptions can lead to biased attitudes and behavior (7, 11, 12). For instance, overestimating the size of the immigrant population is associated with negative views of immigrants and support for restrictive immigration policies (7, 8), while overestimating the percentage of poor people who are Black is associated with greater opposition to welfare programs (13). Understanding the origin of these misperceptions is thus a crucial civic and scientific undertaking.

Two leading theories have emerged, both suggesting that overestimation is due to particular characteristics of the group being estimated. The first, *perceived threat*, posits that people overestimate the size of outgroups that they perceive as threatening (14–16). However, predictions from this theory are at odds with empirical work showing that members of minority groups also overestimate their *own* prevalence (even though they presumably find themselves less threatening) but underestimate the size of majority groups (who they presumably find more threatening) (4). The other theory, *social contact*, posits that interactions with members of a social group—either directly through personal contacts such as close friendships (2) and face-to-face interaction (17) or indirectly through media exposure (8)—influence misperceptions of that group's size, with greater levels of exposure leading to larger overestimates of the group's size (1, 2, 8, 17). Empirical support for this theory is also limited, and it too makes predictions that are out of line with empirical findings. For instance, past work shows that members of majority groups *underestimate* their own prevalence in society, yet social contact theory predicts that

Significance

Misperceptions of minority racial and ethnic group size are widespread. For example, Americans overestimate the immigrant, Black, and Hispanic populations by more than double. These errors are typically interpreted as evidence of group-specific biases. Here, we demonstrate that they result from psychological processes that people use to estimate proportions of any kind. We show that domain-general psychological processes explain widespread errors in demographic estimates, whether the estimated group is racial, nonracial, or entirely nonpolitical (e.g., home owners and passport holders). Our findings call for researchers, journalists, and pundits to adopt a psychologically realistic interpretation of these errors: while bias against minority groups is pervasive, it is not the primary source of error in estimates of their size.

[1]To whom correspondence may be addressed. Email: brianmguay@gmail.com.

https://doi.org/10.1073/pnas.2413064122   **1 of 11**

members of majority groups should *overestimate* the size of their own group, since people tend to socialize with people who are similar to themselves (17).

Here, we show that misperceptions about the size of demographic groups are far more reflective of the psychological process of estimating proportions than of factors related to the specific group whose size is being estimated. We directly test these existing theories against an alternative, rooted in the psychology of how individuals estimate proportions more generally. When people estimate proportions under uncertainty, they rescale their estimates toward a prior expectation; as a consequence, smaller proportions are systematically overestimated and larger proportions underestimated. We describe a psychologically realistic Uncertainty-Based Rescaling model of proportion estimation and show that this model explains much of the systematic errors in people's demographic estimates. Importantly, this alternative explanation is domain-general, meaning that it has nothing to do with characteristics of the specific group being estimated. Unlike existing theories, this account explains a wider range of misperceptions—not only why members of the majority overestimate the size of minority groups but also why members of minority groups overestimate their own prevalence, and why members of both minority and majority groups underestimate the size of majority groups.

Past work's focus on group-specific theories has overlooked the more general psychological mechanisms that can drive people to misestimate the size of any quantity, demographic or not, particularly when estimates are made under uncertainty. Consequently, researchers continue to misinterpret the misperceptions they measure on surveys using proportion estimates: overestimates of the size of minority groups are characteristic of uncertainty, not group-specific bias. Indeed, this explanation is relevant whenever researchers measure beliefs or attitudes by asking people to estimate proportions, a technique that is increasingly popular for measuring everything from perceptions of the risk of contracting COVID-19 (18) and refugees posing a terrorism threat (19), to how much others support climate change policies (20) and democratic values (21).

A Model of Uncertainty-Based Rescaling During Proportion Estimation

Explicit judgments, such as responses on a survey, are seldom direct expressions of respondents' underlying beliefs or attitudes. Since people are unlikely to maintain an explicit estimate of the proportional size of various demographic groups, for instance, they will need to generate such estimates on the spot when prompted (11). To generate an explicit response, individuals must integrate a variety of cues and considerations, and this process of constructing a response can introduce error (22). In the case of proportion estimates—proportion estimates in general, not just estimates of demographic proportions—these errors follow a recurring pattern: individuals overestimate the size of smaller proportions and underestimate larger ones (23, 24). Proportion estimates *in general* consistently follow an inverted s-shaped pattern, with the most dramatic misestimation occurring near the ends of the proportion scale. This pattern appears reliably across domains, whether when estimating the proportion of A's in a random sequence of letters (25), the number of dots on a page that are a specific color (26), the proportion of time intervals containing a specific sound (27), or the proportions represented by bar graphs and pie charts (28) (Fig. 1). Similar forms of misestimation error characterize economic decision-making (29), estimates of general numerical magnitudes (30), and perceptions of the relative frequency of lethal events (31).

A variety of mechanisms have been proposed to account for this general phenomenon (e.g., ref. 24). Here, we describe a model of uncertainty-based rescaling—a model of how individuals adjust or "rescale" their demographic estimates to reflect their uncertainty—that captures features shared by many of these accounts (33). The model formalizes two key features of domain-general numerical cognition.

First, explicit numerical estimates made under uncertainty are Bayesian, in the sense that they incorporate prior expectations about typical values. This is the basic insight behind Bayesian approaches to perception and cognition (34). Thus, our model formalizes the assumption that, when explicitly estimating a proportion, individuals rely not only on information specific to that proportion (e.g., the number of Hispanics living in the United States) but also on their prior expectations about the typical size of such proportions more generally (e.g., the typical size of racial and ethnic groups). As a result, estimates of extreme values should be shifted, or rescaled, toward the center of one's prior (Fig. 2*B*). Importantly, one's prior expectation about demographic proportions need not always be 0.50 (35). For instance, when estimating the size of a group one knows to be a minority, the range of possible estimates is constrained above by 0.50, because a minority group cannot, by definition, account for more than 50% of the population. With no information about
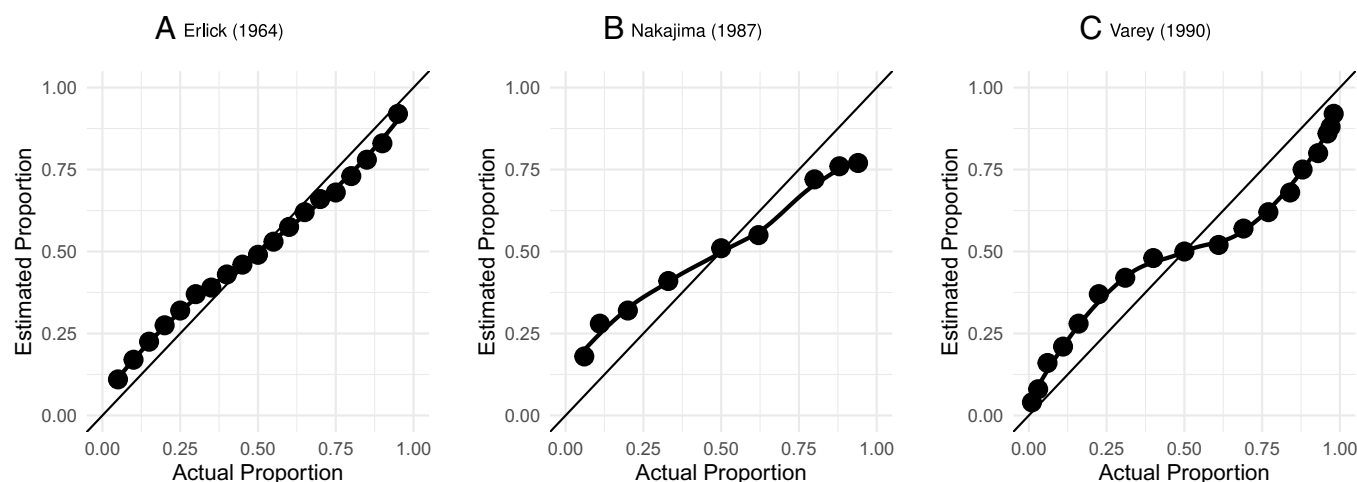
Fig. 1. Examples of systematic estimation error from previous studies of proportion estimation. From left to right: Estimates of (*A*) the proportion of letters in a sentence that are "A," of (*B*) time intervals containing a specific sound, and of (*C*) dots that are a certain color. Recreated from data plotted in ref. 32.

a group other than that it is a minority, a reasonable prior will be less than 0.50. Likewise, because the size of majority groups is naturally greater than 0.5, plausible priors will be constrained to values between 0.5 and 1.

The influence of priors should depend on one's uncertainty: when individuals are less certain about the size of a particular demographic group, they should rely more on their prior expectations about group sizes in general, and should thus increasingly shift or rescale their estimates toward their prior. Thus, from a Bayesian perspective, an estimate of the size of a demographic group reported on a survey should ideally reflect a combination of one's knowledge about the size of that group and one's prior expectations for demographic group sizes in general, with the relative contribution of each weighted by one's uncertainty about the former (see *Materials and Methods* for formal details).

The model's second assumption is that the mental processing of proportions is nonlinear, and in particular that proportions are mentally processed as log-odds (24, 33). The nonlinear processing of numerical quantities has been hypothesized for monetary value since the 1700s and is a central tenet of expected utility theory, prospect theory, and other modern economic models of human decision-making (29, 36). This nonlinear, log-like processing generalizes to many other contexts, including the processing of sound (37) and numbers (38). In each context, a small change in a small quantity feels more salient than the same change in a larger quantity: for instance, it is easy to distinguish a 5 pound weight from a 10 pound weight, but a 105 pound weight may feel indistinguishable from one that is 110 pounds. When people estimate the size of a demographic group as a proportion of the entire population, therefore, their response likely reflects the cognitive processing of representations on a log-odds scale (see Fig. 2C and *Materials and Methods*). To be clear, the claim is not that people are aware of this format or perform this calculation consciously, but rather that the cognitive processing of proportions operates with representations on a log-odds scale, as documented in past research on numerical cognition.

Combining these assumptions gives us a first-principles, psychologically realistic model of how an individual should incorporate uncertainty into their explicit estimates of demographic group sizes. On a log-odds scale, one's estimate should reflect both one's information about the particular group's size and one's prior expectations about the size of demographic groups in general:

$$\Psi_{group} = \gamma r_{group} + (1 - \gamma) r_{prior} \qquad [1]$$

Here, Ψ_{group} is the explicit estimate of the size of a particular group that an individual should make, in log-odds; r_{group} is an uncertain estimate of the group's size based on current information, in log-odds; r_{prior} is the mean of one's prior expectations for the size of demographic groups in general, in log-odds; and γ captures the relative certainty in one's own information versus in one's prior. (In *Materials and Methods*, we show how to express Eq. **1** in terms of probabilities rather than log-odds, which is the model we use in our empirical analyses.)

If the person estimating has unbiased but uncertain beliefs about the actual size of a group, then r_{group} will be the group's actual size, inferred with some uncertainty. This belief could reflect information from a variety of sources, including personal experience in the world, word-of-mouth, popular discussions of demographic trends, and more. However, even if this belief were perfectly accurate, their survey response (Ψ_{group}) will not be equal to r_{group}, as their prior will exert significant influence. In this case, Eq. **1** is the optimal Bayes estimator of the group's size, given that uncertainty. In other words, the model captures how an uncertain person *should* respond on surveys, even when their underlying knowledge is totally unbiased.

The question, then, is whether this psychologically realistic model can explain widespread misperceptions of the size of demographic groups. Attempts to account for these errors in terms of domain-general psychological processes have been limited by the use of aggregated demographic estimates (33), since inverted s-shaped error patterns can arise from averaging, even if estimates by individuals are not s-shaped (*SI Appendix*, *Fig. S2* in section 6.2). Moreover, past work has considered only a limited range of demographic misperceptions, omitting many of the most politically relevant misperceptions, such as estimates of the size of racial groups. More importantly, no work to date has compared domain-general psychological processes
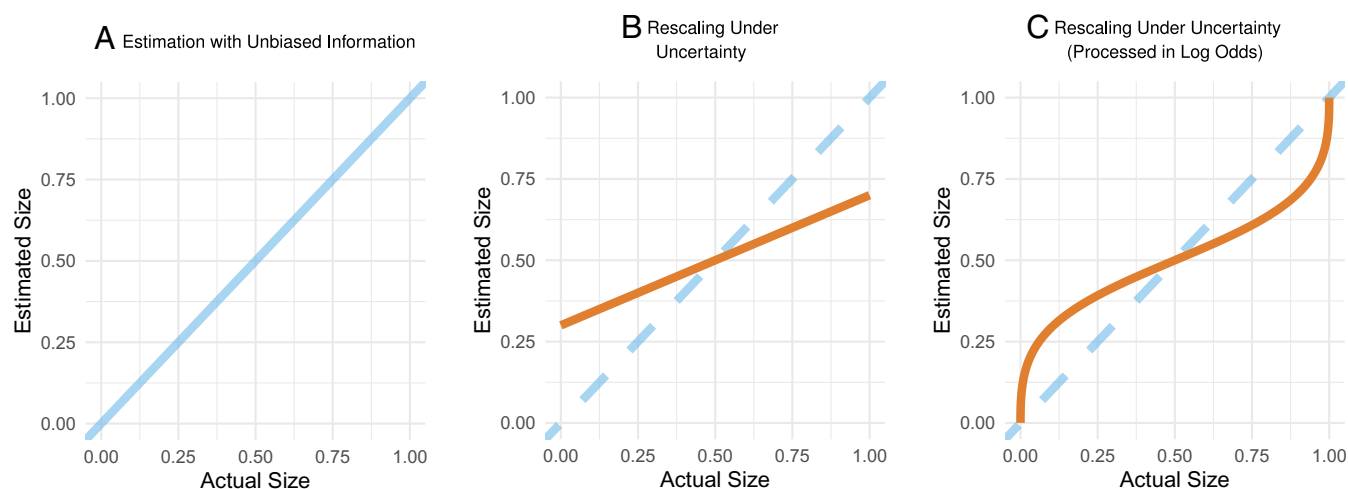
Fig. 2. Illustration of a domain-general model of Uncertainty-Based Rescaling during proportion estimation. (*A*) The estimates of a perfectly informed and completely certain individual (solid blue line) of the proportional size of demographic groups; in the absence of uncertainty in one's information, the estimate is equal to the actual size of the proportion. (*B*) Visual illustration of how, under uncertainty, an individual might shift or "rescale" their proportion estimates toward the center of their prior (e.g., 50%). This shifts estimates of small proportions upward and shifts estimates of large proportions downward (34). This simple model, however, does not account for the way proportions are processed psychologically as log-odds. (*C*) When individuals rescale their estimates while processing proportions on a log-odds scale, their proportion estimates exhibit an s-shaped nonlinearity. The solid orange line shows the predictions of the Uncertainty-Based Rescaling model, which combines the idea of rescaling-under-uncertainty with the psychologically realistic assumption that the human mind processes proportions on a log-odds scale; see *Materials and Methods* for formal details.

to long-standing theories of perceived threat and social contact, which continue to be the primary explanations of demographic misperceptions.

Uncertainty-Based Rescaling Explains a Wide Variety of Demographic Misperceptions

We begin by applying this model of Uncertainty-Based Rescaling to the largest collection of estimates of the size of demographic groups to date, containing a total of 100,170 estimates. These estimates come from 36,130 respondents in 22 countries over a three-decade period. 70% of these estimates come from existing surveys, including those run on large national probability samples—the 1991 American National Election Study Pilot (ANES), 2000 General Social Survey (GSS), and the 2002 European Social Survey (ESS)—and surveys from four previous studies (9, 39–41).

We begin by comparing 63 mean estimates from these surveys to their actual values. Fig. 3*A* shows the pattern of misestimation discussed above: the sizes of all 59 minority groups (i.e., those
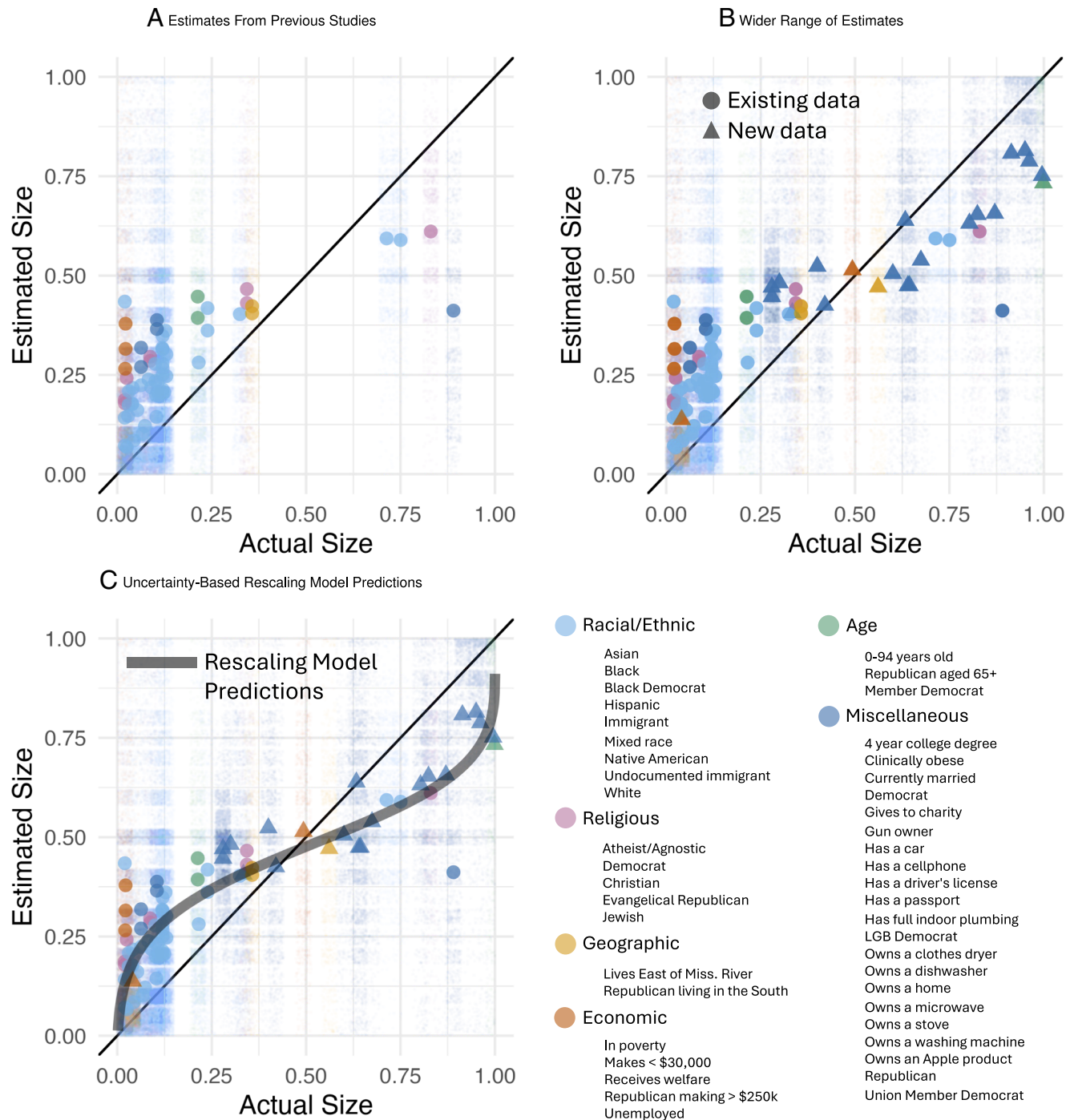
Fig. 3. Demographic group size estimates exhibit a systematic S-shaped pattern of over- and underestimation. Estimates of groups' sizes (vertical axis) are plotted against their actual sizes (horizontal axis). Smaller transparent points represent individual estimates (*N* = 100,170); larger solid points represent means for each estimated group. (*A*) Previous surveys: the 1991 ANES, 2000 GSS, 2002 ESS, and four published studies. (*B*) Additional estimates from two original surveys asking about a wider range of demographic groups. (*C*) Predictions from the Uncertainty-Based Rescaling model specified in Eq. **1**. The model captures the S-shaped pattern of errors across the full range of actual sizes. Mean estimates and actual sizes for all estimated groups are in *SI Appendix*, section 3.

comprising <50% of the population) are overestimated and the sizes of all four majority groups are underestimated.

Two limitations of existing survey data make it difficult to discern whether estimates of demographic proportions follow the inverted s-shaped pattern of over-under estimation described above. First, past work has focused primarily on relatively small minority groups, obscuring overarching patterns that would suggest a domain-general explanation. Second, past work has focused primarily on estimates of minority racial, ethnic, and religious groups, where perceived threat and social contact can often account for the direction, if not the magnitude, of misestimation. Observing systematic errors in demographic groups not influenced by perceived threat and social contact (e.g., the percent of Americans who hold a valid passport) would suggest a more general underlying cause.

We thus conducted two new surveys, which contribute the remaining 30% of estimates in the full dataset analyzed here (*Materials and Methods*). First, we asked 1,262 U.S. adults recruited from Lucid to estimate the size of 19 nonracial groups that cannot be easily explained by perceived threat and social contact, such as the percentage of U.S. adults who are younger than 95, clinically obese, earn less than $30,000 annually, and who possess common objects such as a cell phone, microwave, stove, washing machine, clothes dryer, dishwasher, car, driver's license, and passport. Second, we asked 2,487 US adults from Cloud Research Connect to estimate the size of three demographic groups: the percentage of adults in the United States who are Republican (0.28), Democrat (0.28), and are unemployed (0.04).

When we combine estimates from past studies with our two original surveys in Fig. 3B, the familiar inverted s-shaped pattern characteristic of proportion estimation (Fig. 1) is evident. On average, respondents underestimate the size of majority groups and overestimate the size of minority groups. Indeed, all 67 minority groups are overestimated while 17 of the 18 majority groups are underestimated (the remaining majority group, the percentage of Americans who have a car, is overestimated by less than 1 percentage point). Moreover, the qualitative pattern of errors observed in estimates of racial and nonracial groups is strikingly similar, suggesting that the errors are due to a domain-general process rather than processes that are specific to the perception of racial groups. In *SI Appendix*, section 6.4, we show that ad hoc demographic groups such as passport-holders exhibit the same error pattern as racial, ethnic, and religious groups.

The Uncertainty-Based Rescaling model captures this pattern of over-under estimation (Fig. 3C). We model all respondents' estimates with the two-parameter model given in Eq. 1 (*Materials and Methods*). Model predictions are represented by the solid gray line in Fig. 3C. Across racial and nonracial groups, the model accounts systematically for errors in estimates of the groups' sizes. This two-parameter Uncertainty-Based Rescaling model is thus able to account for estimation errors across a wide variety of groups without any information about the particular groups being estimated besides their actual size. In other words, domain-general psychological processes alone explains much of the error in demographic estimates, without invoking any group-specific considerations such as threat or contact.

Indeed, as reported in *SI Appendix*, section 6.4, rescaling was even more pronounced for estimates of groups that theories of perceived threat and social contact cannot explain—groups unrelated to race, ethnicity, or religion. This follows naturally from our account of uncertainty-based rescaling, since uncertainty is presumably higher for atypical or ad hoc demographic categories such as people who own Apple products or people who have a passport.

Comparison to Existing Theories of Demographic Misperception

Next, we compare the domain-general Uncertainty-Based Rescaling model to existing group-specific accounts of perceived threat and social contact. We use data from the 2000 GSS, which asked a probability sample of 1,398 U.S. adults to estimate the share of the population that is Black, Hispanic, Asian, and White. Since theories of perceived threat and social contact posit that demographic misperceptions are driven largely by everyday, personal interactions and observation, we might expect these theories to be especially successful at explaining misperceptions of local rather than national prevalence.

The GSS data are uniquely suited to a direct comparison of domain-general and group-specific theories of demographic misperception. Respondents were asked to report how threatening they perceive each group to be and how much close, personal contact they have with each group (*Materials and Methods*). Additionally, respondents not only estimated the size of demographic groups in the country but also in their local counties. The local prevalence of racial groups varies widely in the United States (for instance, the actual county-level Black population in our sample ranges from less than 1 to 57%), and according to the Uncertainty-Based Rescaling model, this variation in actual prevalence should systematically explain the direction and magnitude of estimation errors. The GSS thus offers variation in both the actual size of each racial group (invoked by the Uncertainty-Based Rescaling model) and in individual-level group-specific threat and contact (invoked by theories of threat and contact), allowing us to test these theories directly.

An additional benefit of the GSS data is that, unlike most surveys, the GSS asks respondents to estimate not only the size of other racial groups (i.e., out-groups) but also the size of their own racial group (i.e., in-groups). According to theories of social contact, people should *over*-estimate the size of their own group, regardless of the group's size, because social networks are homophilic (i.e., people tend to interact with others who resemble themselves). Theories of perceived threat, on the other hand, do not typically address in-group estimation—but, if anything, they predict that minority groups should underestimate their own prevalence, since people are presumably less threatened by their own group. According to theories of social contact and perceived threat, therefore, errors in in-group estimates should go in the opposite direction from errors in out-group estimates. Our Uncertainty-Based Rescaling model, by contrast, predicts that people should exhibit the same inverted s-shaped pattern of errors whether they are judging the size of their own group or another: *over*-estimate if it is a smaller group, *under*-estimate if it is a larger group.

Fig. 4A plots mean estimates from the GSS data against their actual sizes. We find the same over-under estimation pattern observed in Fig. 3: smaller groups are systematically overestimated while larger groups are underestimated. Panel B features the same data, but the y axis is average estimation *error* (Fig. 4), calculated by subtracting the actual size of each group from each estimate:

$$\text{estimation error} = \text{estimated size of group} \\ - \text{actual size of group}. \quad [2]$$

Whereas in the previous analyses, we have focused on estimates, we focus on *estimation error* from here forward because theories

A Estimates
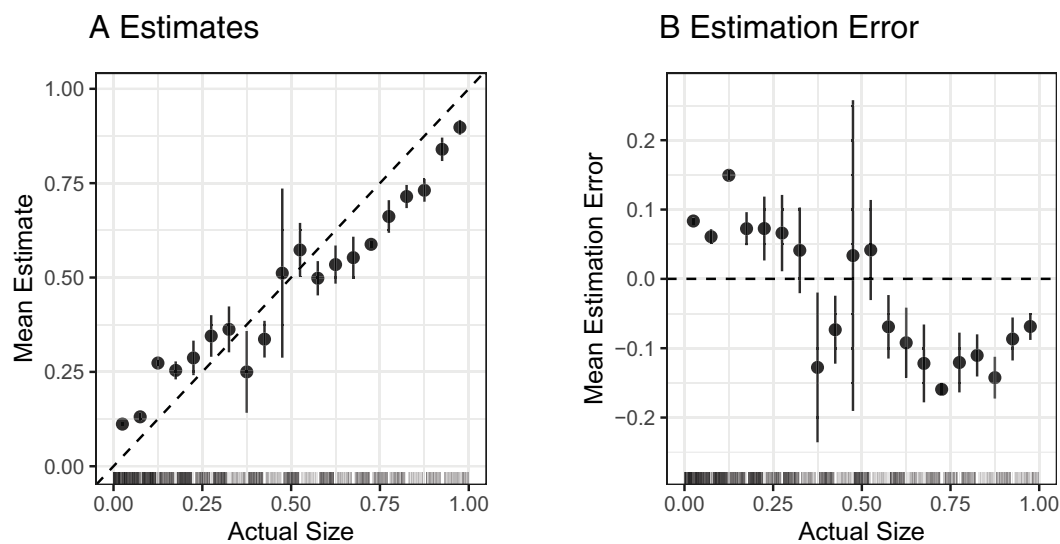
B Estimation Error

Fig. 4. Estimates of the local and national prevalence of racial and ethnic groups in the United States followed the same S-shaped pattern of errors. Respondents separately estimated the percent of the United States and their local county that is Black, Hispanic, Asian, and White. (*A*) The average estimate (Y axis) for local groups of different actual sizes (X axis). To prevent overplotting, we average group sizes in 5% bins (e.g., groups that make up less than 5% of the population; groups that make up between 5 and 10% of the population; etc.). Vertical lines represent 95% CI around each mean. The rugs on the X axes show the (jittered) distribution of actual sizes. We observe the same S-shaped pattern of over- and underestimation. (*B*) Focusing on estimation *error* highlights the systematic pattern of over- and underestimation. We calculated estimation error by subtracting the actual size of a local group from its estimated size (Eq. **2**).

of perceived threat and social contact relate to the direction of error, not raw estimates.

We begin by applying the Uncertainty-Based Rescaling model to four mutually exclusive subsets of the data: respondents' estimates of the size of local out-groups, local in-groups, national out-groups, and national in-groups (for modeling details, see *SI Appendix*, section 1.1; and for regression tables, see *SI Appendix*, section 5). As seen in Fig. 5, we observe the familiar pattern of systematic overestimation for small populations (i.e., positive estimation error) and underestimation for large populations (i.e., negative estimation error) for estimates of both out-groups and in-groups at both the local and national levels. The similarity in this pattern across in-groups and out-groups is predicted by the Uncertainty-Based Rescaling model, but, as discussed previously, runs counter to theories of perceived threat and social contact. Indeed, for each subset of the data, the Uncertainty-Based Rescaling model fits the pattern of average errors made by respondents closely (orange lines in Fig. 5). While the Uncertainty-Based Rescaling model captures the overall, qualitative phenomenon that smaller groups are overestimated while larger groups are underestimated, it also closely predicts the variation in errors among smaller groups. This is illustrated by the *Inset* in Fig. 5*A*, which zooms in on groups that comprise less than 15% of the population, which make up two-thirds of estimated local out-groups in our sample.

Since this pattern is so reliable, our model can account for respondents' estimates of a wide range of group sizes with only two parameters. For instance, estimates of local out-groups and of local in-groups show the qualitatively similar s-shaped pattern of over- and underestimation (Fig. 5).

Separate models of out-group and in-group estimates, moreover, revealed interesting differences in the process of uncertainty-based rescaling (*SI Appendix*, section 5). These differences make sense in light of the types of judgments and who was making them. Because the GSS is a probability sample of U.S. adults, *estimates of out-groups* consist mostly of estimates made by people who belong to the White majority judging the size of minority groups to which they do not belong. By contrast, *estimates of*

in-groups consist mostly of estimates made by people who belong to the White majority judging the size of their own majority group. If people think that a group is a minority, then their prior should reflect that the group will, by definition, make up less than half the population; likewise, if people think a group is a majority, their prior should reflect that the group will make up more than half the population. Moreover, people are presumably more certain in their knowledge of their own group, so they should engage in less uncertainty-based rescaling in estimates of their own group. This is, indeed, what we found (*SI Appendix*, section 5). There is less rescaling for estimates of in-groups (0.35 for local in-groups, 0.30 for national in-groups) than for estimates of out-groups (0.44 for local out-groups, 0.42 for national out-groups).

Finally, we examine whether theories of perceived threat and contact explain any of the error in demographic estimation. We model respondents' estimates as a function of their group-specific perceived threat and group-specific social contact. Both measures are coded to reflect relative differences; for instance, how much more threatening a respondent finds a group (e.g., Hispanics) relative to the other groups (African Americans, Asian-Americans, and Whites). Since theories of perceived threat are typically invoked to explain out-group estimates, we fit this model to estimates of the size of racial out-groups, at both the local and national levels.

As seen in Fig. 6, variation in perceived threat or social contact account for only a small fraction of the error. Greater perceived threat, for instance, is associated with only small amounts of overestimation. At the local level, an increase of one SD (standard deviation) in perceived threat was associated with estimates that are 1.3 percentage points larger; at the national level, with estimates that are 1.9 percentage points larger (Fig. 7). While this is a statistically significant increase, the influence of perceived threat is small relative to the large estimation errors they seek to explain. For instance, the mean estimation error for the size of the African American population at the national level is 19 percentage points (*SI Appendix*, section 3), an order of magnitude larger than the association with
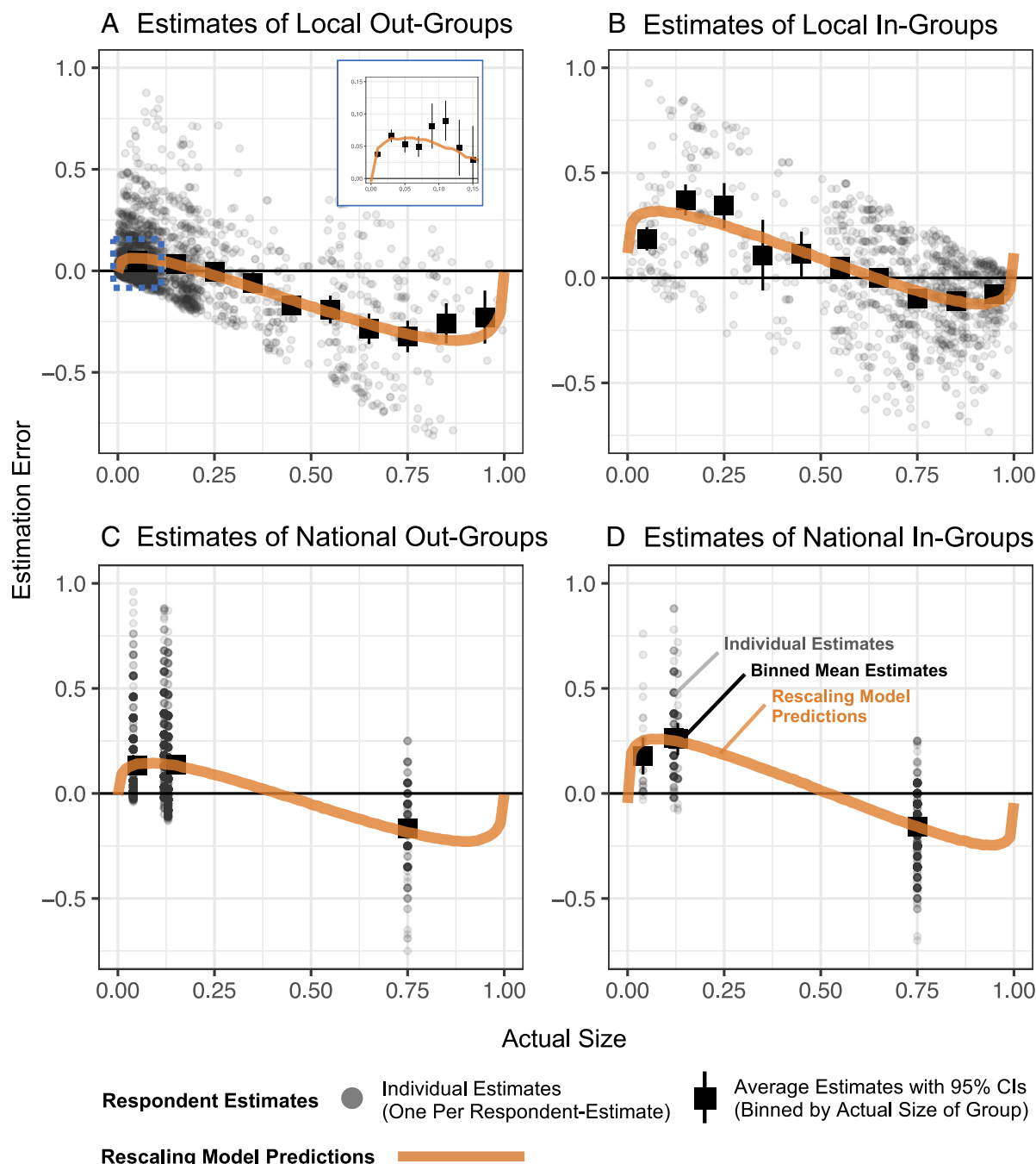
Fig. 5. Misestimation of the local (*A* and *B*) and national (*C* and *D*) prevalence of ethnic and racial groups followed the S-shaped pattern predicted by the Uncertainty-Based Rescaling model. All panels show estimation *error* (i.e., the difference between the estimate and the group size), plotted against actual group size. Predictions from the Uncertainty-Based Rescaling model are overlaid in orange (Eq. **1**). Individual estimation errors are represented as gray points (jittered). Binned mean estimation errors are represented as larger black squares with 95% vertical CI. The *Inset* in Panel *A* zooms in on estimates of smaller groups (those comprising less than 15% of the population), which account for two-thirds of the local out-groups that respondents estimated. Full model results, including model fit statistics for each model, are reported in *SI Appendix*, section 5.

perceived threat. Variation in perceived threat, therefore, might explain some of the overestimation by individuals with extreme views (e.g., the few, extreme individuals who rated the perceived threat of minority groups as seven SD higher than average; see *SI Appendix*, Fig. S7), but it does little to explain the large estimation errors that are observed for most respondents, even those with lower-than-average perceptions of perceived threat.

Moreover, contrary to predictions of social contract theories, greater social contact is associated with *lower* estimates, though this relationship is small (Fig. 6, *Bottom*). A one SD increase in

social contact is associated with a 1.6 percentage point *decrease* in the group size estimate at the local level and a 0.7 percentage point decrease at the national level (Fig. 7). Once again, note that these associations are an order of magnitude smaller than the average estimation errors people make.

One possibility is that much of the error in demographic estimates is due to rescaling, as captured by the Uncertainty-Based Rescaling model, but that the remaining unexplained error is due to perceived threat or contact. To test this possibility, we again model out-group estimates as a function of perceived threat
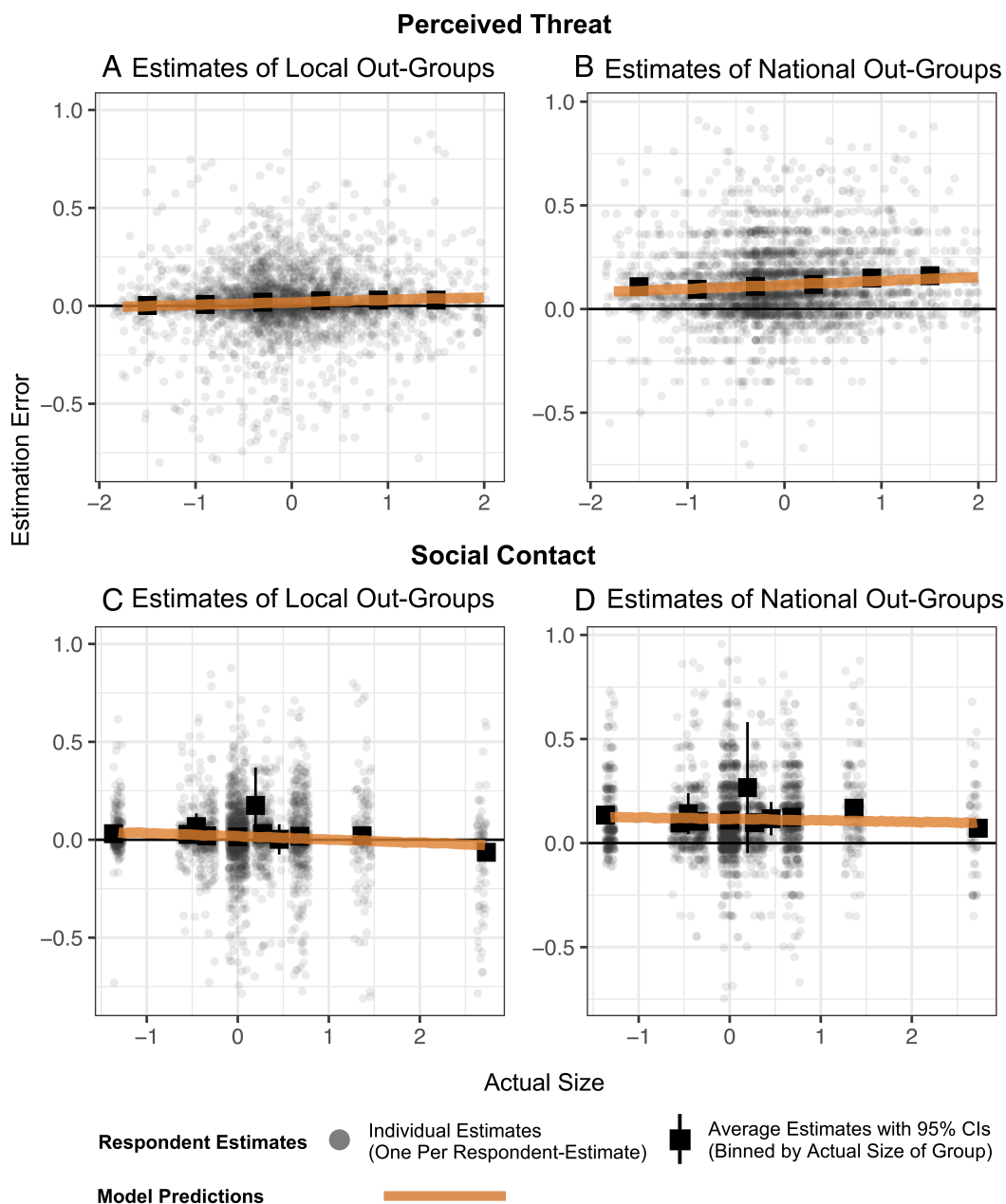
Fig. 6. Variation in perceived threat and contact explained little of the variation in estimation error. Individual estimation errors are represented as jittered gray points, plotted against perceived threat (*A* and *B*) and contact (*C* and *D*). Binned mean estimation errors are represented as larger black squares with 95% vertical CI. Predictions from the perceived threat and contact models are overlaid as orange lines. Full model results, including model fit statistics, are reported in *SI Appendix*, section 5. Plots show the central 95% of values, excluding extreme outliers. See *SI Appendix*, Fig. S7 for plots that include extreme values.

and contact, but this time also accounting for Uncertainty-Based Rescaling (see *Materials and Methods* and *SI Appendix*, section 1.1 for details).* The black points in Fig. 7 report parameter estimates for perceived threat and contact in models that also account Uncertainty-Based Rescaling. After accounting for rescaling, the positive association between perceived threat and overestimation remains significant though smaller. Interestingly, accounting for rescaling results in a parameter estimate for social contact that is in the direction predicted by contact theory; this is consistent with our proposal that group-specific factors may be responsible

for residual error that remains after accounting for domain-general rescaling. Notably, the size of these associations with threat and contact remains substantively small. A one SD increase in social contact is associated with a 1.0 and 1.5 percentage point increase in estimates for local and national estimates, respectively, while a one SD increase in perceived threat is associated with increases of 0.7 and 1.0 percentage points. In sum, after accounting for rescaling, the relationships between estimation error and both perceived threat and contact are in the predicted directions, although they only account for small amounts of error.

To directly compare accounts based on rescaling, perceived threat, and contact, we report fit statistics for all models in *SI Appendix*, section 5: models that predict estimation error with 1) only the demographic control variables that are included in

---

*Note that this is a particularly conservative test of the Uncertainty-Based Rescaling model, since the Uncertainty-Based Rescaling model does not account for any individual differences or group-specific factors; for simplicity, we assume that all respondents engage in rescaling using the same prior and uncertainty. Perceived threat and contact, by contrast, are measured at the level of individual respondents and racial groups.
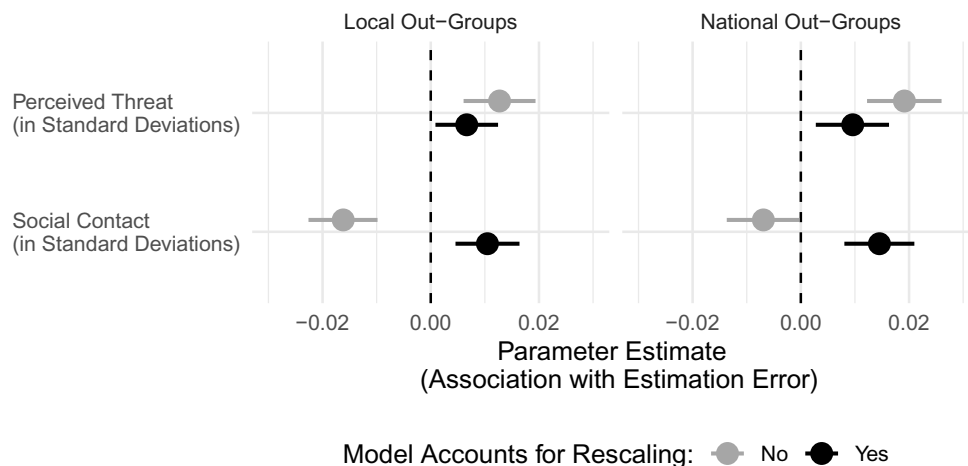
Fig. 7. Parameter estimates for models of group-size estimation that included perceived threat and contact, with and without accounting for rescaling. Note that an increase of one SD in perceived threat or social contact was associated with only about one percentage point of overestimation. The association between social contact and estimation error, moreover, was only in the predicted direction after accounting for rescaling (black circles). Horizontal lines represent 95% credible intervals.

all models (e.g., respondent age, gender, education), 2) models that include perceived threat and contact, 3) models include rescaling, and 4) models that include rescaling, perceived threat, and contact. Across all subsets of the data, models that account for rescaling substantially minimize prediction error compared to those that do not. For instance, accounting for rescaling in estimates of local out-groups increases the leave-one-out Bayesian R^2 by a factor of 3.6 (from 0.091 in the controls-only model to 0.329 in the rescaling model). In contrast, accounting for perceived threat and contact results in only a minuscule improvement in model fit (Bayesian R^2 from 0.091 in the controls-only model to 0.101 in the rescaling model; a factor increase of 1.1). Likewise, adding perceived threat and contact to a model accounting for rescaling does not result in any improvement in model fit. [These results are qualitatively unchanged when we account for random variation between estimated groups in the sources of estimation error (*SI Appendix,* Table S11), except associations with perceived threat and social contact are no longer credibly different from zero.]

Discussion

We examined whether widespread demographic misperceptions are explained by the psychological processes by which people perceive and estimate numerical information more broadly. Our findings demonstrate that a minimal model of Uncertainty-Based Rescaling during proportion estimation—in which individuals rescale their explicit estimates toward prior expectations—accounts for much of the error in demographic estimates. Demographic estimates followed the same inverted s-shaped pattern of systematic error that is characteristic of proportion estimation in nondemographic domains. Moreover, we found that errors in estimates of hot-topic groups (e.g., undocumented immigrants, gay Democrats) looked no different from errors in estimates of mundane demographic groups (e.g., Apple product owners, passport holders). In contrast, we found little empirical support for theories of perceived threat and social contact, and where there was support, these theories explained only a small fraction respondents' estimation errors.

Our findings have implications for how to interpret demographic misperceptions reported on surveys. Previous interpretations have attributed demographic misperceptions to underlying bias or misinformation about the size of particular groups, driven by differential social contact with minority groups or perceptions of certain groups as threatening (1, 14–16). Here, we demonstrate that these estimation errors are quite general, appearing for a wide range of demographic groups, and are explained as the product of a domain-general cognitive model of how people estimate proportions under uncertainty. We show that the errors in demographic estimates that have been observed and widely publicized are, contrary to previous assumptions, precisely what we would expect to see when people have unbiased underlying information, but adjust their estimates toward a reasonable prior expectation due to uncertainty.

Our findings also have implications for how misperceptions about *nondemographic* quantities are interpreted. Social scientists are often interested in people's perceptions of quantities relating to the economy, such as the proportion of government spending dedicated to welfare, the unemployment rate, and inflation (11, 42, 43). Other studies have documented errors in the public's perception of the frequency of lethal events (31), the human and financial cost of armed conflict (44), the likelihood of contracting COVID-19 (18), and the proportion of the federal budget spent on foreign aid (45). More recent work has documented, alarmingly, that both elected representatives and citizens misestimate public opinion, such as support for climate legislation, gun control, and abortion policy (20, 21), as well as others' beliefs more generally (46). Together, these findings have been interpreted as worrying evidence of bias or ignorance among political elites and the voting public alike. For instance, people overestimate the frequency of infrequent causes of death (e.g., botulism) but underestimate the frequency of frequent causes (e.g., heart disease) (31)—the familiar s-shaped pattern of errors. Past accounts of these errors have included both domain-general heuristics such as anchoring-and-adjustment (i.e., generating estimates by adjusting away from a representative value) and item-specific features such as biased newspaper coverage or memorability. Our results suggest that, when explaining errors in such estimates, we should account for topic-neutral psychological processes (such as rescaling under uncertainty) before invoking topic-specific bias or ignorance. A model like the one used here can be used to account for the influence of systematic, topic-neutral processes, thus allowing topic- or item-specific explanations to focus on the model's residuals.

Our results may also explain a pattern of findings in the growing body of research that attempts to change attitudes (e.g., toward immigration policy) by correcting numeric misperceptions (e.g., of the size of the current immigrant population). A recurring pattern across studies is that offering correct information often succeeds in reducing errors in explicit estimates but fails to change downstream attitudes (6, 9, 11, 19, 47). Our account offers an explanation of these failures to change downstream attitudes: errors in explicit estimates, while sometimes quite large, are often the product of the domain-general processes involved in generating explicit estimates, not group-specific misinformation or bias. Thus, errors in explicit estimates are a poor guide to underlying group-specific ignorance in need of correction, unless we first account for errors introduced by domain-general processes such as rescaling. Indeed, one of the key implications of Bayesian models of estimation is that people can make systematic errors in estimation even when their internal perceptions of the world are unbiased.

This is not to undermine the existence of bias and even animus against immigrants, the LGBTQ community, and other marginalized communities. But such bias is not responsible for most of the errors that people make in estimating the demographic structure of their communities. Efforts to reduce animus toward marginalized communities, therefore, are misplaced if they focus on correcting demographic misestimation and are best directed elsewhere.

Though perceived threat and social contact explain little of the error in people's estimates, this does not mean that group-specific information in general plays no role in the formation of people's beliefs. Since people's estimates are related systematically to groups' actual sizes (Fig. 3), people must be using some source of group-specific information to form their underlying sense of any particular group's size. But our findings suggest that much of the error in people's explicit estimates of the structure of society, including its demographic structure, is rooted in the broader psychology of how quantities are estimated. That is not to say that group-specific factors do not play some role in these errors. For example, for the 2000 GSS, the actual sizes of the Black (12%) and Hispanic (13%) communities were similar, but respondents overestimated size of the Black (32%) population considerably more than the Hispanic (25%) population. Our findings suggest that when seeking to explain misestimates of the size of a particular group, future work should first account for any error that appears systematically across estimates of all groups before invoking factors specific to a particular group. Similar reasoning applies to other psychological phenomena, such as the formation of group stereotypes, which may reflect both domain-general cognitive processes and topic- or group-specific factors (48).

Our central finding—that much of the variation in demographic estimates is due to rescaling under uncertainty, not group-specific biases and attitudes—helps to explain why, despite the magnitude of these errors, their correlations with other aspects of political belief and behavior have been so small (7, 13, 39). By first accounting for errors due to the psychology of estimation in general, future work will better identify and understand citizens' beliefs—including their inaccurate beliefs—that are central to their participation in society.

Materials and Methods

Model of Uncertainty-Based Rescaling. The Uncertainty-Based Rescaling model of demographic proportion estimation assumes that implicit psychological processing of a proportion, p, operates with representations on a log-odds scale (24, 33):

$$r_p = \log\left(\frac{p}{1-p}\right). \qquad [3]$$

We assume that respondents have unbiased but uncertain knowledge about each group, formalized as a Gaussian distribution over log-odds that is centered on r_{actual}, the actual group size, but uncertain (i.e., has variance σ^2):

$$r_p \sim \mathcal{N}(r_{actual}, \sigma^2). \qquad [4]$$

We assume the estimator has a prior r_0 centered on some value r_{prior} that captures their prior expectations for the class of demographic groups:

$$r_0 \sim \mathcal{N}(r_{prior}, \tau^2). \qquad [5]$$

In this scenario, the Bayes estimator, r_{bayes}, that incorporates both uncertain knowledge about a particular group and prior expectations for group sizes in general is

$$r_{bayes} = \gamma \cdot r_{actual} + (1 - \gamma) \cdot r_{prior}. \qquad [6]$$

The first term captures how much the estimate reflects one's own knowledge of the size of the particular group, and the second term captures how much one rescales back toward the prior. The weighting parameter (γ) reflects the respondent's relative certainty in their group-specific knowledge versus in the prior; when the variances, σ^2 and τ^2 respectively, are known, then $\gamma = \frac{\tau^2}{\sigma^2 + \tau^2}$. These two terms combine to give the Bayes-ideal estimate under uncertainty [technically, the *minimum mean squared error* Bayes estimator (49)]. This psychologically realistic model formalizes the scenario where a respondent has uncertain but unbiased knowledge and must account for that uncertainty when making explicit estimates.

To generate the estimate on a probability scale, rather than on a log-odds scale, we combine Eqs. **3** and **6**. For notational simplicity, we represent the mean of the prior in odds, denoted by δ:

$$\Psi(p_{actual}) = \frac{\delta^{(1-\gamma)} p_{actual}^\gamma}{\delta^{(1-\gamma)} p_{actual}^\gamma + (1 - p_{actual})^\gamma}. \qquad [7]$$

Here, Ψ is the Bayes-ideal proportion estimate under uncertainty, p_{actual} is the actual group size as a proportion, and γ captures uncertainty.

Demographic Group Size Estimates from Existing and New Surveys. We aggregated estimates of demographic group sizes (Fig. 3) from multiple sources: large national probability samples (the 1991 ANES, 2000 GSS, and the 2002 ESS); four published studies (9, 39–41); and two original online studies that we ran to address limitations of existing data. The first original study was conducted in 2018 with a nonprobability sample of $N = 1,262$ US adults recruited by Lucid, a platform that connects researchers to a pool of online research participants drawn from over 250 respondent providers. The Lucid study received approval from Duke University's IRB (2019-0140). The second original study was conducted in 2025 with a nonprobability sample of $N = 2,487$ US adults recruited from Cloud Research Connect. The Cloud Research study received approval from American University's IRB (E-5539). We obtained informed consent from respondents and they were compensated for their time. See *SI Appendix*, section 1 for details on the Lucid and Cloud Research studies.

Analysis Approach. We fit Eq. **7** to demographic estimates to infer how an estimator would have generated those estimates if they were engaged in Uncertainty-Based Rescaling. All models were fit using the *brms* package in R. See *SI Appendix*, section 1 for modeling details.

Data, Materials, and Software Availability. Data, code, and instructions for reproducing the analyses are available here: https://osf.io/cuvgw/ (50). The portion of our analysis that uses the 2000 General Social Survey relies on both publicly available data (on demographic estimates at the national level) and

restricted data (on demographic estimates at the county level). Instructions for obtaining these restricted data, along with code to reproduce the analysis of these data, are also included in the replication file.

Edderic Ugaddan, Adam Berinsky, David Rand, Ben Tappin, Chloe Whittenberg, Adam Bear, Hause Lin, Antonio Arechar, and Tom Costello.

Author affiliations: [a]Department of Political Science, Stony Brook University, Stony Brook, NY 11794; [b]Department of Cognitive and Information Sciences, University of California, Merced, CA 95343; [c]Department of Political Science, University of Illinois, Urbana-Champaign, Urbana, IL 61801; [d]Netflix, Los Gatos, CA 95032; and [e]Department of Psychological and Brain Studies, Indiana University, Bloomington IN 47405-7007

1. R. Nadeau, R. G. Niemi, J. Levine, Innumeracy about minority populations. *Public Opin. Q.* **57**, 332–347 (1993).
2. L. Sigelman, R. G. Niemi, Innumeracy about minority populations: African Americans and Whites compared. *Public Opin. Q.* **65**, 86–94 (2001).
3. R. Alba, R. G. Rumbaut, K. Marotz, A distorted nation: Perceptions of racial/ethnic group sizes and attitudes toward immigrants and other minorities. *Soc. Forces* **84**, 901–919 (2005).
4. C. J. Wong, "Little" and "Big" pictures in our heads: Race, local context, and innumeracy about racial groups in the United States. *Public Opin. Q.* **71**, 392–412 (2007).
5. B. Duffy, *The Perils of Perception: Why We're Wrong About Nearly Everything* (Atlantic Books, 2018).
6. E. D. Lawrence, J. Sides, The consequences of political innumeracy. *Res. Polit.* **1**, 2053168014545414 (2014).
7. J. Sides, J. Citrin, European opinion about immigration: The role of identities, interests and information. *Br. J. Polit. Sci.* **37**, 477–504 (2007).
8. D. Herda, How many immigrants? Foreign-born population innumeracy in Europe. *Public Opin. Q.* **74**, 674–695 (2010).
9. D. J. Hopkins, J. Sides, J. Citrin, The muted consequences of correct information about immigration. *J. Polit.* **81**, 315–320 (2019).
10. Ipsos, "Perils of perception, prejudice, and conspiracy theories" (Tech. Rep., Ipsos, 2023).
11. J. H. Kuklinski, P. J. Quirk, J. Jerit, D. Schwieder, R. F. Rich, Misinformation and the currency of democratic citizenship. *J. Polit.* **62**, 790–816 (2000).
12. P. Converse, *The Nature of Belief Systems in Mass Publics in Ideology and Discontent* (Free Press, New York, NY, 1964), pp. 206–261.
13. M. Gilens, *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy, Studies in Communication, Media, and Public Opinion* (University of Chicago Press, Chicago, IL, 1999).
14. G. Allport, *The Nature of Prejudice* (Addison-Wesley, Reading, MA, 1954).
15. M. Semyonov, R. Raijman, A. Y. Tov, P. Schmidt, Population size, perceived threat, and exclusion: A multiple-indicators analysis of attitudes toward foreigners in Germany. *Sci. Res. Rev.* **33**, 681–701 (2004).
16. J. C. Dixon, The ties that bind and those that don't: Toward reconciling group threat and contact theories of prejudice. *Soc. Forces* **84**, 2179–2204 (2006).
17. E. Lee *et al.*, Homophily and minority-group size explain perception biases in social networks. *Nat. Hum. Behav.* **3**, 1078–1087 (2019).
18. T. Schlager, A. V. Whillans, People underestimate the probability of contracting the coronavirus from friends. *Hum. Soc. Sci. Commun.* **9**, 59 (2022).
19. E. Thorson, L. Abdelaaty, Misperceptions about refugee policy. *Am. Polit. Sci. Rev.* **117**, 1123–1129 (2023).
20. G. Sparkman, N. Geiger, E. U. Weber, Americans experience a false social reality by underestimating popular climate policy support by nearly half. *Nat. Commun.* **13**, 4779 (2022).
21. M. H. Pasek, L. O. Ankori-Karlinsky, A. Levy-Vene, S. L. Moore-Berg, Misperceptions about out-Partisans' democratic values may erode democracy. *Sci. Rep.* **12**, 16284 (2022).
22. J. Zaller, *The Nature and Origins of Mass Opinion* (Cambridge University Press, 1992).
23. S. S. Stevens, On the psychophysical law. *Psychol. Rev.* **64**, 153 (1957).
24. R. Gonzalez, G. Wu, On the shape of the probability weighting function. *Cogn. Psychol.* **38**, 129–166 (1999).
25. D. E. Erlick, Absolute judgments of discrete quantities randomly distributed over time. *J. Exp. Psychol.* **67**, 475 (1964).
26. C. A. Varey, B. A. Mellers, M. H. Birnbaum, Judgments of proportions. *J. Exp. Psychol.* **16**, 613 (1990).
27. Y. Nakajima, A model of empty duration perception. *Perception* **16**, 485–520 (1987).
28. I. Spence, Visual psychophysics of simple graphical elements. *J. Exp. Psychol.* **16**, 683 (1990).
29. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).
30. H. C. Barth, A. M. Paladino, The development of numerical estimation: Evidence against a representational shift. *Dev. Sci.* **14**, 125–135 (2011).
31. S. Lichtenstein, P. Slovic, B. Fischhoff, M. Layman, B. Combs, Judged frequency of lethal events. *J. Exp. Psychol. Hum. Learn. Mem.* **4**, 551 (1978).
32. J. Hollands, B. P. Dyre, Bias in proportion judgments: The cyclical power model. *Psychol. Rev.* **107**, 500 (2000).
33. D. Landy, B. Guay, T. Marghetis, Bias and ignorance in demographic perception. *Psychon. Bull. Rev.* **25**, 1606–1618 (2018).
34. J. Huttenlocher, L. V. Hedges, S. Duncan, Categories and particulars: Prototype effects in estimating spatial location. *Psychol. Rev.* **98**, 352 (1991).
35. E. B. Schille-Hudson, D. Landy, Scaling uncertainty in visual perception and estimation tasks. PsyArXiv [Preprints] (2020). https://osf.io/preprints/psyarxiv/zu4jx_v1 (Accessed 1 December 2024).
36. D. Bernoulli, Exposition of a new theory on the measurement of risk (1738).
37. G. T. Fechner, Elemente der psychophysik. Vol. 2 (1860).
38. S. Dehaene, The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends Cogn. Sci.* **7**, 145–147 (2003).
39. D. J. Ahler, G. Sood, The parties in our heads: Misperceptions about party composition and their consequences. *J. Polit.* **80**, 964–981 (2018).
40. E. Theiss-Morse, "Characterizations and consequences: How Americans envision the American people" in *Annual Meeting of the Midwest Political Science Association* (Chicago, IL, 2003).
41. J. Citrin, J. Sides, Immigration and the imagined community in Europe and the United States. *Polit. Stud.* **56**, 33–56 (2008).
42. P. J. Conover, S. Feldman, K. Knight, Judging inflation and unemployment: The origins of retrospective evaluations. *J. Polit.* **48**, 565–588 (1986).
43. T. Holbrook, J. C. Garand, Homo economus? Economic information and economic voting. *Polit. Res. Q.* **49**, 351–375 (1996).
44. A. J. Berinsky, Assuming the costs of war: Events, elites, and American public support for military conflict. *J. Polit.* **69**, 975–997 (2007).
45. M. Gilens, Political ignorance and collective policy preferences. *Am. Polit. Sci. Rev.* **95**, 379–396 (2001).
46. L. Bursztyn, D. Y. Yang, Misperceptions about others. *Annu. Rev. Econ.* **14**, 425–452 (2022).
47. T. Marghetis, S. Z. Attari, D. Landy, Simple interventions can correct misperceptions of home energy use. *Nat. Energy* **4**, 874–881 (2019).
48. D. L. Hamilton, R. K. Gifford, Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *J. Exp. Soc. Psychol.* **12**, 392–407 (1976).
49. E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
50. B. Guay, T. Marghetis, C. Wong, D. Landy. Data from "Quirks of cognition explain why we dramatically overestimate the size of minority groups." Open Science Framework. https://osf.io/cuvgw/. Deposited 12 February 2025.