Measuring Perceptions of Numerical Strength of Salient and Stereotypical Groups*

Douglas J. Ahler† Gaurav Sood‡

Misinformation—belief in incorrect information (Luskin, Sood and Blank 2013)—is distressingly common. Despite overwhelming, widely available evidence showing otherwise, some people believe that President Barack Obama was born outside the U.S. (ch. 2), and some that "light" cigarettes are neither addictive nor cancer causing (ch. 4).

Group stereotypes are a type of misinformation—perhaps the most commonly held misinformation. Unlike the examples above, however, many group stereotypes are rooted in kernels of truth. People develop mental images of groups to better understand the social world and their place in it (Lippman 1922). But the social world is complex, and the cognitive capacity, limited. Thus, people rely on mental shortcuts or heuristics when developing these mental images, ending with impressionistic accounts rather than photo-realistic portraits of the world outside. These impressionistic accounts often, however, are systematically biased. People often believe that stereotypical traits are far more common in the stereotyped group than they are, and that differences across groups on stereotypical traits are much larger than they are.

Given exaggerated differences likely inflame intergroup animus (e.g., Ahler and Sood 2016), group stereotypes speak to some of the most fundamental questions in social science. What explains inter-group conflict? What are its consequences? And how do we temper (or inflame) it? And beliefs about how strong a group is, especially a group's numerical strength, are often foundational to these concerns. So, in this chapter, we focus on the measurement of

*We thank Rachel Bernhard, Andy Guess, Stephen Goggin, Eric Lawrence, and Kabir Khanna for useful feedback.
†Doug is Assistant Professor of Political Science at Florida State University, doug.ahler@gmail.com.
‡Gaurav can be reached at gsood07@gmail.com.

beliefs about numerical strength of salient, stereotypical groups in the population and in other stereotype-relevant groups.

It is unlikely that more than a handful of people hold any beliefs about the numerical strength of most groups—for example, how many people keep bees, drive sedans, or own more than two pairs of jeans.[1] But for stereotypical groups or groups that are salient to policy debates, e.g., immigrants, African-Americans in the Democratic Party, many people likely hold crude beliefs about their numerical strength.

We contend that these crude beliefs implicitly map to specific numbers. And that we can elicit these numbers on surveys, though likely only with considerable noise. For instance, a person may hold the belief that "many" undocumented immigrants live in the United States. This belief may affect their attitude toward immigration without they ever cognizing the mapping of "many" to the number that the belief implicitly maps to. However, when asked about the share of undocumented immigrants in the population, the person may report that 15% of the people living in the country do so illegally. (The actual number is 3.5%, as per Krogstad and Passel 2016.) We contend that the reported 15% is a function of the underlying belief and random error.

This conceptualization raises a variety of mechanistic and conceptual concerns. For instance, does it matter how much time we give people to report their beliefs? Do people honestly report what they believe? Or do they instead offer responses that reflect how they feel about the group? Does innumeracy, rather than misperception, for instance, explain the large positive bias in survey reports of perceptions of share of group-stereotypical groups in the group (Ahler and Sood 2016)? Each concern is a threat to the *validity* of measures of perceptions about composition. Each suggests that something comes between people's beliefs about the numerical strength of a group and the survey reports and alters those survey reports systematically.

[1] This isn't to say that most people will draw a blank if asked these questions. Most will use related knowledge to infer these quantities. For instance, a person may reason, "I don't eat honey often as the flavor is unpleasant. Given the unpleasant flavor, I don't think there are a lot of people who eat honey. So, if I had to guess, I would say not many people keep bees."

However, even if concerns like these prove unfounded, it is not clear how to interpret elicited perceptions. If we find that people overestimate the share of group g in subpopulation s—e.g., the percentage of Democrats $[s]$ who are black $[g]$—what do we make of it? Could it be that people think that the share of g in the population is larger than it is, and that causes them to overestimate $p(g|s)$. Or is it that people accurately perceive the share of g in the population but overestimate its share in s?

We shed light on these and other such questions around conceptualization, measurement, and interpretation of perceptions of numerical strength of groups in populations and subpopulations. We provide methods for assessing the validity of reported perceptions, demonstrating the use of some of the tools with two large survey experiments.

Causes and Consequences of Misperceptions of Numerical Strength of Groups

Misperceptions about the numerical strength of groups are thought to affect a host of important variables, including how people feel toward groups and the people in them. But why should that be so? Sometimes attitudes toward groups are rooted in the zero-sum struggle between groups for resources and power. In such circumstances, an increase in the actual (or imagined) share of an out-group amplifies the perceived threat of the out-group seizing power—especially in a democracy—and executing its agenda. Members of the threatened group respond to the increase in perceived threat with greater in-group solidarity and out-group hostility, greater opposition to policies that benefit the out-group, and greater likelihood of preemptive action to curtail power, such as, voting to reduce the power of the threatening group (Nadeau, Niemi and Levine 1993; Sides and Citrin 2007*a*; Key 1949; Blalock 1967; Dancygier 2010).

Social identity theory provides another explanation for how people evaluate groups and why perceptions of group shares may matter. People seek positive distinctiveness. In-group favoritism arises because peoples' self-concepts are tied to their group membership (Tajfel and Turner 2005). So powerful is the need for positive (group) distinctiveness that membership in even arbitrarily assigned groups engenders in-group favoritism (Sherif 2015; Tajfel 1970). And people evaluate novel groups based on their perceptions of how known social groups compose them. For instance, people conceptualize political parties in terms of longer-standing social groups, like those based on race and social class (Green, Palmquist and Schickler 2002). And partisans' perceptions of the share of these longer-standing social groups in the parties affects their feelings toward opposing party supporters (Ahler and Sood 2016). In all, both realistic conflict and social identity theories predict that how large people think a group is affects not only how threatened they feel by the group, but also attitudes toward the group, entangled policy preferences, and behavior toward members of the group (Wong 2007; Wong et al. 2012; Sides and Citrin 2007*a*).

Troublingly, perceptions of numerical strength of groups are often biased in ways that heighten conflict (e.g., Ahler and Sood 2016; Nadeau, Niemi and Levine 1993; Sides and Citrin 2007*a*). These biases are often a result of reliance on accessible information, disinformation, and representativeness. We discuss each in sequence.

We are most likely to receive information that is most readily available. But what is accessible is not always apt for drawing correct inferences. For example, local television news covers violent crime far more than non-violent crime even though non-violent crime is far more common. And watching local news likely leads some people to wrongly believe that violent crime in their area is more common than it is (Romer, Jamieson and Aday 2003). Similarly, it is no surprise that Americans' beliefs about the percentage of the poor who are black hew closely to media depictions of the poor (Gilens 1996), and that perceptions of the percentage of Democrats and Republicans belonging to party-stereotypical groups are the most biased among people who report having the greatest interest in political news (Ahler and Sood 2016).

Accessibility bias implies that people make erroneous inferences about group strength from readily available information; elite communication *indirectly* drives misinformation. By contrast, disinformation implies that people are *directly* misled. Cynical spreading of falsehoods with the aim of persuading the electorate to change their preferences and behavior is not new (Jackson and Jamieson 2007), but the 2016 presidential elections brought this concern to fore. For example, several stories that grossly overstated the total number of immigrants (especially illegal immigrants) in the country circulated during the campaign. And some of these stories were shared (and read) widely on the social media, sometimes cynically disseminated by others; Barthel, Mitchell and Holcomb (2016) find that "14% [said] they shared a story they knew was fake at the time [of sharing]." As such, there is reason to think that some people hold incorrect beliefs about the share of prominent social groups entangled in policy debates because they took as fact disinformation from a trusted source.

Aside from external phenomena like biased accessible information and disinformation, internal processes can also cause misinformation. Foremost among them is the use of representativeness heuristic (Bordalo et al. 2016; Tversky and Kahneman 1974). In the case of perceptions about group composition, this is likely to result in people neglecting base rate information. For example, when people are asked to evaluate the percentage of poor Americans who are black, they are apt to focus on the categorical overlap in memory between blacks and the poor and ignore that just 13% of the U.S. population is black. Bordalo et al. (2016) formalize this logic, asserting that group g is representative of subpopulation s if it scores high on the ratio $\frac{p(g|s)}{p(g|\neg s)}$. They further empirically demonstrate that people overweight representative groups (g) when making judgments about other groups' (s) composition.

Perceptions of numerical strength of groups, thus, are of great interest, not only because they affect intergroup attitudes and relations, but also because biases in them can shed light on the nature of information flows about social groups and events. However, measuring these perceptions presents some challenges.

Conceptual, Measurement, and Interpretation Concerns

A half century of survey data suggests that the average American knows little about politics (Campbell et al. 1960; Delli Carpini and Keeter 1996). Roughly half of the survey respondents fail to identify their member of Congress, and similar numbers fail to correctly place the political parties on major issues (e.g., Freeder, Lenz and Turney 2016). Such disengagement suggests that it is unlikely that most people will encounter—much less remember—specific data on the numerical strength of even salient or stereotypical social groups. Instead, most people likely have crude beliefs—stereotypes founded in representativeness, or impressions based on inferences from accessible information. These crude beliefs, however, likely map implicitly to numbers, which people use to react to situations. And it is these numbers that survey researchers want to elicit. But a variety of concerns and questions attach themselves to measurement of these numerical perceptions.

Top-of-the-Head vs. Considered Responses

The quantity of interest is the numbers that people's beliefs about group composition implicitly map to. And it is likely that the more automatic the response, the better it is at capturing the implicit mapping. Thus, researchers may want to curtail the time respondents have between accessing the crude belief and reporting its numerical mapping. Providing more time to respond may yield reports that reflect additional considerations and reasoning beyond what people would normally engage in when accessing these beliefs in real life. For instance, given additional time, a respondent may reason that their gut response is too large and adjust it accordingly. This process may thus also change the underlying belief. This, in turn, suggests that eliciting more considered responses may be useful in evaluating the degree to which misperceptions can be corrected through slower, more effortful processing alone (e.g., Kahneman 2011; Petty and Cascioppo 1986).

Use of Denominators Larger than 100

When asked to report shares of an exhaustive set of mutually exclusive groups in a population or a subpopulation, peoples' answers often sum to more than 100 (Wong 2007; Lawrence and Sides 2014; Macchi, Osherson and Krantz 1999). Thus, bias in reported perceptions of numerical strength of groups may be a reflection of nothing more than innumeracy about percentages. The concern can be addressed in two different ways. The first is recalibration. For instance, if shares of an exhaustive, mutually exclusive set of categories sum to 125, estimates for each of the categories can be divided by the more appropriate denominator (125). Such recalibration assumes that relative error is the same across categories; it may not be.

Alternatively, one can address the problem by amending the survey instrument in a way that makes respondents more acutely aware of the appropriate denominator. For instance, one may force respondents to sum the shares of comprehensive sets of mutually exclusive groups to 100 (e.g., Ahler and Sood 2016). This solution, however, likely has some side effects. Not only is it likely cognitively taxing for the respondents, it may also cause them to think more effortfully about the quantities than they would do in real life. Alternatively, and more simply, to address the concern, one may rephrase the question stem as "Out of 100, how many..." as opposed to "What percentage..." (Sides and Citrin 2007*b*). However, it isn't clear whether this strategy ameliorates the concern.

Motivated Responding

Respondents may intentionally misreport their beliefs about the numerical strength of groups to express their feelings about the groups referenced in the survey question (see Bullock et al. 2015; Khanna and Sood 2017; Prior, Sood and Khanna 2015). In particular, people may intentionally overstate the share of groups they (dis)like in groups they (dis)like. For example, a white racist may purposefully over-report the percentage of poor Americans who are black, as doing so casts aspersion on a disliked racial out-group.

A bulk of the evidence for motivated responding comes from experiments that pay people for providing the correct answer and for confessing to ignorance. For instance, giving money for giving the correct answer halves the partisan gap in responses to knowledge items with partisan implications, e.g., changes in the unemployment rate under a Democratic president Prior, Sood and Khanna (2015). Giving money for marking "Don't Know" when you don't know and a little more money for marking the correct answer also has the same effect (Bullock et al. 2015).

Interpreting the results of experiments that provide financial incentives, however, can be tricky. Firstly, incentives may encourage respondents to consult outside sources; like Bullock et al. (2015), researchers may want to use placebo questions to gauge the extent of the concern. Secondly, to earn the reward, respondents may revise their responses to line up with their perceptions of researchers' beliefs. Asking respondents to guess how the researcher(s) would answer may be one way to gauge that concern. Thirdly, providing incentives likely yields more considered responses. And as we argue above, top-of-the-head answers may be closer to the beliefs respondents generally carry about salient and stereotypical groups.

Misunderstanding Question or Scale

Ambiguity about what is being asked abrades the validity of the survey item. For instance, the question, "What percentage of Democrats are black?" is somewhat ambiguous. Some respondents may reasonably wonder if by Democrats we mean people who identify themselves as Democrats, or also those who lean toward the party when probed, or instead are referring to those who voted for the Democratic candidate in the last presidential election. All of this assumes that these distinctions appreciably alter the elicited number or the relevant population statistic. Many a times it doesn't matter. For instance, including those who *lean* toward a party doesn't appreciably change the share of prominent party-stereotypical groups in the "party" (Ahler and Sood 2016). But even so, precision is preferred.

There is, however, generally a trade-off between precision, compactness, comprehensibility. And given only a small chunk of respondents are likely to be aware of these finer distinctions, one idea may be to keep the question stem as is and ask an additional open-ended question about the definition of the quantity being estimated. A person's understanding of the quantity being asked can then be used to more clearly interpret the responses.

Ambiguous questions are but one source of problems. Another alleged problem is response scales that don't give information about benchmarks. Ansolabehere, Meredith and Snowberg (2013) argue that "providing respondents with benchmark quantities…can reduce measurement error due to respondents not understanding the scale on which more complex quantities, such as the unemployment rate, are measured." We are skeptical of the argument that providing benchmarks reduces measurement error. It is odd to claim that people can know the unemployment rate and yet not know its scale. If a respondent knows that the unemployment rate is 4.4%, comprehension of the scale is moot. And interpreting the effect of offering a benchmark rate—lower error—as better comprehension of the scale seems unwarranted. Offering a benchmark rate is liable to reduce error not because respondents suddenly realize that the unemployment rate is on a 101 (0–100) point scale, but because respondents can better calibrate their guesses. Another negative side-effect of offering benchmarks is that they act as low-information *anchors*, shrinking variance, and adding bias to the elicited answers (Tversky and Kahneman 1974).

Beliefs About Related Quantities

When interpreting reports of perceptions of share of group g in subpopulation s, for e.g., the percentage of Republicans who earn $250,000 per year or more, researchers face a unique challenge. Someone who only has beliefs about the terms on the right-hand side of the equation below and using the Bayes rule would do the following calculation:

$$p(g|s) = \frac{p(s|g)p(g)}{p(s)} = \frac{p(s|g)p(g)}{p(s|g)p(g) + p(s|\neg g)p(\neg g)}$$

or using the example,

$$p(\text{\$250K}|\text{Republican}) = \frac{p(\text{Republican}|\text{\$250K})p(\text{\$250K})}{p(\text{Republican}|\text{\$250K})p(\$250K) + p(\text{Republican}| < \text{\$250K})p(< \text{\$250K})}$$

As the equation makes clear, Bayesians can overestimate $p(\text{\$250K}|\text{Republican})$ for two very different reasons. They may believe that a larger share of people earning over \$250,000 identify as Republicans than in reality. Or, they may believe that a larger share of Americans earns over \$250,000 than in reality. If the latter is true, people will also likely overestimate the percentage of Democrats and independents who earn a lot. And that would suggest that bias in perceptions of $p(\text{\$250K}|\text{Republican})$ has less to do with party stereotypes than what looking at $p(\text{\$250K}|\text{Republican})$ alone may lead us to believe. Thus, to interpret $p(g|s)$ correctly, we sometimes also need to learn about beliefs about related quantities.

Related quantities can also illuminate psychological mechanisms behind perceptual biases. For instance, failure to attend to $p(g)$ when estimating $p(g|s)$—the *base rate fallacy*—is a sign that people are using the representativeness heuristic. Thus, to assess whether representativeness is behind biased perceptions of $p(g|s)$, researchers may want to track impact of providing information about base rates (see for e.g., Ahler and Sood 2016).

Researchers interested in correcting misperceptions may also find value in priming $p(s|\neg g)$ before asking respondents about $p(g|s)$. For instance, they may ask respondents about the percentage of Republicans who earn less than \$250,000 per year. Or, instead, ask respondents to think about five Republican friends or associates—who, because only a small fraction of the populations earns that much, are unlikely to earn over \$250,000 per year—or to write down something about their jobs or socioeconomic status (e.g., Thorson 2016).

Evaluating Measures of Perceptions of Numerical Strength of Groups

To assess the concerns, we exploit data from two surveys with multiple embedded experiments conducted on Amazon's Mechanical Turk (MTurk). MTurk is a market for trading small services. Compared to the population, survey respondents recruited on MTurk tend to be younger, better educated, and more likely to identify with the Democratic party (Berinsky, Huber and Lenz 2012) (For details about the composition of the samples, see Online Appendix [OA] OA 1.1 and OA 2.1.)

We conducted the first study (Study 1) in November, 2014 to assess concerns about the validity of the finding that Americans overestimate the share of party-stereotypical groups in the two main political parties (see Ahler and Sood 2016). To assess the concerns, we randomly assigned respondents to one of four conditions: a control condition, and three other conditions designed to assess the validity of a specific alternative explanation or interpretation of the results. We leave the description of the conditions to sections discussing the relevant concern.

In the second study (Study 2), conducted in July, 2016, we asked about perceptions of shares of a broader variety of groups: 1) the share of Democrats who are atheist/agnostic, black, gay/lesbian/bisexual, and union members, 2) the share of Republicans who are age 65+, Evangelical, Southern, and earn over $250,000 per year, 3) the share of Americans who drink diet soda (for some respondents), and the percentage of American men and women who do so (all respondents), 4) the percentage of people living in America who were born outside the US (e.g., Alba, Rumbaut and Marotz 2005), and 5) the percentage of the American poor who are black (e.g., Gilens 1996). And once again, we leave the description of the specific experiments to sections discussing the relevant concerns.

Top-of-the-Head vs. Considered Responses

Top-of-the-head answers are likely closest to the numbers that stereotypes implicitly map to. If so, how biased are more considered responses? To evaluate that, in Study 2, we assigned roughly 300 respondents to a timing experiment. Half of the respondents were randomly assigned to the `time pressure` condition in which they were only given 10 seconds to answer each item, and half were assigned to the `time requirement` condition in which they had to wait 15 seconds with just the question text on the screen before they could type their response. Respondents were alerted about the timed portion of the survey before answering the questions. (See the online appendix to Ahler and Sood (2016) for depictions of all treatments in Study 1. See OA for depictions of all treatments in Study 2.)

Use of Denominators Larger than 100

To gauge the impact of implicit use of denominators larger than 100, we conducted two experiments. In Study 1, respondents were assigned to a `sum-to-100` condition in which they not only reported their beliefs about the percentage of partisans belonging to party-stereotypical groups, but also beliefs about a comprehensive set of complementary, mutually exclusive groups. For instance, respondents not only reported their perceptions of the percentage of Republicans who are evangelical Christian, but also the percentage of Republicans who are mainline Protestant, Catholic, and "other/no religion." We required that their responses add up to a 100. An on-screen counter displayed the running total. The difference between results in the `sum-to-100` condition and the `standard estimation` condition give us the extent to which implicit use of denominators larger (different) than 100 affects estimates of $p(\text{group}|\text{party})$.

In Study 2, another 300 respondents were assigned to a wording experiment designed for the same purpose. We randomly manipulated question stems to read either, "Out of every 100 P/S, how many do you think are G?" or, "What percentage of P/S do you think are G?" We expect "Out of every 100 . . . " to make the correct denominator more salient. We manipulated the stems of the party composition items, foreign-born population item, and racial composition of the poor item.

Motivated Responding

We assessed the extent to which motivated responding affects responses by offering `accuracy incentives` to a random subset of respondents in Study 1. Respondents in the `accuracy incentives` condition received an additional five cents, 20% of the compensation for finishing the survey (25 cents) for each response that fell within five percentage points of the truth. While the bonus may seem small, given that respondents answered items on eight groups' shares, they had the opportunity to nearly triple what they made for the survey. If reported perceptions about party composition reflect motivated responding, estimates of respondents assigned to the `incentives` condition should be substantially different from those elicited without incentives.

Beliefs About Related Quantities and Interpretation of Responses

Do misperceptions about $p(\text{group}|\text{party})$ merely reflect misperceptions about $p(\text{group})$ instead of partisan stereotypes? We gauged the possibility in three ways. Firstly, and perhaps most dispositively, in Study 1, we removed ignorance about base rates as a plausible alternative explanation. We did so by anchoring sliders at the base rate for each party-stereotypical group, alerting respondents to this design feature, and then asking them to use the sliders to estimate $p(\text{group}|\text{party})$. Significantly lower estimates in the `base rates` condition would mean that inflated base rates *potentially* explain inflated beliefs about $p(\text{group}|\text{party})$.

Secondly, in the `standard estimation` condition, we asked respondents to estimate the groups' share in the population in addition to their share in a particular party. We can compare reported perceptions of a group's share in the population to reported perceptions of share in the party to test whether misperceptions reflect anything beyond ignorance of base rates. To more clearly interpret misperceptions about subpopulation composition, we can compute a difference-in-differences:

$$(p(\text{group}|\text{representative party})_P - p(\text{group}|\text{representative party})_A) - \\ (p(\text{group})_P - p(\text{group})_A)$$

where P indexes perceived quantities and A indexes actual quantities. Note that the former estimate, $p(\text{group}|\text{representative party})_P - p(\text{group})_P$ measures the extent to which people's misperceptions exceed what we would expect if they were only using their erroneous base rates to assess $p(\text{group}|\text{party})$. The difference-in-differences estimate, by contrast, tells us whether party stereotypes or erroneous beliefs about the group's base rate more strongly color judgments about party composition.

Lastly, in Study 2, we used the diet soda items to test a hypothesis regarding beliefs about related quantities. Although roughly equal percentages of men (23%) and women (24%) report drinking diet soda (Gallup 2013), advertisers often target women (Lin 1998; Yoder, Christopher and Holmes 2008). Thus, we suspect that Americans overestimate the gender gap in diet soda consumption. We randomly assigned half of respondents to provide their belief about the base rate of diet soda consumption in America before answering the items specific to men and women. With between-conditions data, we can assess whether asking about base rates reduces bias in reported perceptions of the percentage of men and women who drink soda.

Results

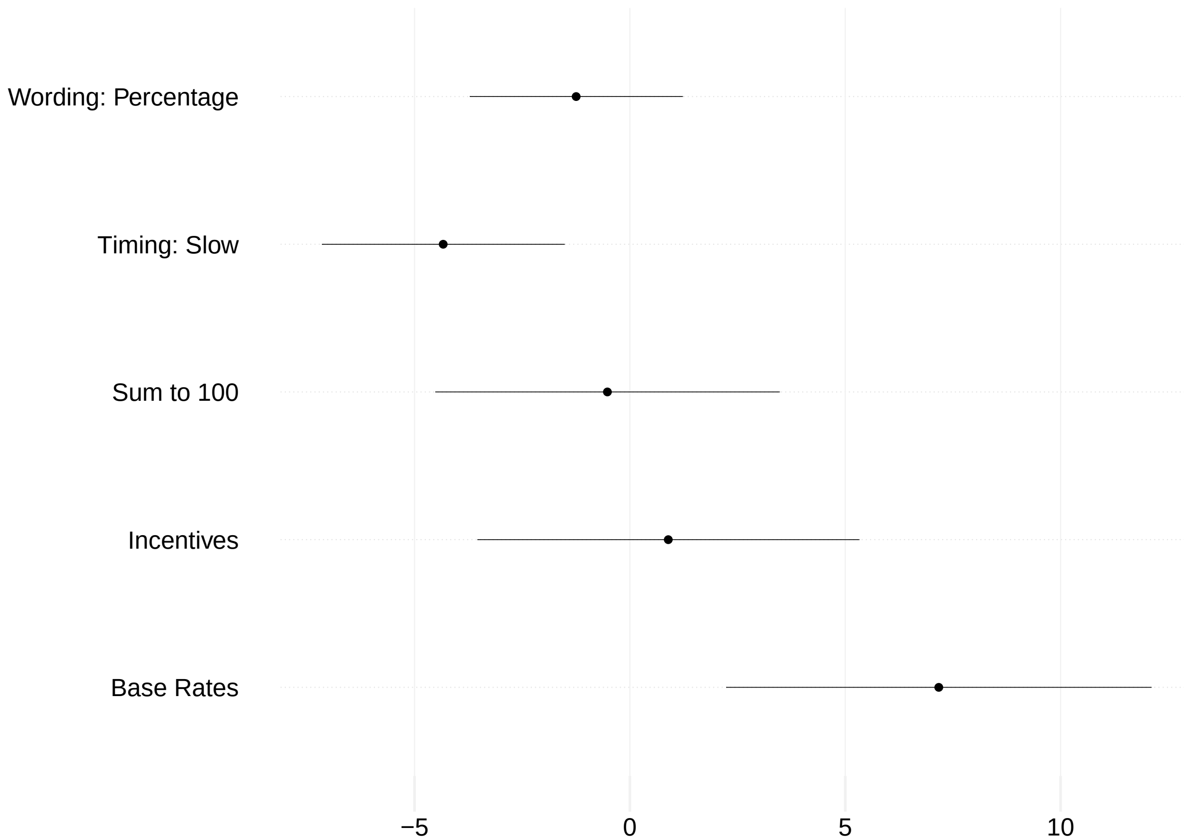
Thinking (About Group Shares) Fast and Slow

For all the ten items in the timing experiment—the eight party composition items, the percentage of foreign-born in the U.S. item, and the percentage of poor who are black item—responses in the `time requirement` condition are more accurate. Not all differences are statistically significant, though half have $p < 0.05$ and 7 of the 10 have $p < 0.1$ (see OA 2.3). To estimate the average difference across conditions across items, we regressed *perceptual bias*, the signed difference between reported perception of the share and our best estimate of the actual share, on `time requirement` treatment and item dummies, clustering the standard errors by respondent.[2] On average, perceptual bias in the `time requirement` condition is 4.3 points lower than in the `time pressure` condition (see Figure 1). However, it is only a 22% decline; average bias in reported perceptions is 19.7 points in the `time pressure` condition and 15.5 points among those assigned to the `time requirement` condition.

Lower bias in the `time requirement` condition, however, may be due to respondents using the additional time to consult outside sources. To assess the concern, we plotted a density curve of the responses to all the items by treatment condition. If lower bias in the `time requirement` condition is a consequence of cheating, we should see spikes in the density plot at the correct answer. But we do not see these spikes (see OA 2.4). To formally test for cheating, we compared proportion correct (within one percentage point of the actual share) across conditions; there were no significant differences (see OA 2.5).

[2] In OA 2.8, we also investigated the impact of the treatments on *absolute error*, average absolute error in respondents' reports across items, and *percentage bias*, average percentage by which respondents err (signed error) across items. Impact of the treatments on both *absolute error* and *percentage bias* is small.

Figure 1: Average Effects of Eliciting Perceptions of Various Groups in Different Ways



NOTE: 95% confidence intervals depicted. Results are from OLS regression of *perceptual bias* on treatment indicators (with baseline conditions noted here), with item fixed effects and standard errors clustered by respondent. The "standard estimation" condition serves as a baseline for the "Sum-to-100," "Incentives," and "Base rates" conditions. The "time pressure" condition serves as a baseline for the "time requirement" (or "slow") condition. The alternative stem wording ("Out of every 100...") serves as a baseline for the "Percentage" (standard stem wording) condition.

Use of Denominators Larger than 100

If people implicitly use denominators larger than 100 when reporting percentages, estimates in the `sum-to-100` condition should be significantly lower than in the `standard estimation` condition. However, when respondents are required to ensure that shares of a comprehensive set of mutually exclusive groups add up to a 100, they are generally just as biased. For just one of the eight items—the black-Democratic group-party dyad—are the responses significantly less biased (see OA 1.1). As Figure 1 shows, average bias in reported perceptions fell by just 0.5 points, from 18.9 points in the `standard estimation` condition to 18.4 in the `sum-to-100` condition. (Results are from OLS regression of error on indicators for Study 1 treatments, with item fixed effects and standard errors clustered by respondent.)

Data from a different research design, changing the question stem to make the correct denominator more salient, also yields little evidence of implicit use of denominators larger than 100 being very consequential. "Out of every 100…" wording reduces bias in just three of the ten items (see OA 2.6 for plots). And pooling across items, on average, the "Out of every 100" wording causes a 1.2-point increase in perceptual bias. (Again, the results are from an OLS regression of perceptual bias on an indicator for the "out of every 100" treatment, with item fixed effects and standard errors clustered by respondent.) Given that the usual "What percent" wording yields an average perceptual bias of 18.9 points, this difference is neither statistically nor substantively significant. In sum, neither making the appropriate denominator more salient nor constraining respondents to the right denominator significantly reduces bias in their estimates, suggesting that respondents' implicit use of denominators greater than 100 doesn't bias reported perceptions of the share of groups asked in the surveys.

Motivated Responding

To deter motivated responding, in Study 1, we offered monetary incentives to a random set of respondents. If reports of p(group|party) are distorted by motivated responding, estimates in the `accuracy incentives` condition should be very different. They are not. As Figure 1 shows, pooling across items, mean perceptual bias in the `accuracy incentives` condition is roughly the same as in the `standard` condition. Not only that, even the distribution of the responses is similar across the conditions (see OA 1.3).

Observational evidence also suggests that motivated responding did not significantly bias responses on these items. In a separate survey that we conducted on MTurk in April 2014, we asked respondents to tell us how they felt toward the eight party-stereotypical groups on a 101-point "feeling thermometer" scale (Ahler and Sood 2016). Later in the survey, we collected perceptions of p(group|party) for the four out-party dyads. If motivated responding explains the responses, feelings about the groups should strongly predict beliefs about their share in the party. For all the party-group dyads, the linear relationship is extremely feeble (see OA 1.4 for tables). Pooling across party-group dyads, a one-point increase in group feeling thermometer rating corresponds to just a .03-point decrease in perceptions of p(group|party)—a relationship that is neither statistically nor substantively significant. In all, it appears that bias in reports of perceptions of shares of groups due to motivated responding is small.

Beliefs About Related Quantities

Even if measures of perceptions of p(group|subpopulation) capture beliefs, they can still be hard to interpret. Without additional data, it is unclear whether the perceptions reflect beliefs about composition of the subpopulation, or beliefs about the group's share in the population.

To clarify the interpretation, we measured beliefs about the share of party-stereotypical groups in the population in Study 1. The data show that people generally overestimate the share of party-stereotypical groups in the population. The bias is especially large for small groups. For example, gays, lesbians, and bisexuals are just 3.8% of the population, but people think they are 14.9% of the population. Similarly, respondents think that those who earn over $250,000 per year are 11.4% of the population—only about 2% of the population actually earns that much (see OA 2.7).

While people tend to overestimate the share of party-stereotypical groups in the population, they think their share in the parties is yet larger ($p(\text{group}|\text{representative party})_P - p(\text{group})_P > 0$). The difference is significant for all eight party-group dyads except for the black-Democratic dyad; like Wong (2007), we find that people greatly overestimate the share of African-Americans in the population.

The difference between the extent to which people overestimate the share of party-stereotypical groups in "their" party and in the population can shed light on the extent to which party stereotypes influence people's assessments of $p(\text{group}|\text{party})$. The difference-in-differences is significantly greater than zero for five of the eight dyads, implying a role of party stereotypes; estimates are not significantly different from zero for the others suggesting a more muted role of stereotypes for those groups.

In addition to these descriptive and observational analyses, we conducted an experiment in Study 1 to more cleanly estimate the extent to which beliefs about base rates explain beliefs about party composition. As we discuss above, we randomly assigned one group of respondents to answer the party composition items with sliders anchored at the groups' base rates, informing respondents we had done so. As Figure 1 shows, reported perceptions become more biased, not less. The results suggest that perceptions are driven by *representativeness*.

Lastly, asking people about their beliefs about base rates before asking about the share of subgroups does not reduce bias. Recall that in Study 2 we had randomly assigned half of the respondents to report their beliefs about the percentage of Americans who drink diet soda (24% in reality) before eliciting their beliefs about the percentage of men (23%) and women (24%) who do so. Respondents in the `no base rate` condition significantly overestimated the gender gap in diet soda consumption, reporting that 46.2% of women drink diet soda compared to 30.0% of men. If anything, those who were asked about the base rate first (average estimate = 41.0%) became somewhat worse: rather than alter their estimates about women (46.7%), they likely changed their beliefs about men's diet soda consumption (27.4%). The difference in the reported gender gap across the two conditions is not statistically significant (95% CI: [-1.0, 7.2]) but comports with the evidence from Study 1. Overall, priming or providing base rates has little effect on reported beliefs about subpopulation composition, and may sometimes even make those reports more biased.

Discussion

A variety of politically salient concerns are linked to misperceptions about the numerical strength of groups in the population or in various subpopulations. For instance, nativist sentiment is associated with biased beliefs about the share of immigrants in the population (e.g., Alba, Rumbaut and Marotz 2005), anti-welfare attitudes are linked to misperceptions about the share of poor who are black (Gilens 1996; 1999), and partisan antipathy to erroneous beliefs about the share of out-party supporters who belong to party-stereotypical groups (Ahler and Sood 2016). These concerns take on additional heft when paired with the frequency with which they are exploited. For instance, both the "Brexit" referendum and Donald Trump's 2016 presidential campaign prominently highlighted claims about immigration levels.

Given the importance of these concerns, the topic deserves sustained careful attention, starting with valid measurement of key variables. In this chapter, we shed light on measurement of one key variable—perceptions of the numerical strength of groups. In particular, we highlight and illustrate some of the inferential strategies that researchers can use to assess the severity of some of the most pressing concerns.

Data suggest that people carry cognitions about the share of some politically salient and stereotypical groups in the population and some subpopulations, and that these beliefs can be reliably elicited by asking respondents to give numerical estimates of the share of these groups. Commonly noted concerns like motivated responding, use of denominators larger than 100, and "cheating," and less commonly noted concerns like considered responding, do not significantly bias survey reports of perceptions. And the impact of addressing some of the concerns may even be in the opposite direction than hypothesized. For instance, providing groups' base rates made reported beliefs about those groups' shares in certain subpopulations slightly worse.

The conclusions may not hold for items other than those discussed here. We focused on items tapping prevalent group stereotypes. And we contend that for such groups, people are liable to have crude but firm beliefs about their numerical strength. And our claim is not that the concerns raised about these measures never apply—we cannot even confidently rule out all the concerns for all the measures presented here without collecting significant additional data.

Much of science reduces to measurement—of a phenomenon, its causes, and its consequences. And much of scientific progress has been built on improvements in measurement. Our hope is that the chapter will spur additional conversation and research on how best to measure perceptions of numerical strength of salient and stereotypical groups. And that a better measurement machinery for assessing numerical perceptions of group strength will lead to progress in our understanding of an important driver of inter-group conflict.

References

Ahler, Douglas J and Gaurav Sood. 2016. "The Parties in our Heads: Misperceptions About Party Composition and Their Consequences.".

Alba, Richard, Ruben G Rumbaut and Karen Marotz. 2005. "A distorted nation: Perceptions of racial/ethnic group sizes and attitudes toward immigrants and other minorities." *Social Forces* 84(2):901–919.

Ansolabehere, Stephen, Marc Meredith and Erik Snowberg. 2013. "Asking about numbers: Why and how." *Political Analysis* 21(1):48–69.

Barthel, Michael, Amy Mitchell and Jesse Holcomb. 2016. "Many Americans Believe Fake News Is Sowing Confusion." *Pew Research Center* . http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 289–300.

Berinsky, Adam J., Gergory A. Huber and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(2):351–368.

Blalock, Hubert M. 1967. "Toward a theory of minority-group relations.".

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli and Andrei Shleifer. 2016. "Stereotypes." *Quarterly Journal of Economics* 131:1753–1794.

Bullock, John G., Alan S. Gerber, Seth J. Hill and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10(4):519–78.

Campbell, Angus, Philip E. Converse, Warren E. Miller and Donald E. Stokes. 1960. *The American Voter.* Chicago: University of Chicago Press.

Dancygier, Rafaela M. 2010. *Immigration and conflict in Europe.* Cambridge University Press.

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters.* New York: Yale University Press.

Freeder, Sean, Gabriel S. Lenz and Shad Turney. 2016. "What Goes with What: Reinterpreting the Evidence on Attitude Stability, Policy Voting, and Multi-Item Issue Scales." Working paper: University of California, Berkeley.

Gallup. 2013. ""Regular Soda Popular With Young, Nonwhite, Low-Income".". http://www.gallup.com/poll/163997/regular-soda-popular-young-nonwhite-low-income.aspx.

Gilens, Martin. 1996. "Race and Poverty in America: Public Misperceptions and the American News Media." *Public Opinion Quarterly* 60(4):515–541.

Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media and the Politics of Antipoverty Policy.* Chicago: University of Chicago Press.

Green, Donald, Bradley Palmquist and Eric Schickler. 2002. *Partisan Hearts & Minds: Political Parties and the Social Identities of Voters.* New Haven, CT: Yale University Press.

Jackson, Brooks and Kathleen Hall Jamieson. 2007. *unSpun: Finding facts in a world of disinformation.* Random House.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* New York: FSG.

Key, Valdimer. 1949. "Southern politics in state and nation.".

Khanna, Kabir and Gaurav Sood. 2017. "Motivated Responding in Studies of Factual Learning." *Political Behavior* pp. 1–23.

Krogstad, Jens Manuel and Jeffrey S. Passel. 2016. "5 Facts About Illegal Immigration in the U.S." *Pew Research Center* 19 November. http://www.pewresearch.org/fact-tank/2015/11/19/5-facts-about-illegal-immigration-in-the-u-s/.

Lawrence, Eric D and John Sides. 2014. "The consequences of political innumeracy." *Research & Politics* 1(2):2053168014545414.

Lin, Carolyn A. 1998. "Uses of Sex Appeals in Prime-Time Television Commercials." *Sex Roles* 38(5/6):461–475.

Lippman, Walter. 1922. *Public Opinion*. Transaction Publishers.

Luskin, Robert C, Gaurav Sood and Joshua Blank. 2013. The Waters of Casablanca: Political Misinformation (and Knowledge and Ignorance). In *Annual Meeting of the Midwest Political Science Association*.

Macchi, Laura, Daniel Osherson and David H Krantz. 1999. "A note on superadditive probability judgment." *Psychological Review* 106(1):210.

Nadeau, Richard, Richard G Niemi and Jeffrey Levine. 1993. "Innumeracy about minority populations." *Public Opinion Quarterly* 57(3):332–347.

Petty, Richard and John Cascioppo. 1986. "The Elaboration Likelihood Model of Persuasion." *Advances in Experimental Social Psychology* 19:123–181.

Prior, Markus, Gaurav Sood and Kabir Khanna. 2015. "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions." *Quarterly Journal of Political Science* 10(4):489–518.

Romer, Daniel, Kathleen Hall Jamieson and Sean Aday. 2003. "Television News and the Cultivation of Fear of Crime." *Journal of Communication* 53(1):88–104.

Sherif, Muzafer. 2015. *Group conflict and co-operation: Their social psychology.* Vol. 29 Psychology Press.

Sides, John and Jack Citrin. 2007*a*. "European opinion about immigration: The role of identities, interests and information." *British journal of political science* 37(03):477–504.

Sides, John and Jack Citrin. 2007*b*. How large the huddled masses? The causes and consequences of public misperceptions about immigrant populations. In *annual meeting of the Midwest Political Science Association, Chicago, IL.*

Tajfel, Henri. 1970. "Experiments in intergroup discrimination." *Scientific American* 223(5):96–102.

Tajfel, Henri and John C. Turner. 2005. *Political Psychology.* New York: Psychology Press chapter The Social Identity Theory of Intergroup Behavior.

Thorson, Emily A. 2016. "Some of My Best Friends are Poor? Income Misperceptions and Policy Attitudes." Working paper.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185:1124–1131.

Wong, Cara J. 2007. ""Little" and "Big" Pictures in Our Heads Race, Local Context, and Innumeracy About Racial Groups in the United States." *Public Opinion Quarterly* 71(3):392–412.

Wong, Cara, Jake Bowers, Tarah Williams and Katherine Drake Simmons. 2012. "Bringing the person back in: Boundaries, perceptions, and the measurement of racial context." *The Journal of Politics* 74(04):1153–1170.

Yoder, Janice D., Jessica Christopher and Jeffrey D. Holmes. 2008. "Are Television Commercials Still Achievement Scripts for Women." *Psychology of Women Quarterly* 32(3):303–311.

Online Appendix

OA 1 Study 1

Study 1 utilizes data from Ahler and Sood (2016). The results presented in SI I are also part of the appendix in Ahler and Sood (2016), and provided here only for convenience.

OA 1.1 Study 1: MTurk Sample Demographics

Table OA 1.1: Characteristics of the MTurk Sample

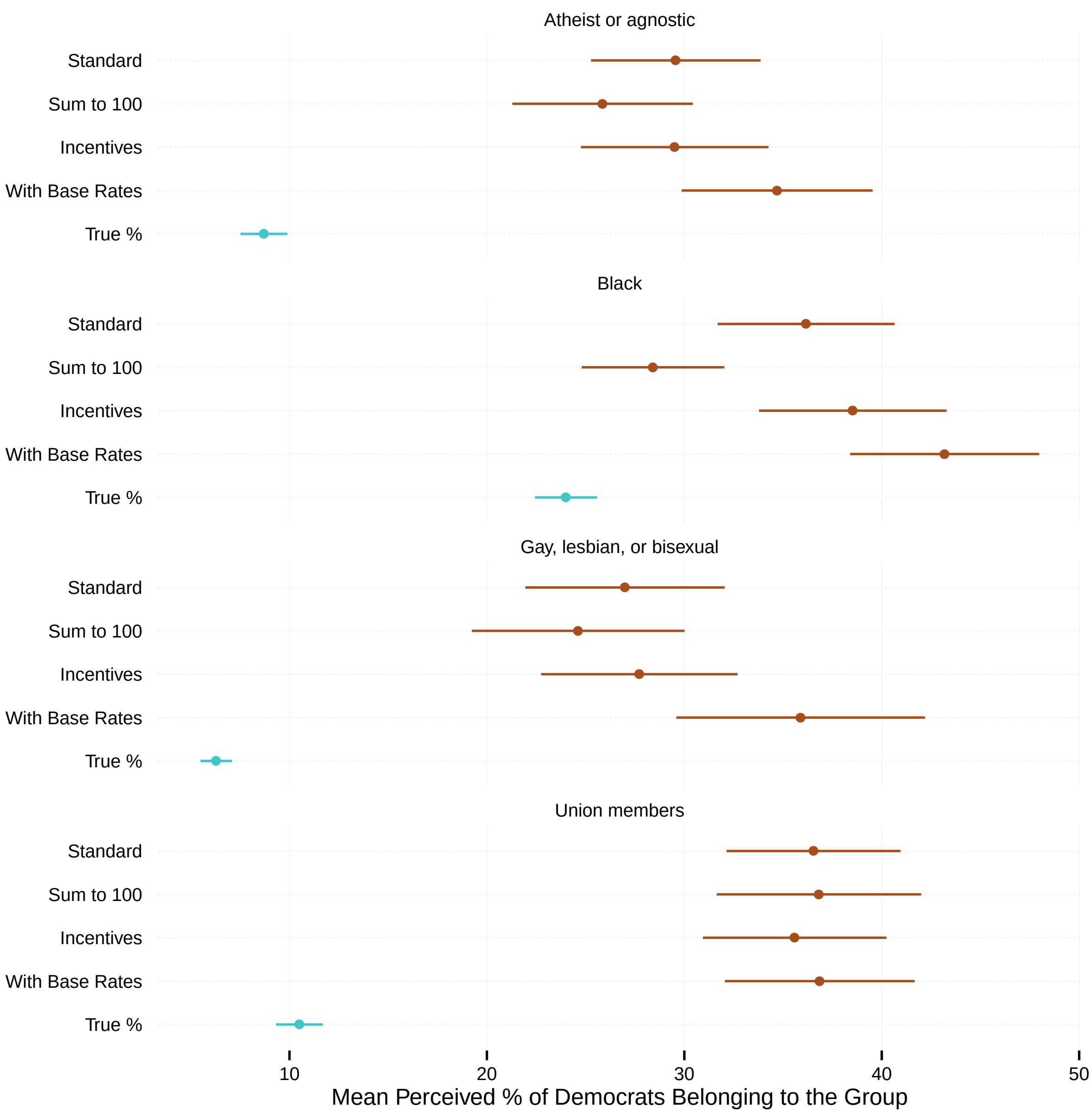
	Sample	Population
Partisanship		
Democratic (inc. leaners)	58.9%	49.0%
Republican (inc. leaners)	22.3%	39.0%
Non-leaning Independent	18.9%	11.9%
Gender		
Female	50.5%	50.9%
Male	49.5%	49.1%
Race/Ethnicity		
White/Caucasian	76.2%	63.7%
Black/African-American	8.1%	12.2%
Asian-American/Pacific Islander	10.1%	4.8%
Native American/Native Alaskan	1.6%	1.1%
Latino/Hispanic	9.7%	16.4%
Education		
Less than high school	0.5%	8.9%
High school diploma (or equiv.)	9.7%	31.0%
Some college	46.1%	28.0%
4-year degree	34.8%	18.0%
Advanced degree	8.9%	9.3%
Age		
18-39	79.1%	39.1%
40-64	19.1%	43.7%
65+	1.8%	17.2%

NOTE: Population estimates come from the 2010 US Census, except for partisanship, which comes from the 2012 ANES.

OA 1.2 Study 1 Results by Item

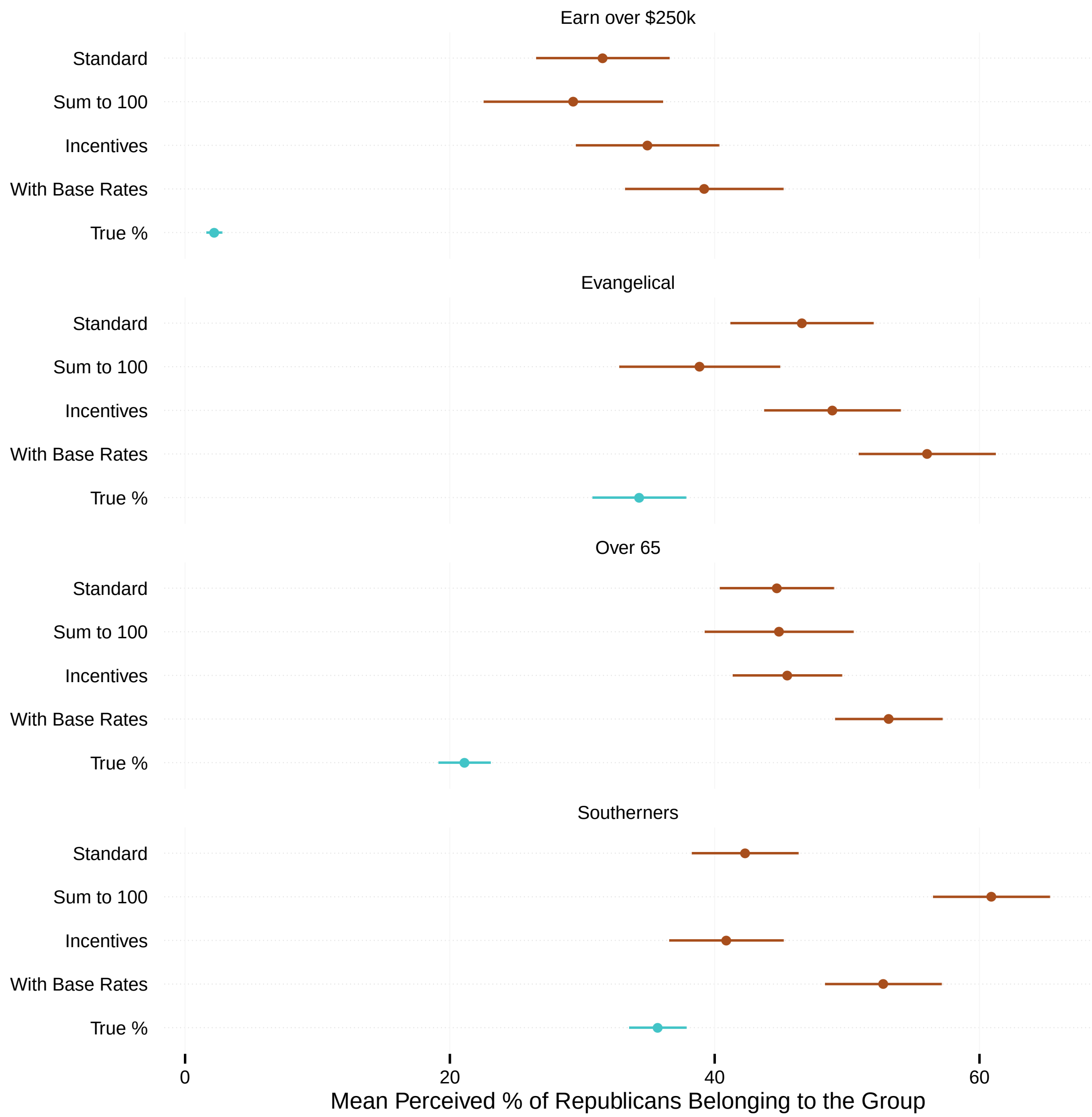
Figure OA 1.1: Expressive Responding, Innumeracy, and Poor Knowledge of Base Rates Don't Explain Apparent Perceptual Errors

(a) Perceived Composition of Democratic Identifiers, by Experimental Condition



NOTE: 95% confidence intervals depicted.

(b) Perceived Composition of Republican Identifiers, by Experimental Condition



NOTE: 95% confidence intervals depicted.

OA 1.3 Kolmogorov-Smirnov Tests

Table OA 1.2: Kolmogorov-Smirnov tests of Distributions

	Incentives		Sum-to-100		Base Rates	
Party-Group	D	p(Same dist.)	D	p(Same dist.)	D	p(Same dist.)
Dem.-Ath./ag.	0.14	0.25	0.11	0.78	0.17	0.12
Dem.-Black	0.08	0.85	0.20	0.12	0.23	0.01*
Dem.-LGB	0.10	0.64	0.19	0.64	0.19	0.06
Dem.-Union	0.09	0.80	0.11	0.8	0.09	0.74
Rep.-$250k+	0.11	0.57	0.09	0.9	0.18	0.07
Rep.-Evangelical	0.12	0.48	0.21	0.08	0.20	0.03
Rep.-Age 65+	0.07	0.96	0.49	0.00***	0.19	0.06
Rep.-Southern	0.07	0.93	0.07	0.93	0.19	0.05

NOTE: Statistics are based on K-S tests comparing the distribution of responses under the condition named at the top of each column to the distribution under the "standard estimation" condition. Asterisks denote statistical significance under the Benjamini and Hochberg (1995) method for controlling the false discovery rate. (Family defined as each set of eight K-S tests comparing responses under two distinct conditions.) * = $p < \alpha^*$ when $\alpha = 0.1$, ** = $p < \alpha^*$ when $\alpha = 0.05$, *** = $p < \alpha^*$ when $\alpha = 0.01$.

OA 1.4 Affect Toward Groups is a Weak Predictor of Perceptions of Share of Groups in Parties

Expressive benefits (rather than misperception) are a potential alternative explanation for the apparent misperceptions we find. If expressive responding explained our findings, we would expect perceptual errors to be associated with how much a respondent likes a group. At the start of the extremity perceptions experiment we had asked respondents to rate the eight groups related to party prototypes on a 101-point feeling thermometer. (We separated these feeling thermometers and the experimental content with a lengthy demographics and political information battery.) If dislike of the groups and the out-party explains our primary descriptive finding, we should find a significant correlation between respondents' feeling thermometer ratings for group g and the reported perceptions of how prevalent g is in party p. However, as the table below shows, we fail to find relationships that are distinguishable from zero for any of the eight group-party dyads. (Note that in this study we swapped one party-stereotypical group, "people over 65" for Republicans, for a counter-stereotypical group, "people under 35.")

Table OA 1.3: Feeling Thermometer Ratings for Groups Don't Predict Perceptions of Share of Groups in Parties

	Reg. coefficient	Std. error	95% conf. interval	n
Democratic Party Groups				
Blacks	-0.01	0.05	[-0.11., 0.10]	297
Union members	-0.04	0.05	[-0.15, 0.06]	297
Gay, lesbian, & bisexual	-0.07	0.05	[-0.17, 0.03]	297
Atheist/Agnostic	-0.03	0.05	[-0.13., 0.06]	297
Democratic Party Groups				
The rich/earn over $250,000	0.01	0.04	[-0.07., 0.10]	659
Evangelicals	-0.02	0.04	[-0.09, 0.05]	659
Southerners	-0.03	0.03	[-0.09, 0.04]	659
The young/people under 35	0.02	0.02	[-0.03, 0.07]	659

Note: The coefficient is from the regression of response to the question, "What percentage of supporters of party p do you think are members of group g?" on feeling thermometer rating of g.

OA 2 Study 2

OA 2.1 MTurk Sample Demographics

Table OA 2.4: Characteristics of the MTurk Sample

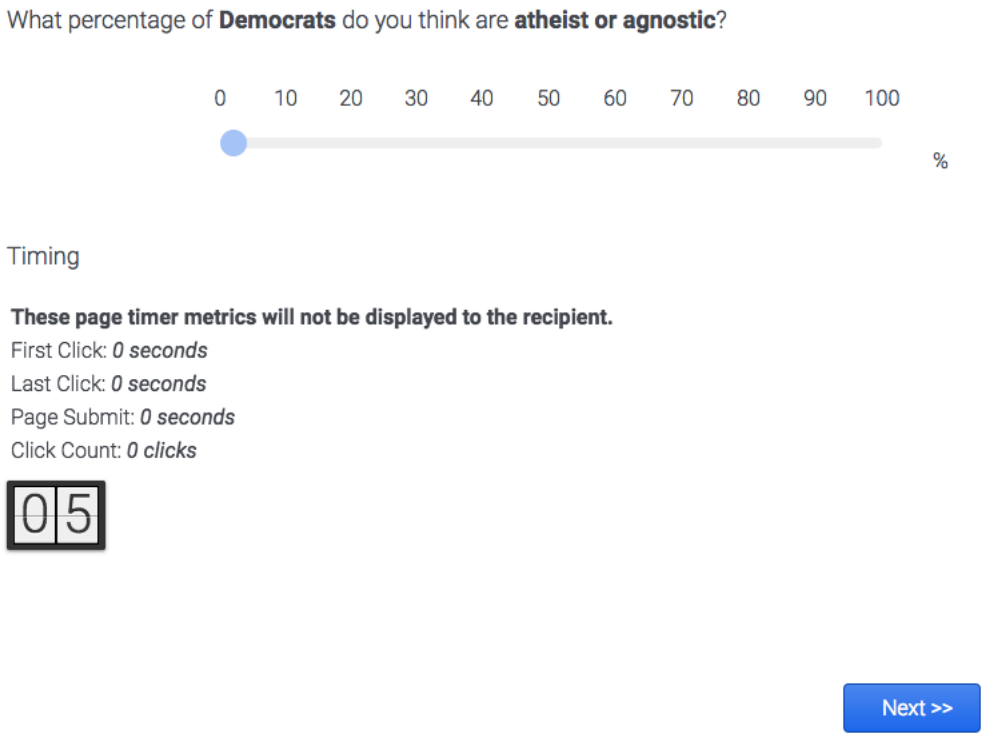
	Sample	U.S. Population
Partisanship		
Democratic (inc. leaners)	56.0%	49.0%
Republican (inc. leaners)	28.0%	39.0%
Non-leaning Independent	16.0%	11.9%
Gender		
Female	50.5%	50.9%
Male	49.5%	49.1%
Race/Ethnicity		
White/Caucasian	78.0%	63.7%
Black/African-American	6.7%	12.2%
Asian-American/Pacific Islander	7.8%	4.8%
Latino/Hispanic	19.4%	16.4%
Age		
18-39	66.8%	39.1%
40-64	30.9%	43.7%
65+	2.4%	17.2%

NOTE: Population estimates come from the 2010 US Census, except for partisanship, which comes from the 2012 ANES.

OA 2.2 Depiction of Study 2 Treatments

Figure OA 2.1: Timing Experiment

(a) "Time Pressure" Condition



(b) "Time Requirement" Condition

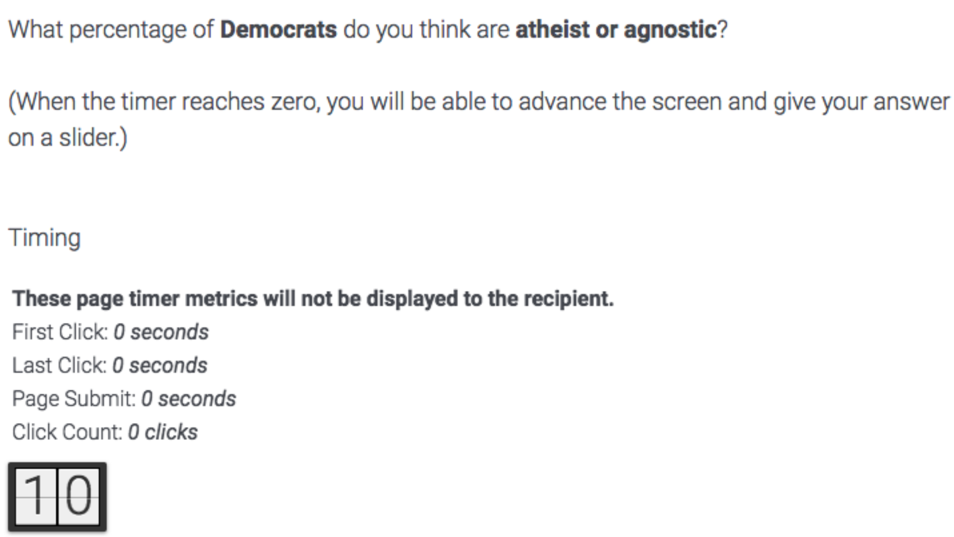
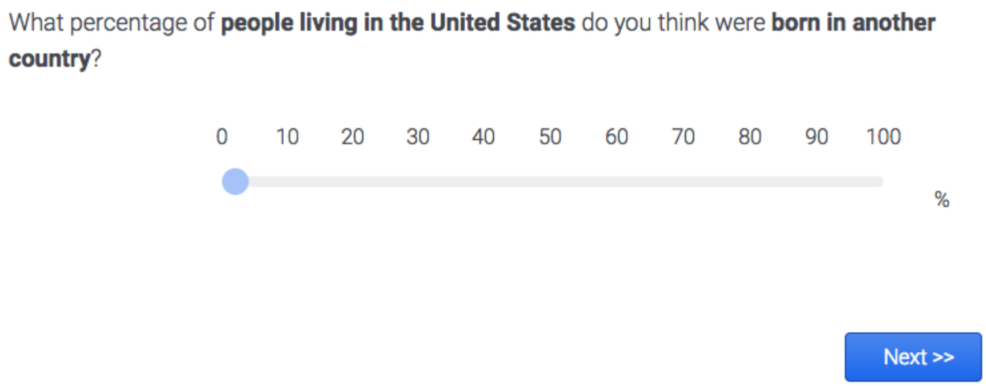


Figure OA 2.2: Wording Experiment

(a) "Percentage" Condition



(b) "Out of 100" Condition

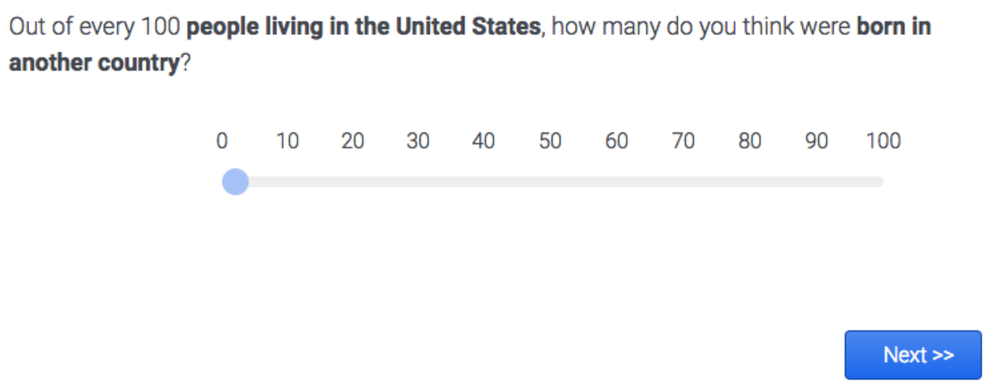
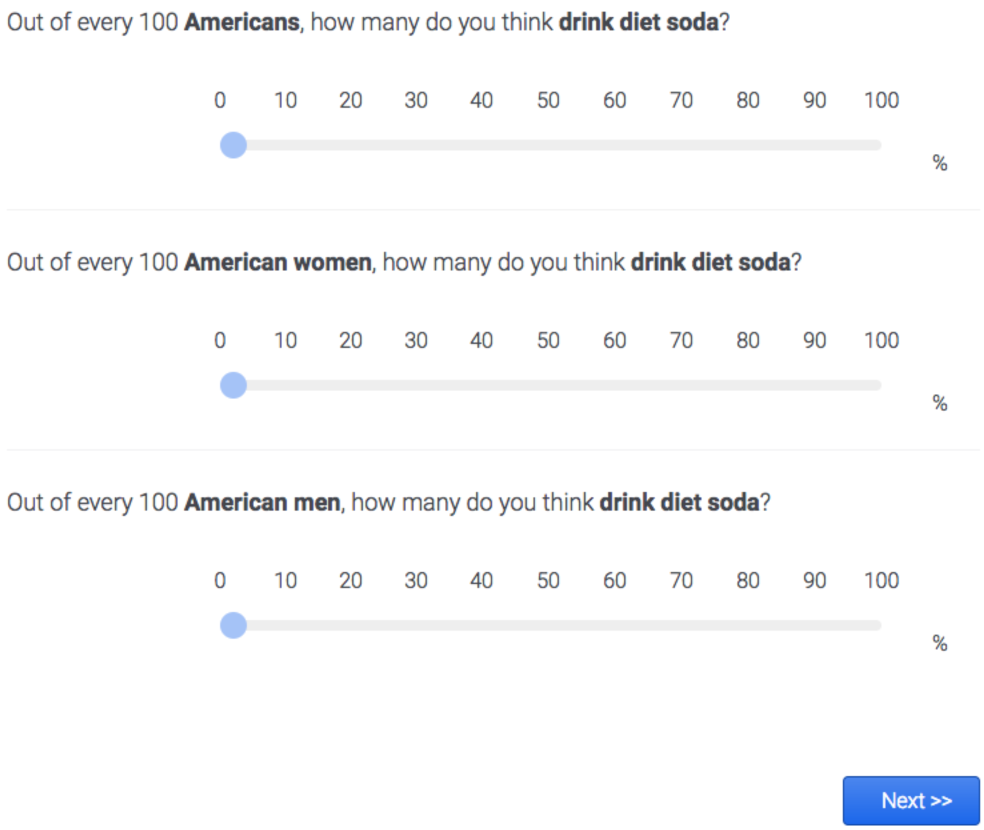


Figure OA 2.3: "Base rates and diet soda" experiment—"Control" condition omits top item


Out of every 100 **Americans**, how many do you think **drink diet soda**?

0 10 20 30 40 50 60 70 80 90 100

%

Out of every 100 **American women**, how many do you think **drink diet soda**?

0 10 20 30 40 50 60 70 80 90 100

%

Out of every 100 **American men**, how many do you think **drink diet soda**?

0 10 20 30 40 50 60 70 80 90 100

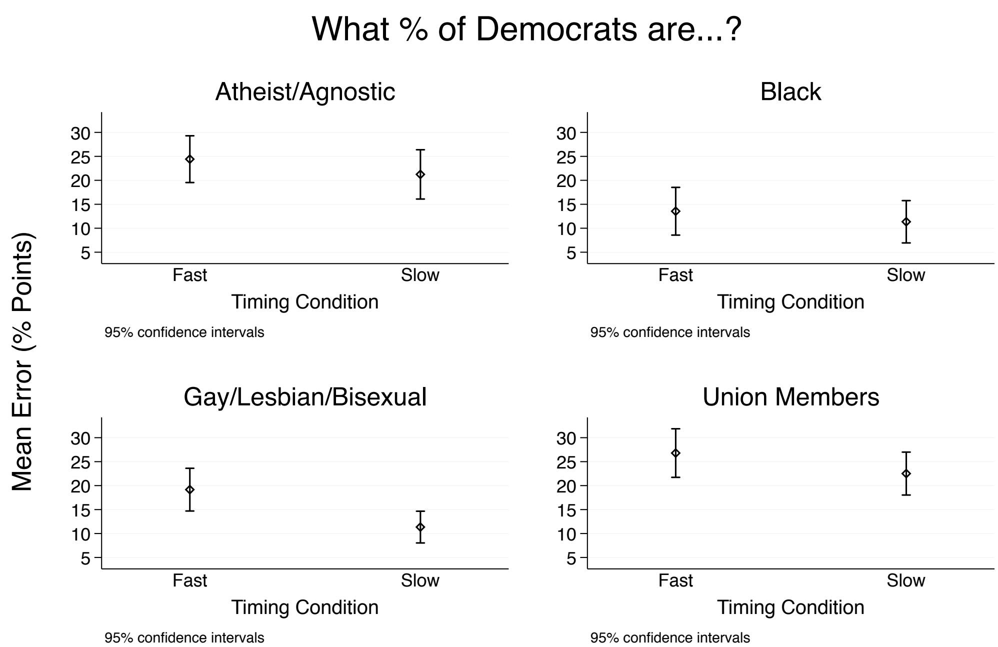
%

Next >>

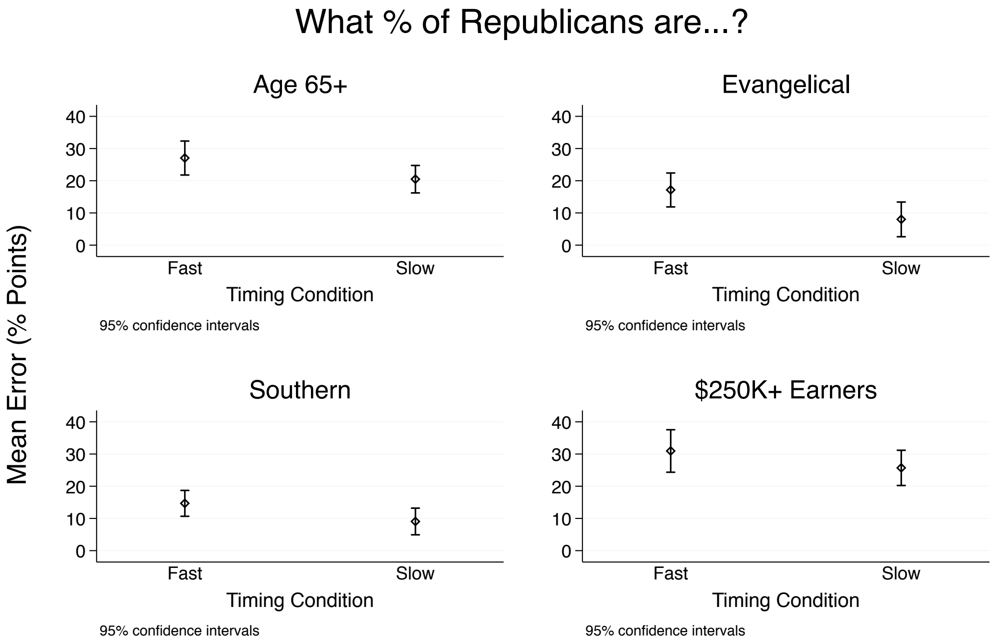
OA 2.3 Timing Experiment: Average Estimates by Item and Condition

Figure OA 2.4: Average Responses (and Perceptual Error) Tend to Be Lower in the "Slow" Condition

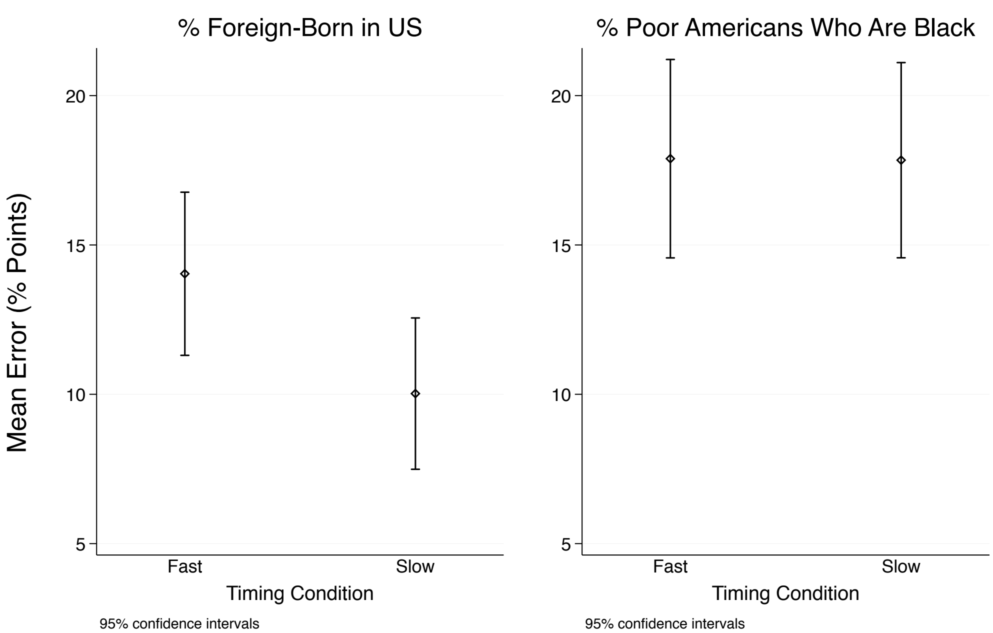
(a)



(b)



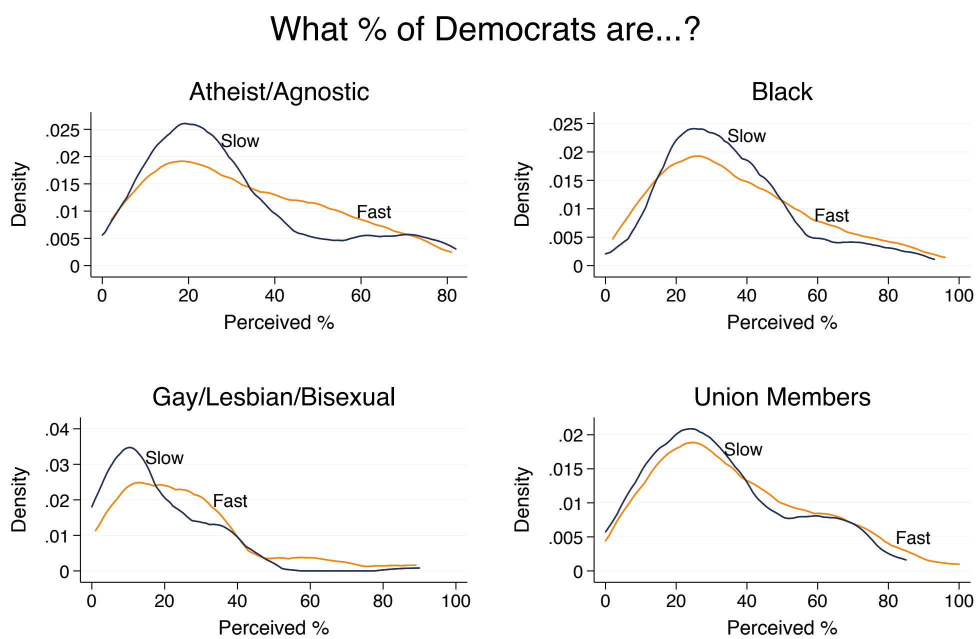
(c) Estimates for the Immigration and Race/Poverty Items



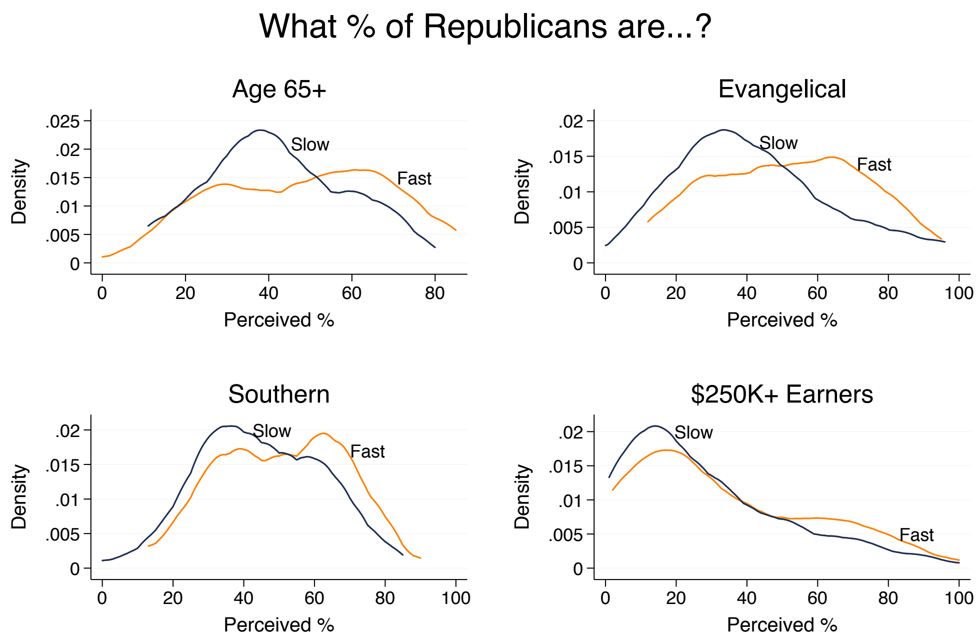
OA 2.4 Timing Experiment: Density Plots by Item and Condition

Figure OA 2.5: Average Responses (and Perceptual Error) Tend to Be Lower in the "Slow" Condition

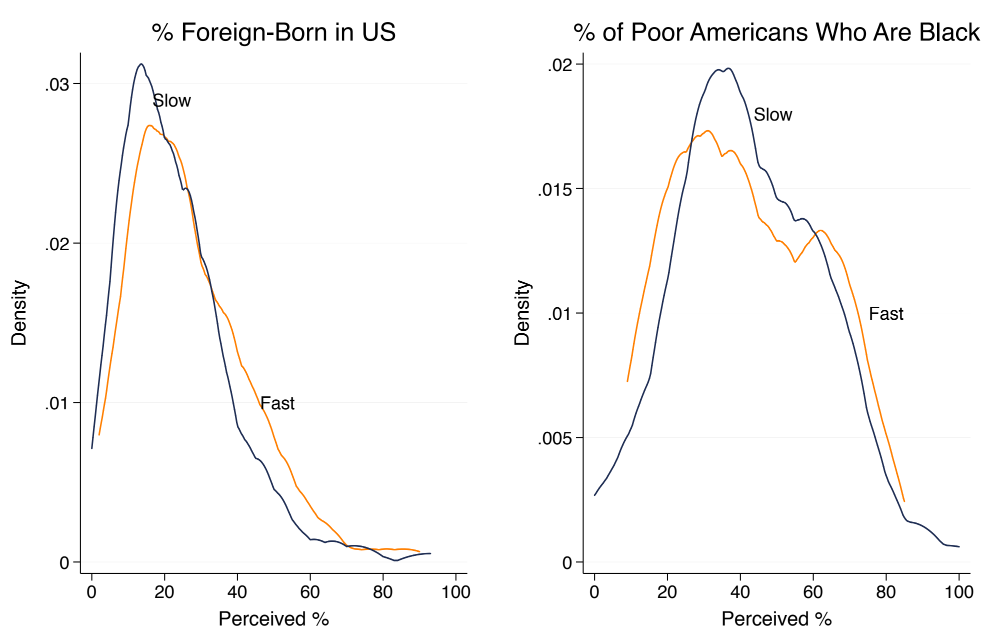
(a)



(b)



(c) Estimates for the Immigration and Race/Poverty Items



OA 2.5 Timing Experiment: A Statistical Test Showing No Evidence of Cheating

Respondents given more time to answer may use the time to consult outside sources. This is undesirable because we want to measure people's perceptions. To assess the extent respondents consulted outside sources, we calculated the percentage of respondents in each condition (slow, or "time requirement," versus fast, or "time pressure") responding with either integer adjacent to the correct response. (E.g., if someone said that 8 or 9 percent of Democrats are atheist or agnostic, they would be "correct" since the correct response is 8.7.) As the table below shows, there are few significant differences. Only in the case of responses to the percentage of Republicans who are from the U.S. South are responses significantly more accurate ($p < .1$) in the "slow" condition. Aggregating across items (including item fixed effects and clustering standard errors by respondent) lends more power, and as the bottom line of the table indicates, we do detect a significant difference across conditions. 1.3% more responses are "on the nose" in the "slow" condition. However, the difference is substantively very small. Even in the "slow" condition, very few responses are close-to-correct—just 3.5% (compared to 2.2% in the "fast" condition). This suggests that even if there is cheating, only 1% of respondents are doing so. Further, the increase is not consistent across all items—three items show a decrease in the "slow" condition, which suggests there was no systematic cheating. More likely, respondents simply became less biased in the "slow" condition as a result of having time to consider the numbers they called to mind.

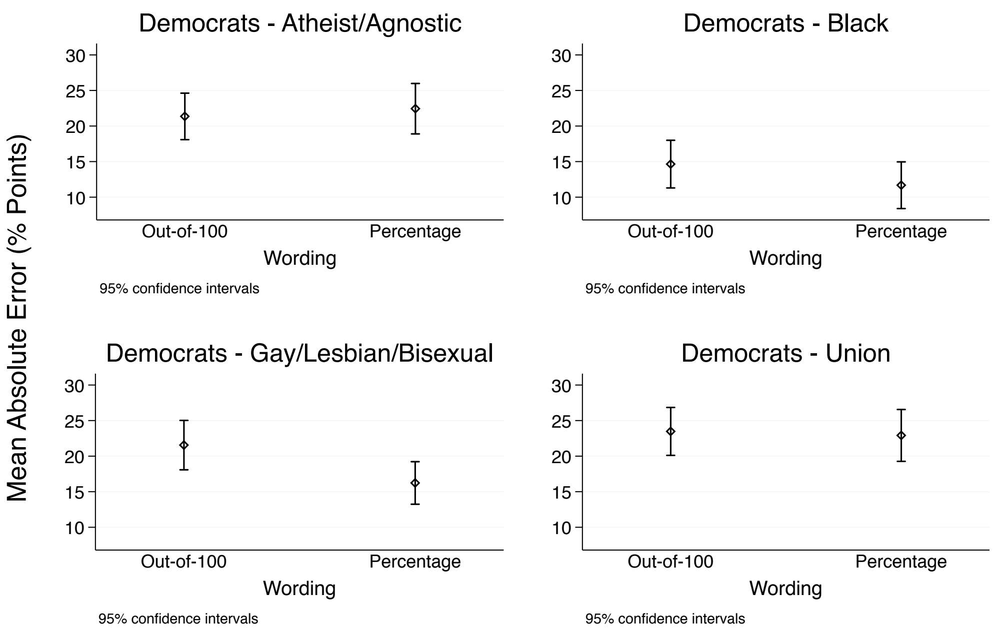
Table OA 2.5: Few Respondents Answer Perceptual Items Correctly, and Differences Across Timing Conditions are Small

	Fast	Slow	Difference	Std. Error of Dif.	n	$P > \|t\|$
% Dems. − Ath./Ag.	.02	.01	-.01	.01	284	.32
% Dems. − Black	.04	.06	.03	.03	284	.27
% Dems. − LGB	.02	.06	.04	.02	284	.13
% Dems. − Union	.02	.04	.03	.02	284	.48
% Reps. − Age 65+	.01	0	-.01	.01	284	.32
% Reps. − Evang.	.02	.03	.01	.02	284	.70
% Reps. − Southern	.01	.05	.04	.02	284	.09
% Reps. − $250K+	.02	.01	-.01	.01	284	.32
% U.S. foreign-born	.02	.05	.03	.02	284	.20
% U.S. poor − Black	.04	.06	.02	.02	284	.40
Across all items	.02	.04	.01	.01	2840	.05

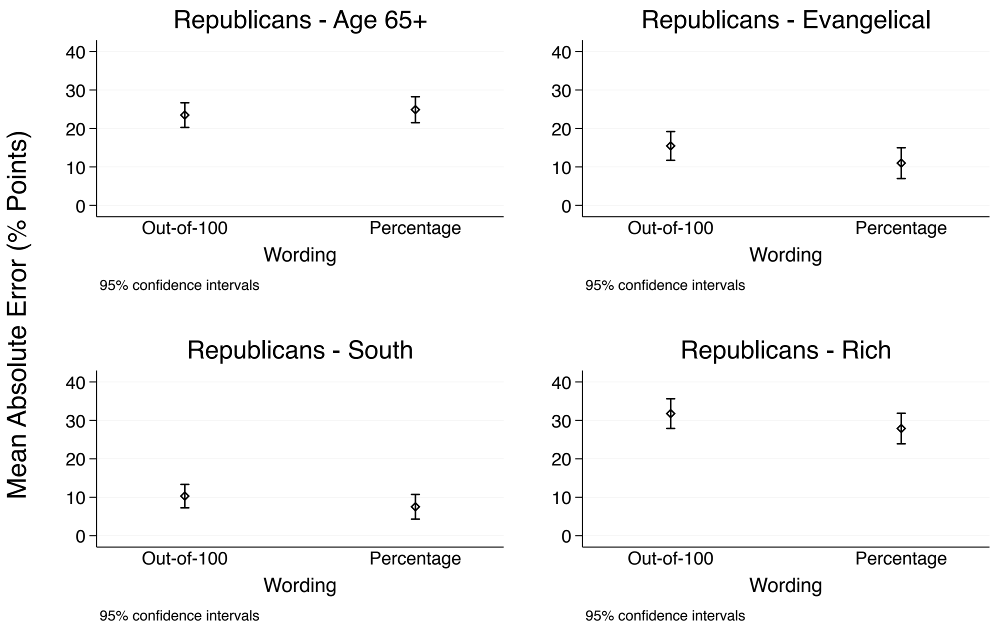
OA 2.6 Wording Experiment: Average Estimates by Item and Condition

Figure OA 2.6: Little Apparent Relationship Between Perceptual Error and "Out of 100" or "What Percentage" Wording

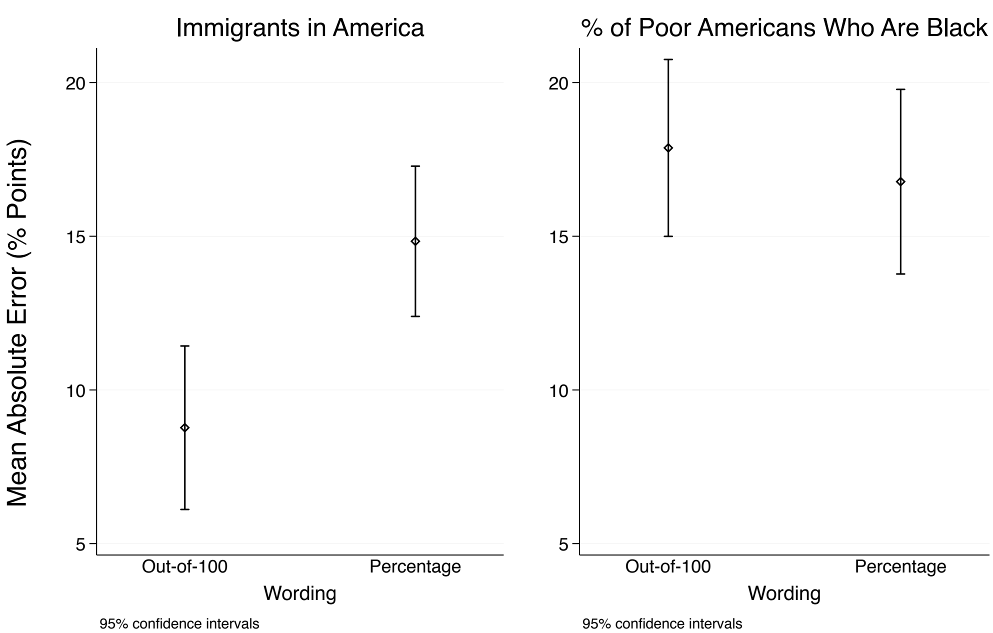
(a)



(b)



(c) Estimates for the Immigration and Race/Poverty Items



OA 2.7 Perceptions of Base Rates of Party-Stereotypical Groups in the Population at Large

In the Study 1's `standard estimation` condition, after measuring party-specific perceptions, we asked respondents to estimate the percentage of the US adult population that belongs to a randomly-assigned subset of the eight party-stereotypical groups. As the table below shows, consistent with previous work (e.g., Wong 2007), respondents tend to overestimate the prevalence of these groups. However, misperceptions do appear to be party-specific: perceptions of the share of these groups in their associated parties are significantly higher than those for the population. And, importantly, the substantive difference between these party-specific and base rate perceptions tend to be quite large.

Table OA 2.6: Comparison of Party-Specific Perceptions to Perceptions of Population Base Rates of Party-Stereotypical Groups

Group	Mean Perceived Base Rate	Mean Perceived Party Rate	Difference
Southerners	32.74%	41.94%	-9.20**
Over 65	30.36%	46.54%	-16.18***
Evangelical	35.5%	49.98%	-14.48***
Earning Over $250K	11.4%	28.6%	-17.19***
Black	31.38%	35.96%	-4.58
Atheists or Agnostics	22.93%	28.04%	-5.11+
Union Members	25.74%	33.52%	-7.78**
LGBT	14.86%	27.33%	-12.47**

+ $p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$

Difference-in-Differences Estimator

To better understand misperceptions about subpopulation composition, we can compute a difference-in-differences:

$$(p(\text{group}|\text{representative party})_P - p(\text{group}|\text{representative party})_A) -$$
$$(p(\text{group})_P - p(\text{group})_A)$$

where P indexes perceived quantities and A indexes actual quantities. While the results in the above table confirm that base rates alone do not drive these perceptions, and that they reflect something about party stereotypes, the D/D estimator can tell us whether one of those factors—misperceived base rates or party stereotyping—more heavily influences the reported beliefs. If the D/D estimate is significantly negative, then the misperception is more a function of base rate error, in the aggregate. By contrast, if the D/D estimate is significantly positive, then the base rate error contributes relatively little to the misperception in the aggregate. The table below presents the results, showing that base rate error contributes less than apparent party stereotyping for five of the eight party-group dyads, and contributes more for none of them.

Note that the quantity being estimated (and presented in the table) is the raw error of the party-specific perception minus the raw error of the base rate perception.

Table OA 2.7: Difference-in-Differences Estimates: Do Misperceptions of Group Base Rates Contribute Significantly More/Less Than Beliefs About Parties to the Misperceptions We Observe?

Party-Group Dyad	D/D Estimate
Republicans—Southerners	7.04 **
Over 65	7.88 **
Republicans—Evangelical	-0.82
Republicans—Earning Over $250K	16.99 ***
Democrats—Black	4.56
Democrats—Atheists or Agnostics	2.10
Democrats—Union Members	8.58 ***
Democrats—LGBT	9.97 **

+ $p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$

OA 2.8 Alternative Specifications for Respondent Error

In the paper, the main dependent measure is *perceptual bias*, average signed error in the respondents' reports across items. Table OA 2.8 presents results for the effect of the treatments on other theoretically relevant variations of the dependent variable: absolute error, average absolute error in respondents' reports across items, and percentage bias, average percentage by which respondents err (signed error) across items.

Table OA 2.8: Effects on Different Theoretically Relevant Dependent Measures

	Timing Experiment			Wording Experiment		
	Perceptual bias (1)	Percentage bias (2)	Absolute error (3)	Perceptual bias (4)	Percentage bias (5)	Absolute error (6)
Timing: slow	-4.34***	-48.1**	-3.58***			
	(1.43)	(20.2)	(1.24)			
Wording: percentage				-1.25	-23.7	-1.10
				(1.26)	(18.0)	(1.05)
Constant	19.7	235.0	22.0	18.9	264.6	22.2
	(1.12)	(15.5)	(1.24)	(1.03)	(13.8)	(0.86)
R^2	0.08	0.47	0.07	0.08	0.49	0.05
n perceptions	1696	1696	1696	3119	3119	3119
n respondents	284	284	284	312	312	312