Ordinal ranking as a method for assessing real-world proportional representations

Crystal Trueborn (dhaussec@indiana.edu) David Landy (dhlandy@gmail.com)
Department of Psychological and Brain Sciences, 1101 E. 10th St.
Bloomington, IN 47405 USA

Abstract

Across two experiments, we use ordinal ranking to examine the processing and representations involved in the estimation of large-scale, real-world proportions. Specifically, in two experiments people estimated two kinds of important real-world proportions: the demographic makeup of their communities, and spending by the U.S. Federal government. Our goal was to assess the metric scaling properties that characterize perceptions of these quantities. In particular, previous work in numerical proportions has posited logarithmic or linear representations (Opfer & Siegler, 2007), or linear representations with task-dependent rescaling (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011). The current context differs markedly from this prior work in that the values we are examining are not explicitly presented to participants, nor directly experienced, but must be estimated on the basis of masses of complex experiences. Ordinal ranking of the quantities, combined with a Thurstonian modeling approach, allows a unique means for estimating the internal scale properties of numerical structures. We find that people largely rely on mixed representations that emphasize log-odds transformations of these vaguely known, but socially important values.

Keywords: numerical reasoning; proportion estimation; probability weighting; mathematical cognition

Introduction

How do people make sense of quantities that go beyond the possibility of direct experience? While people have many systems for dealing with everyday, simple quantities (Feigenson, Dehaene and Spelke, 2004), many socially important quantities go beyond what can easily be experienced directly, and are processed only with difficulty (Resnick, Newcombe, and Shipley, 2017), or by processing intermediate representations (Landy et al, 2013). Here we explore two sets of socially important quantities: demographic proportions in the U.S., and spending in the U.S. federal budget, with the aim of better understanding how information about these vague, large, inaccessible quantities is stored and used.

Demographic proportions--the proportion of people fitting into a particular category--are important factors to consider in weighing of public policy, and their perception has, accordingly, been heavily explored in political science (e.g., Wong, 2007; Kuklinski et al, 2000). For instance, roughly 13% of people living in America are immigrants (about 40 million people). If people massively misunderstand this proportion, they may go on to endorse inefficient government policies that, for instance, address imagined problems of over-immigration. As a result, there is widespread concern about the impact of political ignorance on the voting public (Somin, 2014). Other work has examined perceptions of government spending (e.g., Gilens, 2001). We consider both demographic and budgetary as subtypes of *civic proportions*, and consider the processes potentially involved in their estimation.

When asked to give direct numerical estimates of proportions, people misestimate many critical issues. On average, Americans estimate that 25% of Americans immigrated; (Citrin & Sides, 2008); similarly, people overestimate the proportion of Americans who identify as LGBT (Newport, 2015), Muslim, or who vote (Ipsos SRI, 2014). People across the world make similar errors (Citrin & Sides, 2008; Ipsos SRI, 2014).

Numerical Representation and Processing

Although it is clear that both bias and misinformation (or uncertainty of information) are factors in population estimation data, our most recent research (Landy et al., 2017) suggests that the standard measure--asking people for direct estimates of proportions--may fail to reflect underlying beliefs in a direct way. Instead, decades of research in both psychophysics (Stevens, 1957) and numerical estimation (Opfer & Siegler, 2007) suggest that people systematically transform quantities in ways that significantly affect responses.

All psychophysical models comprise the composition of three mathematical functions (see Figure 1): a perceptual function, a psychological or cognitive function, and a response function. The perceptual function transforms data in the world into some kind of perceived format--which one might call a *mental representation*, though this specific pattern may or may not be neurally realized. Secondly a cognitive function (such as memory decay) is applied to the mental 'representation'. Finally, the result of this cognitive function is transformed by the requirements of a response (such as a button press, vote for a political candidate, Likert scale rating, or ordinal ranking). All the theoretical frameworks we consider assume these three functions, but they account for responses in very different ways.

In a direct estimation task, one assesses the composed function $\Psi \circ c \circ r$. In order to differentiate these theoretical models, it is necessary to decompose the functions. One can *assume* a response function, and use parameter estimation to calculate a best-fitting mental representation--but this representation will be model-dependent and therefore ambiguous. A common alternative, followed here, is to change the task. We replace numerical estimates with a rank-ordering task, which we assume to be executed using a Thurstonian approach.

The specific transformational function applied to encode stimuli depends on the details of the task and situation. In many circumstances, the function is taken to be a compressive one, specifically a logarithmic function . In the context of specifically numerical stimuli, Opfer and colleagues (2007) have suggested that how and when people apply particular transformations depends critically on the situation, the number range, and the students' age and expertise. Taken in the context of demographic proportions specifically, this suggests that numerical estimations may not accurately reflect participants' beliefs and knowledge. However, it is entirely unclear whether and how people might transform such stimuli. We consider three possible psychophysical transformations:

log functions: the oldest proposed form of psychological scaling (Fechner, 1860), log functions have specifically been proposed in the case of children's representations of numbers (Opfer & Siegler, 2007).

log odds functions: Another possibility is that people attend to the log odds of the relevant *proportion*, rather than the value of the quantity itself (Gonzalez & Wu, 1998; Landy et al, 2017). This extends suggestions that numerical estimation tasks rest critically on comparisons between the focal quantity and its complement (Spence, 1990).

identity functions: The simplest possibility, and the default assumption in nearly all investigations of public demographic perceptions, is that people use the literal proportions or values when considering inaccessible quantities. Adults working with typical number representations have also been posited to accurately represent quantity (Opfer & Siegler, 2007) at least over sufficiently large numbers, Landy, Silbert, & Goldin, 2013).

Of course, it may not be the case that every individual uses the same psychophysical transformation, and it may even be that one individual mixes in a continuous way information that has undergone different transformations (Kim & Opfer, 2017). Moreover, it is quite possible that people have substantial biases in information about these socially relevant quantities. It is therefore quite challenging to distinguish different underlying transformation functions.

Rank ordering vs. Direct Estimation

Previous research in civic proportions has almost exclusively used one measure: proportion estimation. Any single measure confounds the impact of perception, cognition, and response. Accordingly, while existing work has postulated, in various forms, log-odds transformations for demographic proportions (Lee & Danileiko, 2017; Landy et al, 2017; see also Gonzalez & Wu, 1998), this assumption has not been tested explicitly. Here we begin providing constraining information by using a different response: *rank ordering* of different items (see Figure 1).

In the rank ordering task employed here, people take a list of items, and rank order them from least to greatest or from greatest to least. In this method, participants use quantitative beliefs, but do not make an explicit numerical judgment an an explicit--and perhaps poorly understood--metric scale (in rank ordering, people must only remember which end of the scale is which). Compared to binary comparisons, an alternative way of bypassing psychometric transformations, many decisions are made with each ranking, so data collection is extremely efficient (Lee et al, 2012, Johnson & Kuhn, 2013).

Rank ordering is well-captured by a Thurstonian model (Luce, 1994). As shown in Figure 2, both the direct estimation model used previously by Landy and colleagues, and the rank ordering mode begin from a common assumption of a internal scale. As is common in Thurstonian models, we assume that 'beliefs' of a person about a quantity are stored as normal distributions in that metric space. Response begin by sampling values from this distribution. The two response forms then diverge: in a Thurstonian rank-ordering model, the values are placed in a monotonic sequence which is then the produced order;

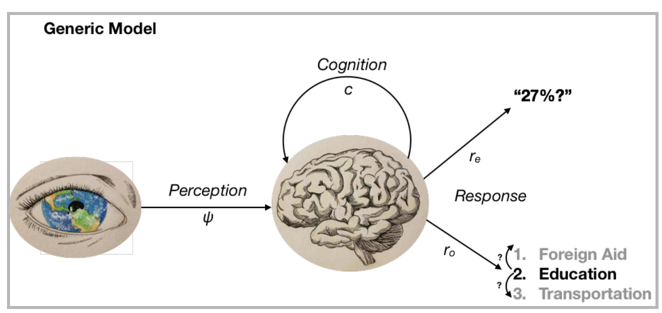

Figure 1: Relationships among various theoretical models considered here. every psychophysical model composes three functions, a perceptual transformation (Ψ), a cognitive updating/storage function, (c), and a response function (r). Extant work in political science typically assumes that numerical responses (e.g., 27%) are identity functions, explaining errors in terms of bias in Ψ and c.

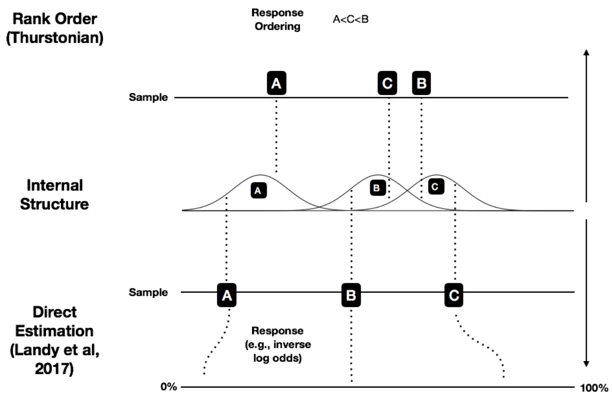

Figure 2: Comparison of different response tasks:Internal magnitudes are represented as normal distributions. On the bottom is the model of direct estimation posited by Landy et al 2017, in which these magnitudes are transformed to responses using a psychophysical response transformation. On the top is the rank-ordering task, in which samples are drawn and ranked. The relative metric positions of the samples can be estimated from how often two items are inverted, e.g., B and C will be inverted more often than A and B, due to greater overlap in their distributions.

numerical estimations are realized by transforming the sample values into the public units (e.g., by inverting the perceptual function used to initially encode them).

The simplicity of the (presumed) response function means that rank ordering provides a powerful technique for examining encoding functions. It might appear as though this method would only provide ordering information, but in fact this is not so. Metric information can be retrieved from a Thurstonian model because inversions are a result of the overlap between distributions.

To sum up, in what follows we present two experiments asking people to rank order civic proportions, and use MCMC sampling in JAGS to fit Thurstonian models assuming equal category variance that contain components for linear, log, and log-odds transformed proportions (plus bias). We evaluate the relative contributions of these transformations to rank ordering across two tasks.

Experiment 1: U. S. Budget Ranking

Materials

Participants: 75 participants were recruited through Amazon Mechanical turk and were compensated $1.25.

Procedure: Participants were shown a list of US 2015 federal budget categories in a unique random order on the survey platform Qualtrics. They were given instructions to rank the 15 items on the list from greatest (rank 1) to least (rank 15). The items listed categories of the US federal budget such as "The percentage of the 2015 U.S. federal budget that was dedicated to social security and unemployment (housing assistance, food and nutrition assistance, etc)" and "The percentage of the U.S. federal budget that was dedicated to foreign aid (international humanitarian assistance, development, etc)"; categories were drawn from the U.S. federal budget functions.

Rank ordering requires remembering correctly which list endpoint is "most," and which "least." Pilot data indicated that a small number of participants indeed confused the correct list ordering despite instructions so additional questions were added in which participants were explicitly asked where they put the "largest" and "smallest" categories. The model used a 'flip' parameter as well, to estimate whether it was more likely that individuals had reversed the intended ordering.

After the main task, people completed the updated 2017 version of Delli Carpini & Keeter (1993), a standard measure of political knowledge. This five-item scale asks basic questions about politics: for example, "which party is more conservative at a national level?", another is "How much of a majority is required for the U.S. Senate and House to override a presidential veto?". This scale is scored from 0-5 with equal weight given to each question. Finally, participants filled out a questionnaire indicating political affiliations, gender, political party, ethnic background, income, and self-reported political involvement.

Analysis: To model contributions of different perceptual transformation functions, we used a mixture model to generate mean internal normal distributions for each item, which combined weights from the 3 transformations:

$$\mu_i = \delta_1 p_i + \delta_2 log(p_i) + \delta_3 log(\frac{p_i}{1-p_i})$$

Here the δ's represent weights distributed as a dirichlet over a simplex, so that the sum of the δ values is one.

Biases were fitted per item, *i*, and distributed with a tight double-exponential (i.e., LASSO) prior. Precisions were kept fixed across items (see general discussion), but were allowed to vary by individual, using a normal distribution at the group level. Political knowledge, which has been implicated in other studies as a predictor of precision of estimates, was also included as a parameter on individual precision (see Marghetis et al, under review).

Results

People were, on average, fairly inaccurate at estimating the US Federal budget items (Figure 3). People over-ranked small items and under-ranked large items, an unsurprising pattern in light of previous psychophysical work showing this general pattern in estimation tasks (Landy et al, 2017). Nevertheless, people *did* show meaningful knowledge of the structure of US expenses. Indeed, the correlation between mean and true rank order was 0.71, while the median individual correlation was 0.48.

Overall, the model confidently estimated only a very small contribution for the identity component, while the components of the log-odds and log models were large and very equal (see Figure 4). This may reflect a balance

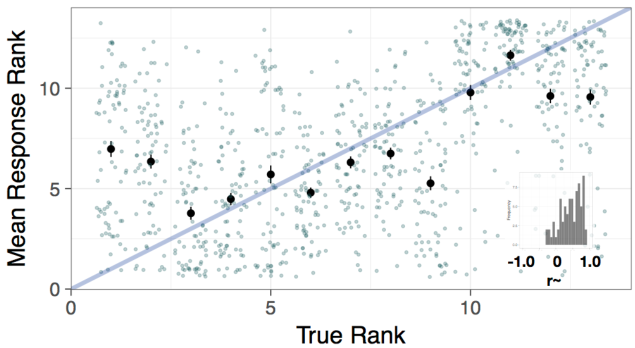

Figure 3: Mean response rank vs. true rank for federal budget expenses in Experiment 1. Perfect correlation is shown by the blue line. The inset shows correlations for each individual participant.

between the two representations, but the large uncertainty bars suggests that the data is simply insufficient to distinguish the two cases in this set.

Posterior predictive checks confirmed that the model Failed to capture significant biases in human judgments on specific items. As has previously been reported, people over-estimated the rank order of spending on foreign aid (Gilens, 2001), but this was not their only error. People, on average, also massively overestimated US spending on agriculture relative to other expenses, and ranked military spending lower than spending on health care/medicare and social security, even though military spending is in fact barely more than half the smaller of these.

Discussion

The model and data indicated that people do perceptually transform their impressions of U.S. federal budget spending, but it did not distinguish the relative contributions of log-style compression of large values relative to low ones, typically found in absolute estimates, and compression of the middle relative to both extremes, often found on proportion judgments. This may be simply because the distribution of budget spending contains no

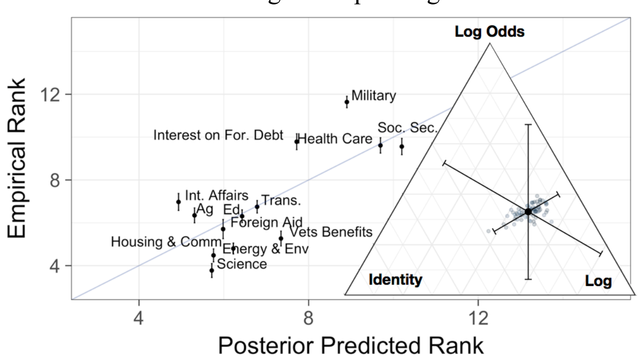

Figure 4: Raw rank means against predicted rank means. Errors are standard errors The posterior estimates of δ values for each model component are shown in the inset in barycentric coordinates.

very large proportions (the largest is social security, which occupies around one third of the federal budget). Another complicating factor is that people on average misranked several items, suggesting meaningful errors in people's errors, which may contribute to systematic biases in estimations of internal parameters. To further explore the perceptual transformation function, Experiment 2 had people rank order demographic proportions which, because they overlapped, could contain many very large, small, and middle-sized proportions; separate work suggested that people had, roughly speaking, relatively unbiased estimates of these values (Marghetis et al, under review).

Experiment 2: U. S. Demographics

Materials

Participants: 125 participants were recruited through Amazon Mechanical turk compensated $1.25.

Procedure: Participants were shown a list of US subpopulations in a unique random order on the survey platform Qualtrics. They were given instructions to rank the 30 items on the list from least to greatest The items listed titles of US subpopulations such as "Percentage of Americans that identify as Japanese" and "Percentage of American workers that make less than $30,000 per year".

As mentioned above, rank ordering requires remembering correctly which list endpoint is "most", and which "least" We again used a 'flip' parameter, to estimate whether it was more likely that individuals had reversed the intended ordering. After the main task, people completed the updated 2017 version of Delli-Carpini & Keeter (1993).

After the task, participants filled out a debriefing questionnaire indicating political affiliations. In this questionnaire we recorded for name, gender, political party, ethnic background, income, and perceived political involvement. We do not consider these factors here.

Figure 5: Rank orderings given by subjects were plotted against the true rank order of the 28 items. Perfect rank order is shown by the blue line. Two U.S. subpopulations (items 9 and 10) had identical population statistics of 13.3% and were spaced in order to show response distribution for each item. Inset shows a histogram of individual-level correlations.

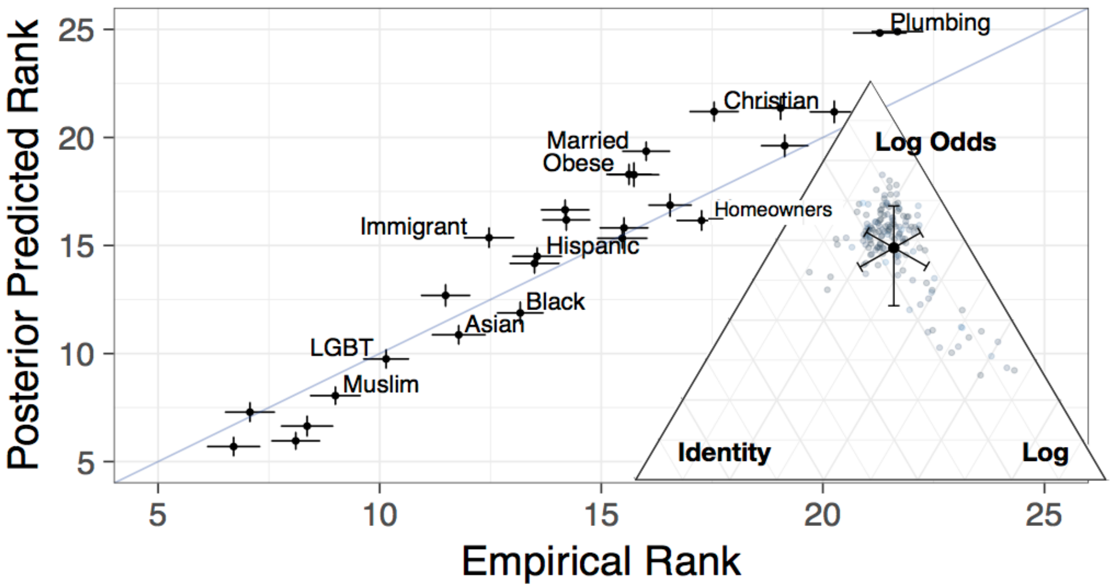

Figure 6: Raw rank means against predicted rank means. Errors are standard errors The posterior estimates of δ values for each model component are shown in the inset in barycentric coordinates. The barycentric inset shows the per-individual weighting of three model components.

Results

People were substantially more accurate at ranking U.S. demographics than budget items. Generally, people centered around the correct ranking, with some tendency to over-rank low items and under-rank high items (Figure 5), in line with other data sources (e..g, Maghetis et al, under review). The median correlation between the true ordering and the response ordering was 0.72, suggesting a moderately strong degree of knowledge about these items.

Figure 6 shows the primary results of the model-fitting. The model strongly favored the log-odds model, although a few individual participants, were, best fit by the log model.

Discussion

The results confirm and extend the results of Experiment 1. Like Experiment 1, these results suggest that literal, untransformed values play little role in guiding ranking judgments. However, the presence of very large proportions let us distinguish the upper end, where we found that people where quite sensitive, in a manner more in line with log-odds than log transformations; at the same time, the best supported models involved a mixture of multiple components, and a small number of individuals were much better fit by a log-heavy component, suggesting that individuals may vary in their processing strategies.

Overall bias on these items was much smaller and less systematic than that for the budget items of Experiment 1. Some salient trends were that people overestimated marriage rates, obesity, and immigration, while underranking the LGBT-identifying proportion, and the proportion of the population that is Black and Asian. However, it is worth remembering that this is not a representative sample, and that individual characteristics such as respondent race and have a large effect on how people make demographic estimates (Wong, 2007; Guay, Wong, & Landy, in prep). Nonetheless, it is worth noting that here people underestimated some demographics, like LGBT rates, that have previously been identified as the target of strong overestimation (Gallup, 2014). This accords with separate reports suggesting that previous accounts that do not take psychophysical transformations into account may misconstrue the nature of social biases that affect demographic and other quantitative estimations (Landy et al, 2017; Brower et al, under review).

General Discussion

Two studies used rank ordering to evaluate the perceptual manifold used in evaluating civically important but vaguely known quantities. Elements of the federal budget were well matched by either a log or log-odds model, while demographic data that covered a larger range of true values was uniquely captured by a mixture emphasizing the log-odds model. Taken together, these results support the conclusion that log-odds are useful ways to capture proportional judgments.

Substantial recent work has gone into exploring mental representations of quantities using line production and estimation techniques (Opfer & Siegler, 2007; Barth & Paladino, 2011; Dale Cohen & Blanc-Goldhammer, 2011), debating whether children’s responses indicate log scaling of numerosity, or proportional reasoning. Children’s scaling of number is very different than adults estimates of quantities, and no firm conclusions can be drawn; nevertheless, our approach suggests a clear reconciliation under the banner that proportional judgment models, such as those proposed by Barth & Paladino (2011) are naturally conceptualized as the result of comparing logs of odds. That is, it may be both true that people in our experiments consider proportions (as odds), *and* that they log scale those proportions, in reaching judgments.

One caveat is worth noting: the internal metric is itself fundamentally undefined: there is no *natural* measure over that scale. The technique we use--Thurstonian modeling--presumes equal precision of each knowledge component. We are finding the perceptual transformation under which all items in our set are equally well-known. It may of course turn out that this assumption fails to be true over these real-world items. Future work should evaluate these representational conclusions in lab situations which can better control the objective experiences of participants.

Ordinal ranking is a method worthy of attention. Despite widely studied errors in numerical estimates of demographics, the rank ordering task reveals that people have substantial implicit knowledge of demographic proportions, and that some of the observed error in budget knowledge may stem from failures to appreciate the importance of perceptual and response scaling functions in shaping judgment patterns. We feel that the direct estimation paradigm common in the literature may actually impede progress in understanding public knowledge. Rank ordering is a way to quickly collect data, which we have shown can be usefully analyzed using Thurstonian models.

More broadly, competent engagement in civic society depends on access to correct--at least roughly--information. Ameliorating widespread misinformation depends in part on our ability to assess people's beliefs, and to communicate with people about them.. Both of these goals are enhanced to the degree that we can understand the core cognitive constructs that undergird people's interactions with the tenuously experienced quantitative information that makes up so much of the civic environment.

Acknowledgments

Thanks to Brian Guay, Tyler Marghetis, Mitali Nag, Atrayee Mukherjee, and Eleanor Brower for making significant suggestions on this work as it unfolded. Thanks to Sam Robinson for drawing sketches for the figures.

References

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Sci*, *14*(1), 125-135

Brower, E. & Landy, D.. (under review) Bias in the self-knowledge of communities. *Proc of the Cog Sci Society.*

Citrin, J., & Sides, J. (2008). Immigration and the imagined community in Europe and the United States. *Political Studies, 56*, 33-56.

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic bulletin & review*, *18*(2), 331-338.

Delli Carpini, M., & Keeter, S. (1996). *What the American public knows and why it matters* (Yale University Press, New Haven, CT)

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cog. Sci.* , *8*(7), 307-314.

Gilens, D. (2001). Political ignorance and collective policy preferences. *American Poli Sci Review, 95,* 379-396.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*(1), 129-166.

Ipsos Social Research Institute. (2014). *Ipsos MORI Perils of Perception Survey.*

Johnson, T. R., & Kuhn, K. M. (2013). Bayesian Thurstonian models for ranking data using JAGS. *Behavior research methods, 45*(3), 857-872.

Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. *Developmental Psychology*, *53*(6), 1088.

Kuklinski, J. F., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R-F. (2000). Misinformation and the currency of democratic citizenship. *Journal of Politics, 62,* 790-816.

Landy, D., Guay, B., & Marghetis, T. (2017). Bias and ignorance in demographic perception. *Psychonomic bulletin & review,* 1-13.

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive science*, *37*(5), 775-799.

Danileiko, I., & Lee, M. D. (2017). A Model□Based Approach to the Wisdom of the Crowd in Category Learning. *Cognitive science.*

Lee, M. D., Steyvers, M., De Young, M., & Miller, B. (2012). Inferring expertise in knowledge and prediction ranking tasks. Topics in cognitive science, 4(1), 151-163.

Marghetis, T., Guay, B, & Karlapudi, A, Landy, D. (under review) [[*Title here]]. Proc of the Cog Sci Society.*

Newport, F (2015). Americans Greatly Overestimate Percent Gay, Lesbian in U.S. *Gallup.* Retrieved from http://www.gallup.com/poll/183383/americans-greatly-overestimate-percent-gay-lesbian.aspx.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive psychology, 55*(3), 169-195.

Resnick, I., Newcombe, N. S., & Shipley, T. F. (2017). Dealing with big numbers: Representation and understanding of magnitudes outside of human experience. *Cognitive science*, *41*(4), 1020-1041.

Somin, I. (2014). Why political ignorance undermines the wisdom of the many. *Critical Review, 26,* 151-169.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal Experimental Psychology: Human Perception & Performance, 16,* 683-692.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153-181.

Luce, R. D. (1994). Thurstone and sensory scaling: Then and now. *Psychological Review 101*(2), 271-277.

Wong, C. 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*, 392-412.