# Modeling Judgment Errors in Naturalistic Numerical Estimation

**Wanling Zou (wanlingz@sas.upenn.edu)**
Department of Psychology

**Sudeep Bhatia (bhatiasu@sas.upenn.edu)**
Department of Psychology, Wharton Marketing
University of Pennsylvania
Philadelphia, PA 19104, USA

## Abstract

We quantitatively modeled and compared two types of errors in numerical estimation for naturalistic judgment targets: *mapping errors* and *knowledge errors*. *Mapping errors* occur when people make mistakes reporting their beliefs about a particular numerical quantity (e.g. by inflating small numbers), whereas *knowledge errors* occur when people make mistakes using their knowledge about the judgment target to form their beliefs (e.g. by overweighting or underweighting cues). In two studies, involving estimates of the calories of common food items and estimates of infant mortality rates in various countries, we found that knowledge error models predicted participant estimates with very high out-of-sample accuracy rates, significantly outperforming the predictions of mapping error models. The knowledge error models were also able to identify the objects and concepts most associated with incorrect estimates, shedding light on the psychological underpinnings of numerical judgment.

**Keywords:** judgment errors; numerical estimation; word embeddings; word vectors; knowledge representation; cognitive model

## Introduction

Decades of research on judgment and decision making has established that people make systematic errors when estimating numerical quantities, such as the frequencies of lethal events, proportions of demographic groups, or the calories of food items (Chernev & Chandon, 2011; Landy, Guay, & Marghetis, 2017; Lichtenstein et al., 1978). Researchers studying these judgment errors have identified a number of factors responsible for numerical mis-estimation, such as the use of non-linear weighting functions (e.g. Gonzalez & Wu, 1999; Hollands & Dyre, 2000; Landy et al., 2017; Tversky & Kahneman, 1992) or the use of heuristic cue-aggregation rules (Brown & Siegler, 1993; von Helversen & Rieskamp, 2008).

These factors can be understood in terms of two types of errors – *mapping errors* and *knowledge errors*. *Mapping errors* occur when people make mistakes reporting their beliefs about a particular numerical quantity. In other words, people may have the correct belief about the numerical quantity but incorrectly map this belief into a response. For example, prior literature on numerical estimation has found an inverse-S-shape pattern when plotting participant estimations against objective statistics, with overestimation of small values and underestimation of large values (e.g. Erlick, 1964; Hollands & Dyre, 2000; Landy et al., 2017; Varey, Mellers, & Birnbaum, 1990). Such patterns have typically been modeled using non-linear functions, e.g. power functions and their variants (Curtis, Attmeave, & Harrington, 1968; Hollands & Dyre, 2000), log-odds transformations (Shepard, 1981; Zhang & Maloney, 2012), and probability weighting functions (Fennell & Baddeley, 2012; Tversky & Kahneman, 1992). These models all assume that a systematic distortion takes place when transforming correct internal beliefs into an explicit numerical response.

In contrast, *knowledge errors* occur when people make mistakes using their knowledge about the judgment target to form their beliefs. These can lead to the formation of incorrect beliefs (e.g. through the biased use of memory cues), though people may still be able to accurately report these beliefs. For example, Chernev and Chandon (2011) have documented halo biases in food calorie estimation, according to which health-related cues are given an incorrectly high weight, which can then lead to the underestimation of food calories. Media coverage or word frequency has also been shown to be used as a cue in probability estimation (Dougherty, Franco-Watkins, & Thomas, 2008; Tversky & Kahneman, 1974) and frequency estimation (Hertwig, Pachur, & Kurzenhäuser, 2005; Lichtenstein et al., 1978), which can lead to the overestimation of the size of minority groups (Gallagher, 2003; Herda, 2013). More generally, many researchers in cognitive psychology have proposed that people use heuristics to weigh and aggregate judgment cues, which can, at times, lead to systematic errors in numerical estimation. These heuristics simplify the decision process by ignoring certain cues (and thus assigning them incorrectly low weights), or by using equal weights for all cues (and thus over-weighting irrelevant cues and under-weighting relevant cues) (see Hertwig, Hoffrage, & Martignon, 1999; Juslin, Olsson, & Olsson, 2003; von Helversen & Rieskamp, 2008). Our division of numerical judgment errors into mapping and knowledge errors has precedent. For example, Lichtenstein et al. (1978) suggested that there are two types of biases in frequency estimation – a primary bias (i.e. overestimation of small numbers and underestimation of large numbers) and a secondary bias (which may due to media bias, disproportionate exposure, imaginability, etc.). Likewise, Brown and Siegler (1993) argued that there are two types of knowledge that come into play in quantitative estimation – metric knowledge (e.g. statistical induction) and mapping knowledge (e.g.

heuristics). Von Helversen and Rieskamp (2008) extended this work by showing that people are likely to sample objects that are similar to the judgment target (where *knowledge errors* may occur) and make estimation based on some transformation of the sampled objects' values (where *mapping errors* may occur). Finally, Landy et al. (2017) showed the presence of two features that lead to errors in demographic estimation – domain-general bias (i.e. a log-odds mental representation of proportion) and domain-specific bias (e.g. media bias and xenophobia).

Although this work has greatly expanded our understanding of numerical estimation, most of it pertains to estimates of simple frequencies, rather than more general (and complex) numerical quantities. Additionally, experiments that do examine such complex numerical estimates, typically use artificial experimental stimuli – such as toxicity of fictional bugs (Juslin et al., 2003), percentage of dots in a mixture of black and white dots (Varey et al., 1990) and proportion of letters (e.g. "A") in a random letter string (Erlick, 1964) – and/or experimenter-generated cues that provide only an abstract representations of the rich knowledge present in the human mind (Brown, 2002; Juslin et al., 2003; von Helversen & Rieskamp, 2008). Although artificial stimuli and simplified knowledge representations help establish theoretical foundations, it is also necessary to model how people make quantitative estimates in the real world where they usually possess rich and complex knowledge and apply it at their discretion.

Our goal in this paper is to address these issues by formalizing, fitting, and comparing mapping and knowledge error models of numerical estimation with policy-relevant consequences. We consider two domains: estimates of food calories and estimates of infant mortality rates in countries. For the mapping error model, we drew insights from prior work and fit various non-linear functions to predict participant estimates from correct answers. For the knowledge error model, we used word embeddings – models trained on large text corpora that preserve semantic knowledge of words and phrases in high-dimensional vectors – to obtain rich quantitative representations for food items and countries, and then attempted to predict participant estimates from these representations by implicitly learning cue weights on semantic knowledge. Word embeddings have been shown to mimic representations at play in human cognition, such as associative judgment (Bhatia, 2017; Caliskan, Bryson, & Narayanan, 2017), free recall in memory (Manning & Kahana, 2012; Steyvers, Shiffrin, & Nelson, 2004), priming (Günther, Dudschig, & Kaup, 2016), and stereotypes (Garg, Schiebinger, Jurafsky, & Zou, 2018). Thus these representations are likely to capture common knowledge about the judgment targets that may hinder or facilitate numerical estimation. More importantly, these representations will offer insights into the psychological qualities and cues that most contribute to over- and underestimation.

# Experimental Methods

## Participants

We recruited a total of 101 participants – 50 participants (mean age = 30 years, 52% were female) in Study 1 and 51 participants (mean age = 31.4 years, 60.78% were female) in Study 2 from Prolific Academic, an online experiment platform. All participants were from the U.S. and had an approval rate of 80% or above. They were paid at a rate of approximately $6.50 per hour.

## Stimuli

For Study 1, we obtained 200 food items and their true calorie amounts from a United States Department of Agriculture (USDA) database. Sample items include lamb, butter, mint, etc. For Study 2, we obtained the infant mortality rates of 200 countries from the Central Intelligence Agency (CIA)[1]. These countries include Denmark, Nepal, Estonia, etc.

## Procedure

In Study 1, participants were asked to estimate how many calories (in kcal) there are in 100 grams of a particular food item; in Study 2, they were asked to estimate the infant (child under 1 year old) mortality rate in number of deaths per 1,000 live births in a particular country. Each participant estimated all 200 stimuli and saw only one item on each screen. The order of the 200 stimuli was randomized and there was a 30-second break after every 50 stimuli. After completing all questions, participants were asked for their age and gender.

# Predicting Estimates

## Computational Methods

For each target $i$ (e.g. peanuts), we obtained both the average participant estimate $y_i$ (e.g. estimated calories in peanuts) and the correct answer $z_i$ (e.g. actual calories in peanuts). To quantitatively study mapping explanations for these errors, we fit three different mapping models that transformed correct answers into participant responses. Formally, our mapping error models predicted $y_i$ as some function (linear or nonlinear) of $z_i$. The first function we used was a simple linear function (Eq.1); the second function was a third-degree polynomial (Eq.2); and the third function was a power function with a constant term (Eq.3)[2]. Parameters were estimated by minimizing the residual sum of squares.

$$y_i = \beta_0 + \beta_1 z_i \tag{1}$$

$$y_i = a z_i^3 + b z_i^2 + c z_i + d \tag{2}$$

[1] There are 224 countries and regions in CIA database. We excluded the ones that do not have a vector representation in our word embedding model (see the computational methods in the next section) and those whose public data are limited (e.g. no electricity usage data, no literacy rates, etc.)

[2] Although linear pattern was rarely found in previous literature, we included the linear model here as a baseline. A third-degree polynomial served to model any potential S-shape or inverse-S-shape pattern. We incorporated a power function due to its prevalence in prior work.

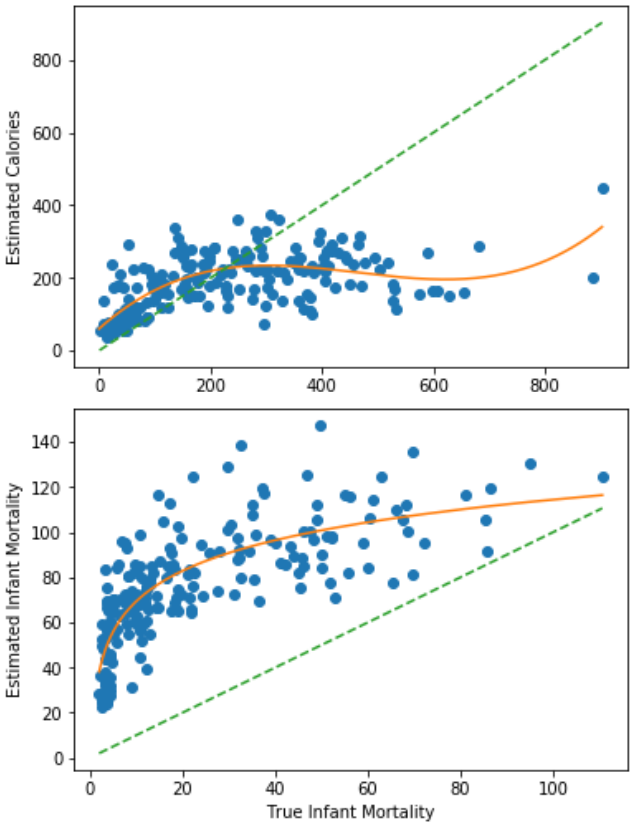


Figure 1: Scatterplots of average participant estimates vs. correct answers for Study 1 (top) and Study 2 (bottom). The green dotted lines indicate perfect calibration where participant estimates are equal to the correct answers. The red curves represent the best fitting mapping error models – third-degree polynomial (Eq.2) for Study 1 and power function (Eq.3) for Study 2.

$$y_i = \lambda z_i^{\delta} + \gamma \tag{3}$$

To examine knowledge errors, we used pretrained Word2Vec word embeddings (Mikolov et al., 2013) to obtain rich quantitative representations for food items and countries. These gave us a 300-dimensional vector $x_i$ for each target $i$. Our knowledge error model involved fitting a (regularized) linear function with a 300-dimensional weight vector $w$ ($w_1$ for Study 1 and $w_2$ for Study 2), to predict $y_i$ using $w * x_i$[3]. The weight vectors (300-dimensional) transform semantic knowledge represented in a 300-dimensional space to an one-dimensional numerical estimation line. Intuitively, each dimension of $x_i$ can be seen as a semantic cue that participants might rely on to facilitate estimation and these weight vectors can be seen as capturing cue weights on semantic knowledge. We compared mapping and knowledge error models through leave-one-out cross-validation (LOOCV)[4], on both the aggregate and the individual level.

[3]Specifically, we implemented a ridge regression in the Scikit-Learn Python machine learning library (Pedregosa et al., 2011). There was a set of hyperparameters in this library. To avoid manipulating the hyperparameters to improve model performance, we took the default values of all these hyperparameters. We focused on ridge regression because previous results (e.g. Bhatia, 2019; Richie, Zou, & Bhatia, 2018, Dec) suggested that compared to other models such as lasso and support vector regression, ridge regression often works best in predicting human judgments from word embeddings.

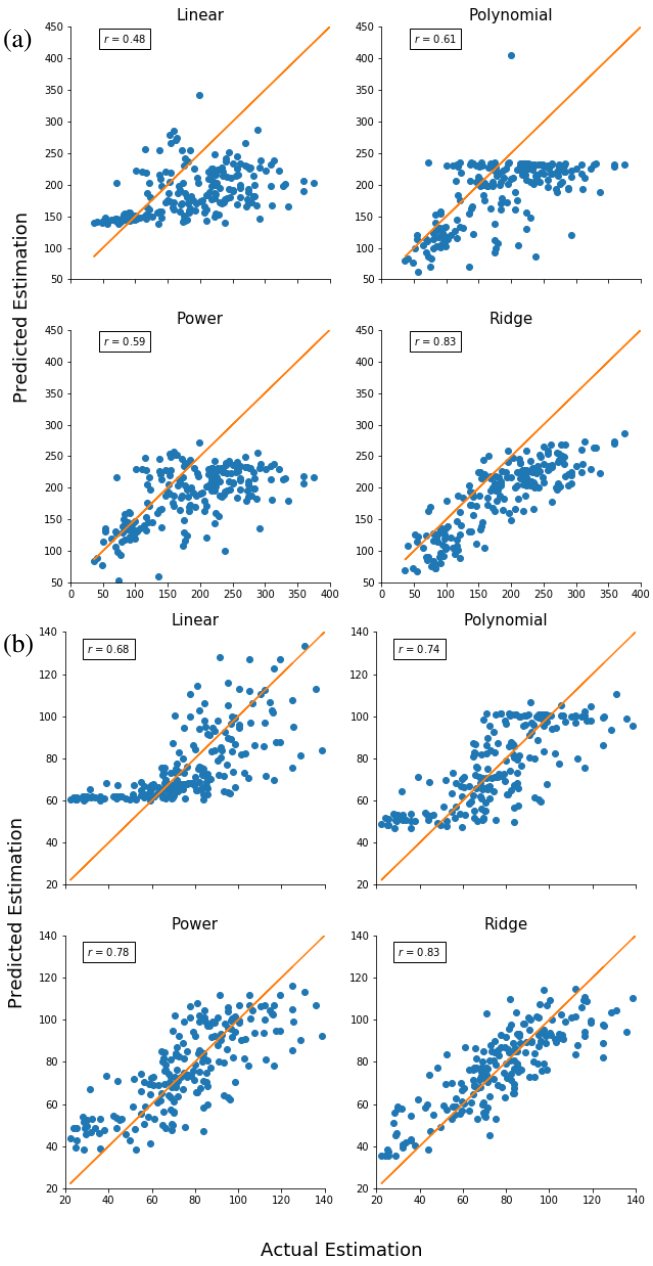


Figure 2: Scatterplots of predicted estimates using leave-one-out cross-validation (LOOCV) vs. actual participant estimates for (a) Study 1 and (b) Study 2, along with Pearson correlations.

## Results

Study 1 and 2 elicited a total of 40,400 participant estimates for 400 distinct naturalistic entities in two domains calories

[4]For each domain, we trained our models on 199 judgment targets and then used the trained model to predict participant estimates of the left-out target. This procedure was repeated for each judgment target to get LOOCV predictions

Table 1: Aggregate level predictive accuracy of the three mapping error models – linear (Eq.1), polynomial (Eq.2), power (Eq.3), and knowledge error model (ridge) using leave-one-out cross-validation (LOOCV) for Study 1 and Study 2.

| | Study 1 | | | Study 2 | | |
|---|---|---|---|---|---|---|
| | Correlation | $R^2$ | RMSE | Correlation | $R^2$ | RMSE |
| Linear | 0.48 | 0.23 | 4609.04 | 0.68 | 0.46 | 348.72 |
| Polynomial | 0.61 | 0.37 | 3772.07 | 0.74 | 0.55 | 291.82 |
| Power | 0.59 | 0.35 | 3902.78 | 0.78 | 0.60 | 258.64 |
| Ridge | 0.83 | 0.68 | 1924.11 | 0.83 | 0.68 | 208.10 |

of 200 common food items and infant mortality rates of 200 countries. Consistent with prior work, we found that participants made substantial errors. The average absolute differences between the average participant estimate, $y_i$, and the correct answer, $z_i$, for food calories and infant mortality rates were -45.28kcal per 100g ($se = 10.8$) and 53.44 deaths per 1,000 live births ($se = 3.78$) respectively, indicating an overall underestimation of food calories and overestimation of infant mortality rates. Figure 1 reflects some overestimation of low calories, significant underestimation of high calories, and overall overestimation of infant mortality rates.

Table 1 summarizes the aggregate level performance of the three mapping error models and one knowledge error model. We evaluated model performance using the Pearson correlation between observed $y_i$ and predicted $y_i$, $R^2$, and root mean square error (RMSE), in the out-of-sample tests. Figure 2 shows scatterplots of predicted estimates using LOOCV and average participant estimates, along with Pearson correlations. We found that the knowledge error model was able to predict average participant estimates fairly accurately, with out-of-sample correlation rates of .83 for both domains on the aggregate level. In contrast, the best mapping error models were only able to achieve aggregate-level out-of-sample correlation rates of .61 for foods and .78 for countries (all $p < 10^{-22}$). We obtained similar results on the individual level. The best mapping error model achieved average individual-level out-of-sample correlation rates of .37 for foods and .43 for countries, while the knowledge error model achieved .51 for food and 0.47 for countries. Our results showed statistically significant improvements in predictive accuracy when using the knowledge error model compared to the mapping error models on both the aggregate and the individual level.

## Traces of Judgment Errors

### Computational Methods

In the previous section, we showed that the word-embedding-based vector representations could be used to predict estimates of food calories and infant mortality rates by multiplying $x_i$ (the vector representations for the foods and countries) with different weight vectors $w_1$ (Study 1) and $w_2$ (Study 2). As mentioned in computational methods of last section, these weight vectors can be seen as capturing cue weights on semantic knowledge. In this section, we hope to better understand the psychological substrates of the judgment errors that these weights generate. What are the features that lead to the overestimation or underestimation of food calories and infant mortality rates?

To address this, we took the 5,000 most frequent words from the corpus of contemporary American English (`http://corpus.byu.edu/coca/`) that were not judgment targets and for each word $j$, we also obtained a 300-dimensional vector, $s_j$, from the Word2Vec model. Intuitively, the weight vector $w$ in the previous section could be seen as a function that projects the semantic knowledge represented by $x_i$ onto a numerical estimation line $y_i$. By multiplying $s_j$ by the weight vector $w$, we got a vector representation $e_j$ for these 5,000 words in the numerical estimation line. Similarly, we also trained a weight vector $w'$ to predict the correct answer, $z_i$, using $w' * x_i$. Multiplying $s_j$ by this new weight vector $w'$ would give us a vector $t_j$ that pinpoints the location of the 5000 words in a line of correct answers. The difference between $e_j$ and $t_j$ then informs us of what words and concepts might be overweighted (or underweighted) in the estimation process. In other words, this difference would offer a quantitative measure of how much any given word contributes to overestimation (or underestimation).

### Results

Figure 3 has word clouds of 50 words[5] that greatly contribute to over- and under- estimation for both domains. These words reveal potential conceptual underpinnings of judgment biases that match our intuition. For example, words related to dining out (e.g. restaurant, menu, chef, wine) bias toward overestimation of calories; words appearing to be small in portion (e.g. flour, candy, powder, dust) bias toward underestimation of calories; developing-country-related words (e.g. Iraqi, Cuban, Palestinian, Arab) contribute to overestimation of infant mortality rates; and European-country-related words (e.g. Dutch, German, French, European) contribute to underestimation of infant mortality rates.

[5]We included 50 words because that was the maximum number of legible words that could be fit into the graphs.

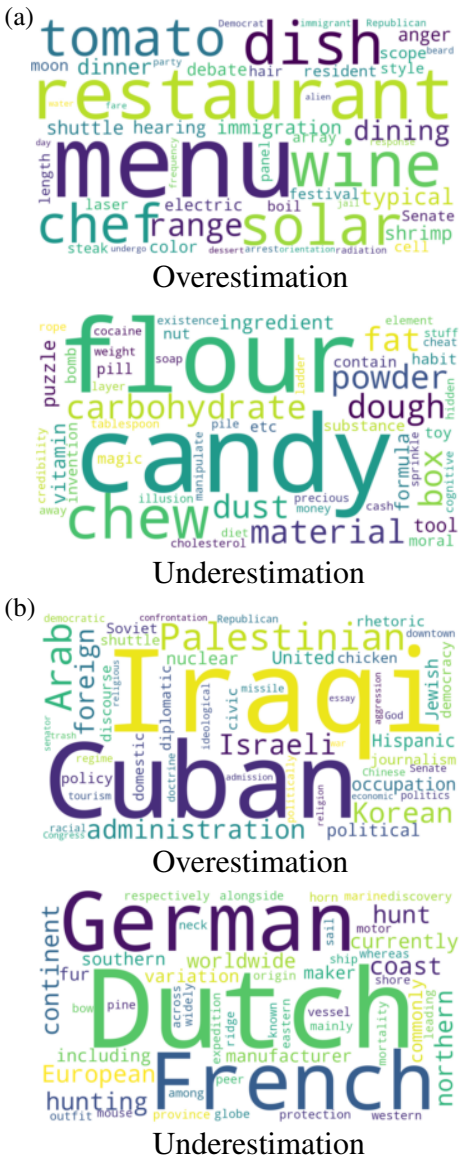


Figure 3: Word clouds of words that greatly contribute to overestimation and underestimation of (a) food calories and (b) infant mortality rates. Font reflects the magnitude of over- and under-estimation. Color has no meanings.

## Discussion

We built computational models to compare two types of errors in numerical estimation: *mapping errors* and *knowledge errors*. We applied these models to study naturalistic numerical estimates in two studies involving judgments of calories of food items and judgments of infant mortality rates of countries. Consistent with previous findings, the best fitting mapping error models in both studies were not linear and drastically outperformed simple linear regression (baseline). We found the common inverse-S-shape pattern in food calorie estimation but not in the infant mortality domain. Although almost all countries' infant mortality rates were overestimated, the magnitude of overestimation appeared to be larger for countries with low infant mortality than for countries with high infant mortality.

Although our mapping error models were able to provide a good account of our data, we obtained even higher predictive accuracy rates from our knowledge error models. This indicates that we can predict participant estimates better if we assume some flexible (potentially incorrect) use of memory cues instead of some flexible (potentially incorrect) use of correct beliefs, and that judgment errors appear to stem primarily from the incorrect use of knowledge, rather than the incorrect mapping of the true quantities. Here it is also useful to note that our knowledge error models, unlike our mapping error models, did not know the correct responses. Rather they were able to predict participant estimates merely by proxying the rich semantic knowledge that participants used in their own judgments. This showcases both the descriptive power of the models as well as their domain-general applications: We can use these models to predict participant estimates even when we (the researchers) do not know the correct answers.

So far, we've only studied *mapping errors* and *knowledge errors* separately. It is likely that these two types of errors take place simultaneously as suggested by prior work (e.g. Brown & Siegler, 1993; von Helversen & Rieskamp, 2008). Future work should investigate the interaction between the two types of errors and also the interplay between error type and judgment domain – how do people balance between these two types of errors to minimize overall errors and in which domains are people more likely to make one type of errors than the other?

For our knowledge error model, we used semantic knowledge which was captured by word embeddings to predict participant estimates. Building upon recent successful applications of such models in naturalistic judgments (e.g. Bhatia, 2019; Richie et al., 2018, Dec), we showed how a single knowledge representation derived from natural language data was able to predict participant numerical estimation with high out-of-sample accuracy. We acknowledge that due to its high dimensionality, word-embedding-based representations are likely to contain more knowledge about judgment targets than what people actually use to estimate numerical quantities, and due to its generality, they also don't account for individual differences, such as personal experience and level of expertise. One way to address this may involve training different word embeddings for different populations, which is a potential topic for future work.

Although our mapping error models and knowledge error models lack some interpretability in terms of cognitive process underlying numerical estimation, the different non-linear mapping error models for food calories and infant mortality rates suggest different transformations from internal beliefs to external responses. Likewise, the words generated from the knowledge error models reveal intuitive conceptual bases of judgment errors, e.g. a misuse of food size associations in food judgment and a misuse of poverty associations in infant mortality judgment. These results show that our knowledge

error model has explanatory value, and can shed light on the types of associations that contribute to judgment errors across different domains.

Finally, we would like to emphasize the naturalism of the two domains examined in this paper. Our approach is unique in that it can be applied to numerical estimates for arbitrary natural entities, such as food items and countries. This opens up new avenues for applying cognitive science theory to policy-relevant applications, such as those pertaining to health-related and humanitarian issues. We look forward to future work that extends our approach to model the types of errors at play in the many important judgments that people make on a day-to-day basis.

## References


Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological review*, *124*(1), 1–20.

Bhatia, S. (2019). Predicting risk perception: new insights from data science. *Management Science*.

Brown, N. R. (2002). Real-world estimation: Estimation modes and seeding effects. *Psychology of learning and motivation*, *41*, 321–359.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological review*, *100*(3), 511–534.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Chernev, A., & Chandon, P. (2011). Calorie estimation biases in consumer choice. In R. Batra, P. Keller, & V. Strecher (Eds.), *Leveraging consumer psychology for effective health communications: The obesity challenge* (pp. 104–121). New York, NY: M.E. Sharpe.

Curtis, D. W., Attmeave, F., & Harrington, T. L. (1968). A test of a two-stage model of magnitude judgment. *Perception & Psychophysics*, *3*(1), 25–31.

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, *115*(1), 199–213.

Erlick, D. E. (1964). Absolute judgments of discrete quantities randomly distributed over time. *Journal of Experimental Psychology*, *67*(5), 475–482.

Fennell, J., & Baddeley, R. (2012). Uncertainty plus prior equals rational bias: An intuitive bayesian probability weighting function. *Psychological Review*, *119*(4), 878–887.

Gallagher, C. A. (2003). Miscounting race: Explaining whites' misperceptions of racial group size. *Sociological Perspectives*, *46*(3), 381–396.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644.

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

Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *The Quarterly Journal of Experimental Psychology*, *69*(4), 626–653.

Herda, D. (2013). Too many immigrants? examining alternative forms of immigrant population innumeracy. *Sociological Perspectives*, *56*(2), 213–240.

Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 209–234). New York: Oxford University Press.

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 621–642.

Hollands, J., & Dyre, B. P. (2000). Bias in proportion judgments: the cyclical power model. *Psychological review*, *107*(3), 500–524.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156.

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

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory*, *4*(6), 551–578.

Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory*, *20*(5), 511–517.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Vanderplas, J. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(Oct), 2825–2830.

Richie, R., Zou, W., & Bhatia, S. (2018, Dec). Semantic representations extracted from large language corpora predict high-level human judgment in seven diverse behavioral domains. Retrieved from `https://psyarxiv.com/g9j83` doi: 10.31234/osf.io/g9j83

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of direct psychophysical measurement. *Journal of Mathematical Psychology*, *24*(1), 21–57.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. *Experimental cognitive psychology*

*and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*, 237–249.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 613–625.

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, *137*(1), 73–96.

Zhang, H., & Maloney, L. T. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, *6*, 1–14.