Does Precision Affect Categorization? Magnitude Categorization and Measurement Scales

Ling Sun (ls44@iu.edu)
Department of Linguistics,
Indiana University

Soyoung Kim (sk135@iu.edu)
Department of Linguistics,
Indiana University

Robert L. Goldstone* (rgoldsto@iu.edu)
Psychological and Brain Sciences,
Indiana University

Abstract

How do systems of measurement influence our conceptualization of relative magnitudes? This study investigates the cognitive interplay between measurement precision and magnitude categorization. By employing morphed shapes organized by an arbitrary dimension, we examine whether exposure to high- vs low-precision numerical systems affects conceptual judgments and well-known phenomena such as semantic distance and semantic congruity effects as found for familiar dimensions. Participants trained on novel scales revealed differences in their sensitivity that depended on the precision of the trained measurement system, consistent with high-precision systems leading to relatively expanded dimensional encodings compared to low-precision systems. Our findings also shed light on other topics such as the interplay of perception and language in learning novel dimensions and the association of directionality with a mental number line.

Keywords: magnitude; relative adjectives; measurement; numerical cognition; perception

Introduction

How do different systems of measurement, such as inches versus centimeters, pounds versus kilograms, or Fahrenheit versus Celsius, influence how we perceive and conceptualize relative magnitudes like height, weight, and temperature? These systems differ in precision, linear relationships, and reference points and are typically acquired after the concepts are introduced. For example, children learn words/concepts like *tall* and *short* by age three, yet their ability to associate these adjectives with numerical scales such as “5 meters” or “10 feet” develops much later (Tiemann, Hohaus, & Beck, 2012). This delayed mapping raises a key question: how do numerical systems affect conceptual understanding of magnitudes like *tall* versus *short* or *hot* versus *cold*?

Although much research has explored magnitude perception (Čech & Shoben, 1985; Leibovich & Henik, 2013) and numerical cognition, such as counting and symbolic number processing (Gallistel & Gelman, 1992; Ansari et al., 2003; Le Corre & Carey, 2007), little attention has been given to the relationship between them. Concepts like *tall* are loosely defined, with judgments becoming more variable as objects approach the contextual standards (Tribushinina, 2011).

For familiar dimensions, studies have already shown that people discriminate concepts or objects more easily when they differ greatly, a phenomenon known as the semantic distance effect (Banks, Clark, & Lucy, 1975; Čech & Shoben, 1985; Cantlon & Brannon, 2005). Similarly, the semantic congruity effect reveals that decisions are quicker when the cue for comparison aligns with the magnitude of the stimuli (Ryalls & Smith, 2000; Cantlon & Brannon, 2005). For example, when given small collections of dots, people respond faster to ”choose smaller” than to ”choose bigger.”

While these effects are well-established for familiar dimensions, it remains unclear whether they extend to arbitrary dimensions. If linguistic labels can be applied to arbitrary dimensions and still exhibit semantic distance and semantic congruity effects, this suggests that the mapping between perception and language is more flexible than previously assumed. Building on this, our findings show that arbitrary dimensions not only support conventional semantic effects after training, but are also sensitive to learning precision: when symbolic measures were less precise, these effects were notably reduced. This study thus addresses the generality and fluidity of cognitive mechanisms underlying magnitude representation.

Contribution

This paper contributes to the broader literature on categorization and concept/manifold learning (Goldstone & Steyvers, 2001; Rogers, Kalish, Harrison, Zhu, & Gibson, 2010). Building in particular on Rogers et al. (2010)’s claim that arbitrary dimensions and manifolds are learnable but challenging, our study investigates whether arbitrary dimensions – constructed by morphing between shapes – can be learned in a similar way to familiar dimensions like height and weight. Specifically, it explores the following questions:

- Can people acquire an arbitrary dimension with the guidance of symbolic labels, such as novel adjectives and numeric measures?
- Do people internalize the newly introduced dimension similar to how they internalize familiar dimensions? In other words, can the semantic distance and/or congruity effects

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

still be observed after the acquisition of an unfamiliar dimension?

- Does measure precision impact categorization along the dimension?

Additionally, this work contributes to linguistic theories on adjective pairs. For adjective pairs like “tall” and “short,” only the ‘positive’ member “tall” naturally accept direct measurements (e.g., “5 feet tall” versus the unnatural “5 feet short”). This validates the ‘positive’ member “tall” as unmarked and ‘negative’ member “short” as marked in direct measurement contexts, and such an asymmetry is called a markedness effect (Beck, Oda, & Sugisaki, 2009). To foreshadow our study, our results shows that directional asymmetry (e.g., “tall” vs. “short”) emerges even from learning a novel label pair.

More broadly, our findings support and extend the proposal by Goldstone and Steyvers (2001) that training can lead to the formation of new psychological dimensions beyond the reweighting of existing ones. We demonstrate that arbitrary dimensions can come to support conventional semantic phenomena, including semantic distance and semantic congruity effects, and may be conceptualized in a number-line-like manner after training. Furthermore, our results reveal that precision of learning matters: when symbolic measures were presented with low precision, such semantic effects were attenuated.

Methodology

Subjects

A total of 108 students from the course PSY-P 101: Introductory Psychology at Indiana University Bloomington were recruited via Psychological and Brain Sciences (PBS) Experiment Management System (SONA). All were native speakers of English with no history of vision, speech, or hearing disorder. Out of the total participants, only those who had an accuracy of 65% or higher in the differentiation testing block were considered for analysis.

Stimuli

20 stimuli from a morph sequence representing a novel scale were generated for participants to learn. The two ends of the scale are labeled with a pair of nonce words (for instance, one end is labeled as most “frimble”, the other as most “orpital”). “Orpital” and “frimble” are always put in opposition – the more “orpital” something is, the less “frimble” it is. The assignment of specific labels was randomized across participants.

The stimuli were created as morph sequences within a two-dimensional morph space using Bezier curves. One dimension determined the target concept pair (e.g., “orpital” versus “frimble”), with distortions globally controlled by a percentage parameter ranging from 0 to 1 in even increments of 0.05. The second dimension introduced noise to simulate natural variability in categorization tasks, encouraging participants to focus on the primary target continuum. Noise values were randomly sampled within defined ranges during both training and testing. Figure 1 illustrates the stimulus continuum without the application of noise.

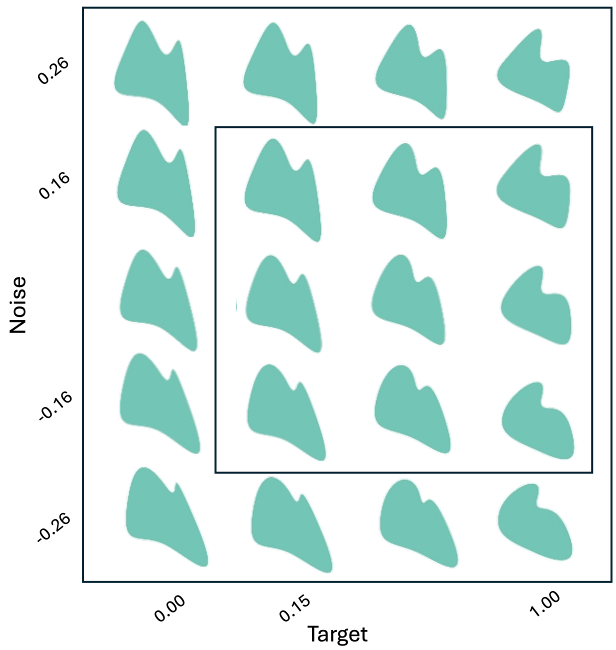

Figure 1: The space of morphed stimuli. Participants during training are only exposed to the figures in the inner square.

The experiment was broken down into blocks, and each block contained pairs of stimuli generated with parametric distances between pairs carefully controlled. The controlled distances between the pairs vary by block type (Learning and Testing), as described later. In all learning blocks, the controlled distances between the pairs ranged between 0.25 and 0.45, whereas they ranged between 0.10 and 0.90 in all testing blocks. A wider range of distances was applied to the testing stimuli in order to assess participants’ ability to extrapolate their training to new values along the trained dimension. Finally, both the stimulus generation and presentation programs were written in JavaScript with the jsPsych library.

Experiment design and conditions

The experiment was designed to include two comparison groups, termed the High Precision group and the Low-Precision group. Participants in both groups were introduced to a novel pair of labels in a label familiarization block (Learning Block I). They were then trained on degree measurements using a numerical scale in a degree-learning block (Learning Block II), where subjects were divided into a High-Precision group and a Low-Precision group via random assignment. The 0.25-0.45 range of distances between stimuli was divided into 4 vs. 2 categories in the High- versus Low-Precision conditions, respectively.

These conditions were designed to test the effect of precision on the learning and generalization of the morph continuum, as well as the use of linguistic labels.

Experiment procedure

The experiment consisted of two learning blocks and three testing blocks.

Learning blocks Learning blocks consisted of 1) the Label Familiarization Block (15 passive learning trials and 35 active learning trials) and 2) the Degree-Learning Block (15 passive learning trials and 45 active learning trials).

- **the Label Familiarization Block (Learning Block I)** acquainted participants with the novel labels "orpital" and "frimble." On passive trials, pairs of stimuli were shown with a description (e.g., "Compared to the left one, the right one is more orpital"). On active trials, participants were asked to decide if the right stimulus was more "orpital" or "frimble," receiving feedback ("correct" vs "incorrect") and only proceeding to the next trial upon a correct answer.

- **the Degree-Learning Block (Learning Block II)** trained participants on degree measurements using a numerical scale from 1 to 4. On passive trials, participants were informed of the distance between stimuli pairs (e.g., "Compared to the left one, the right one is 3 degrees more orpital"), varying by condition (high-precision vs. low-precision). On active trials, participants were asked to specify the distance between stimuli pairs, proceeding only upon a correct answer.

Testing blocks There were three testing blocks: Pretest (before Learning Block I), Post label-familiarization test (between Learning Block I and II), and Post-degree test (after Learning Block II). There were three types of trials in these testing blocks:

- **XAB**: Participants were presented with three items and asked to identify which of the latter two items (A or B) was identical or more similar to the first item (X).

- **Differentiation**: Participants were asked to decide if one item from a pair was more A-like or more B-like, deciding which item exhibited a higher degree of a specific characteristic (A-er or B-er).

- **Comparison**: Participants were asked to judge whether an item compared to the one next to it was more A-like, more B-like, equally A-like or equally B-like, thus evaluating the degree to which items exhibit characteristics A or B, as in figure 2.

The pretest included 16 XAB trials, while the other two testing blocks included all three types of tasks (16 XAB, 40 differentiation, and 30 comparison trials). Identical stimuli were presented across all tests but in a randomized order.

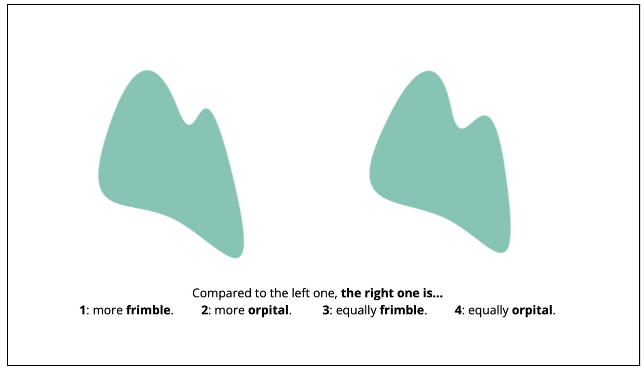

Figure 2: A sample trial from the Comparison task

Results

Out of the 108 total participants, 67 were included in the analysis based on a criterion of test trial accuracy exceeding 65%. 35 participants were in the high precision condition, and 32 participants in the low precision condition.

Accuracy and Extrapolation

Participants demonstrated effective learning, maintaining solid accuracy even when tested on entirely new dimension values beyond their training experience. During training, they only encountered stimuli with distortions from 0.1 to 1.0, noise values from -0.16 to 0.16, and stimulus pairs with distances along the target dimension ranging from 0.25 to 0.45, with the noise dimension differing by at most 0.16. However, during testing, they successfully generalized to a much broader range: stimuli spanned the entire continuum from 0 to 1, noise extended from -0.26 to 0.26, target-dimension distances ranged from 0.1 to 0.9, and noise-dimension distances increased to 0.26.

Remarkably, participants demonstrated robust generalization across all dimensions, with performance remaining high, especially for larger distances (> 0.45). Even for smaller distances (< 0.25), accuracy stayed above chance, indicating effective learning beyond trained conditions. As shown in Fig. 3 (top row), participants successfully extended their knowledge to a broader stimulus and noise range. This ability to extrapolate beyond the training conditions highlights the robustness of their learned dimension.

Precision and Semantic Distance Effect

A semantic distance effect was observed across Differentiation, and Comparison trials (see fig 3). On Differentiation trials, accuracy significantly increased with greater distance ($p = 0.000, p < 0.05$, Pseudo-$R^2 = 0.1438$; $\texttt{correct} \sim \texttt{distance}$), while reaction time decreased, as indicated by an Ordinary Least Squares (OLS) regression ($p = 0.000, p < 0.05$, $R^2 = 0.003$; $\texttt{RT} \sim \texttt{distance}$). Similarly, in Comparison trials, participants were significantly less likely to judge stimulus pairs as equal with increasing distance ($p = 0.000, p < 0.05$, Pseudo-$R^2 = 0.098$; $\texttt{sensitivity} \sim \texttt{distance}$).

Measure	Predictor	μ (Posterior)	95% CI	$P(\beta > 0)$	p-value
Accuracy	`distance`	4.329	$[3.867, 4.843]$	> 0.99	0.000
	`distance` $\times$ `precision group`	-1.410	$[-2.338, -0.509]$	< 0.01	0.003
Sensitivity	`distance`	12.814	$[11.462, 14.205]$	< 0.00	0.000
	`distance` $\times$ `precision group`	-1.723	$[-3.769, 0.032]$	> 0.97	0.000

Table 1: Comparison of Bayesian and Frequentist Results for Accuracy and Sensitivity

Interestingly, a by-group difference in the semantic distance effect emerges for both differentiation and Comparison trials (see fig 3). Specifically, the group exposed to low-precision degree comparisons shows a diminished semantic distance effect compared to the high-precision group. For differentiation trials, the low-precision group demonstrates lower accuracy as the distance increases, compared to the high-precision group ($p = 0.025, p < 0.05$, Pseudo-$R^2 = 0.1462$; frequentist logit model: `correct` $\sim$ `distance` $\times$ `precision group` $\times$ `task`), with the effect observed more clearly on post-training trials($p = 0.003, p < 0.01$), Pseudo-$R^2 = 0.1383$; logit model: `correct` $\sim$ `distance` $\times$ `precision group`).

To further validate these findings, a Bayesian logistic regression was employed to compute posterior estimates and credible intervals for the interaction between **distance** and **precision group**. For differentiation trials, the Bayesian model indicates that the posterior mean of the interaction term (`distance` $\times$ `precision group`) demonstrates a substantial effect (95% credible interval: [-2.338, -0.509]), with the probability of the coefficient being less than zero ($P(\beta > 0)$) exceeding 95% (see table 1).

On the Comparison trials, the results indicate that the low-precision group is significantly more likely than the high-precision group to respond 'equally frimble/orpital' for stimuli pairs with larger distances. ($p < 0.001$, Pseudo-$R^2 = 0.1013$; logit model: `sensitivity` $\sim$ `distance` $\times$ `precision group`). Bayesian analysis further supports this finding, with the posterior estimate for the interaction term ($\beta_{PD} = 1.723$) showing a credible interval $[-0.032, 3.769]$. Although the interval includes zero, it trends positive, providing moderate evidence for an effect of precision group on distance sensitivity (see table 1).

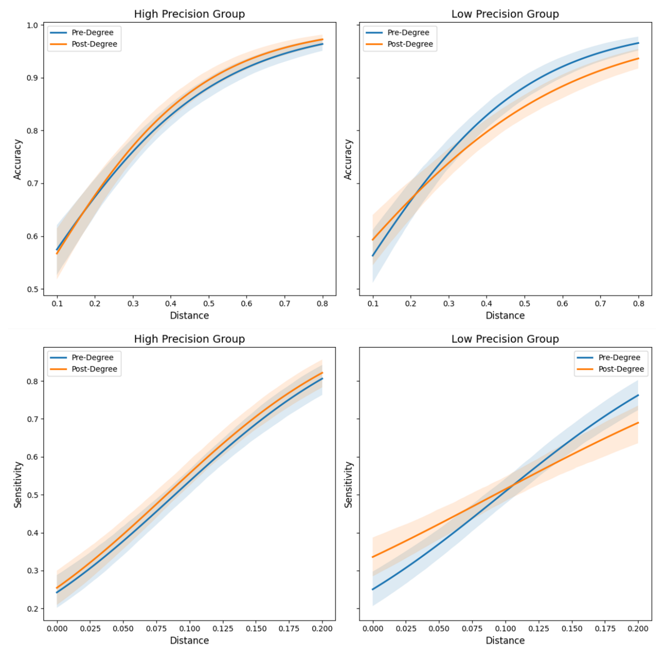

Figure 3: Comparison of semantic distance effects between high precision versus low precision groups. Top row: semantic distance effect on Differentiation trials. Bottom row: semantic distance effect on Comparison trials.

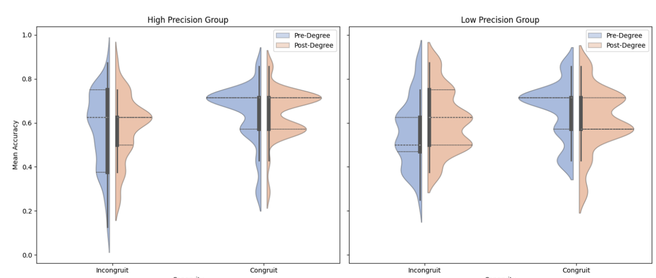

Figure 4: The congruity effect in differentiation trials

Precision and Semantic Congruity

Semantic congruity effect was also observed across Differentiation trials. According to a semantic congruity effect, if the tested shapes are closer to the "orpital" than "frimble" end of the continuum, then responses to "orpital" (congruent) questions should be more accurate than "frimble" questions (incongruent), and *vice versa* for "frimble" questions.

For response accuracy, a significant semantic congruity effect was reported with paired-samples t-test (t-statistic = -2.9389, p-value = 0.00 < 0.05) with mean accuracy for congruent trials at 0.6429 (SD = 0.1033) and for incongruent trials at 0.5777 (SD = 0.1171). However, no significant effect was found for reaction time.

This congruity effect slightly reduced after degree training for both groups — both improved their performance on incongruent cases while showing worse performance on congruent cases. This effect was more pronounced in the low-precision group, though no significant between-group differences were observed (see fig 4). However, this result should be interpreted with caution, as the pseudo-R^2 for the model was negative, indicating weak explanatory power, and the interaction term in the logistic regression model (`accuracy` $\sim$

`congruity × training`) was not significant.

Inherent directionality

Finally, after the Label Familiarization block, greater accuracy was consistently observed in both differentiation trials, post-label and post-measure trainings. for the stimuli which were displayed on the right side of the screen (position vector $> .5$) than the ones on the left (position vector < 0.5) (See fig 5).

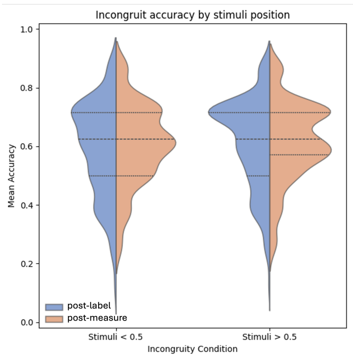

Figure 5: The directionality effect in incongruity trials

Using the logit regression model (`correct ∼ incongruity condition`) with the 2-level incongruity condition (one if stimuli position $< .5$, one if $> .5$), this directionality effect seems to be significant (pseudo-R^2 = 0.0161, p-value = 0.000, coefficient = 0.6135).

Discussion

This study investigated whether arbitrary or otherwise unfamiliar dimensions are indeed learnable, and how such an acquisition influences people's conceptualization and/or perception of dimensions. By introducing high- and low-precision numerical scales in a novel morph space task, we explored the interplay between measurement precision, magnitude categorization, and linguistic concepts. The findings highlight the general interplay and alignment between perception and language, while offering other notable insights on how measurement precision impacts categorization and the relationship between dimension magnitude and the mental number line.

Alignment between Perception and Language

Humans readily organize familiar dimensions, such as height and width, in terms of spatialized magnitude but often struggle to learn arbitrary manifold structures. However, research suggests that they can acquire these structures when conditions strongly favor their formation (Rogers et al., 2010). In this study, participants, when strongly prompted to engage with arbitrary dimensions and novel categories, demonstrated significant generalization across novel contexts. This suggests that humans are highly adaptable and capable of learning complex dimensions beyond their prior experience. Our findings further demonstrate that learning abstract shapes and unfamiliar dimensions is achievable with the support of linguistic labels and measurement systems, highlighting a flexible relationship between perception and language.

Acquired Equivalence

Our results align with the acquired equivalence literature (Folstein, Palmeri, & Gauthier, 2014), demonstrating how category learning influences representational sensitivity. Using pre-degree learning as a baseline, we found that the low-precision group exhibited a reduced semantic distance effect after training, while the high-precision group showed a slightly greater (though non-significant) distance effect after training. Similarly, for semantic congruity, the low-precision group exhibited a smaller congruity effect compared to the high-precision group. Both groups demonstrated signs of acquired equivalence within categories, likely due to receiving fewer numerical categories (i.e., degree values) during learning, which may have led them to treat more stimuli as equivalent or similar.

This pattern mirrors an expansion-based mechanism: high-precision training effectively stretches the relevant dimension, enhancing sensitivity to differences. As in plasticity effects observed in the somatosensory cortex (Kaas, 1991), where neural representations expand with increased sensory discrimination demands, high-precision training appears to expand the conceptual space for magnitude comparisons, making differences more distinguishable. Conversely, the low-precision group, showing less expansion and even compression, showed reduced sensitivity in categorization, leading to a weaker distance effect and reduced semantic congruity effect. These findings suggest that the scale of learned distinctions influences perceptual resolution, shaping the way novel dimensions are processed and categorized.

Mental Number Line

Finally, our results revealed that accuracy was higher for stimuli corresponding to the right-hand key response (i.e., 'p' for stimuli with greater vector positions and 'q' for smaller positions), suggesting a directional sensitivity aligned with the mental number line (Fias & Bonato, 2018). This bias is consistent with SNARC-like effects (Wood, Willmes, Nuerk, Fischer, et al., 2008), where numerical or magnitude-related judgments tend to be spatially mapped, reinforcing the idea that novel dimensions, just like familiar ones, are mentally structured along a left-to-right continuum.

Moreover, participants consistently learned "orpital" and "frimble" as direct opposites, such that the more "orpital" something was, the less "frimble" it was. This opposition suggests that participants treated these novel adjectives as polar ends of a single underlying magnitude dimension, much like "tall" and "short" in linguistic markedness. The fact that "orpital" (or "frimble," depending on condition) was consistently associated with the right side of the keyboard further supports the idea that learned dimensions adopt a spatial structure, mirroring established cognitive frameworks for representing magnitude.

This finding provides further evidence that humans automatically map newly acquired dimensions onto a mental number line, reinforcing the idea that perception, language, and magnitude cognition are deeply interconnected.

Conclusion and Future Directions

Our study demonstrates that an entirely arbitrary dimension can be learned and categorized using novel adjective labels. Crucially, this categorization exhibits key properties of familiar dimensions, such as the semantic distance and semantic congruity effects, indicating that newly acquired dimensions integrate into existing cognitive structures. Furthermore, we show that measurement precision plays a critical role in shaping the acquired dimension. Participants trained with fewer numerical categories (low-precision degrees) exhibited reduced stimulus discrimination, as reflected by their weaker semantic distance and congruity effects.

Additionally, our findings suggest that participants may be mapping the arbitrary dimension onto a mental number line, pointing to a more general and flexible mechanism underlying magnitude and adjective acquisition. This aligns with prior research on the spatial representation of magnitude and extends it to arbirary dimensions, highlighting a deep-rooted cognitive bias toward structuring information along a continuum.

Future studies should explore how these effects manifest in different learning conditions. One promising avenue is to include a control condition without linguistic labels to examine whether explicit labels are necessary for structuring the arbitrary dimension. Additionally, alternative tasks, such as forced-choice comparisons ("Which of these two objects is more?"), relative estimation ("How many more is X than Y?"), or absolute estimation ("How many is X?"), could provide deeper insights into how individuals encode and map novel dimensions.

References

Ansari, D., Donlan, C., Thomas, M. S., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology*, *85*(1), 50–62.

Banks, W. P., Clark, H. H., & Lucy, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(1), 35.

Beck, S., Oda, T., & Sugisaki, K. (2009). Parametric variation in the semantics of comparison: Japanese vs. english. *Journal of East Asian Linguistics*, *18*(3), 225–271. Retrieved from `https://doi.org/10.1007/s10831-009-9040-y` doi: 10.1007/s10831-009-9040-y

Cantlon, J. F., & Brannon, E. M. (2005). Semantic congruity affects numerical judgments similarly in monkeys and humans. *Proceedings of the National Academy of Sciences*, *102*(45), 16507–16511.

Čech, C. G., & Shoben, E. J. (1985). Context effects in symbolic magnitude comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 299.

Fias, W., & Bonato, M. (2018). Which space for numbers? In *Heterogeneity of function in numerical cognition* (pp. 233–242). Elsevier.

Folstein, J., Palmeri, T., & Gauthier, I. (2014, 12). Perceptual advantage for category-relevant perceptual dimensions: the case of shape and motion. *Frontiers in Psychology*, *5*, 1394. doi: 10.3389/fpsyg.2014.01394

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1-2), 43–74.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of experimental psychology: General*, *130*(1), 116.

Kaas, J. H. (1991). Plasticity of sensory and motor maps in adult mammals. *Annual review of neuroscience*.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438.

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, 375.

Rogers, T., Kalish, C., Harrison, J., Zhu, J., & Gibson, B. (2010). Humans learn using manifolds, reluctantly. *Advances in neural information processing systems*, *23*.

Ryalls, B. O., & Smith, L. B. (2000). Adults' acquisition of novel dimension words: Creating a semantic congruity effect. *The Journal of General Psychology*, *127*(3), 279–326.

Tiemann, S., Hohaus, V., & Beck, S. (2012). Crosslinguistic variation in comparison: evidence from child language acquisition. *Empirical approaches to linguistic theory: Studies of meaning and structure*, 115–146.

Tribushinina, E. (2011). Once again on norms and comparison classes. *Linguistics*, *49*(3), 525–553. Retrieved 2025-02-02, from `https://doi.org/10.1515/ling.2011.016` doi: doi:10.1515/ling.2011.016

Wood, G., Willmes, K., Nuerk, H.-C., Fischer, M. H., et al. (2008). On the cognitive link between space and number: A meta-analysis of the snarc effect. *Psychology science quarterly*, *50*(4), 489.