From "Few" to "Many": Explicit Spatial Associations for Language vs. Automaticity for Numbers? Evidence from Card-Sorting, Lexical Decisions, and Numerical Classification

Greta Gaimarri (greta.gaimarri@unitn.it)[1] & Jakub Szymanik[1,2]
[1]Center for Mind/Brain Sciences, University of Trento, Italy
[2]Department of Information Engineering and Computer Science, University of Trento, Italy

Abstract

Do linguistic representations of magnitude (e.g., few, many, cold, hot) share the same spatial architecture as numerical quantities? While the Spatial–Numerical Association of Response Codes (SNARC) effect shows that numbers are automatically mapped onto a mental number line, it remains unclear whether linguistic magnitudes elicit similar automatic spatial associations. We conducted three experiments to address this issue. Experiment 1 showed that Italian quantifiers and adjectives are explicitly mapped onto a horizontally compressed scale. Experiment 2 found no evidence for an implicit Spatial–Linguistic Association of Response Codes (SLARC) effect, with Bayesian analysis providing strong support for the null hypothesis (BF_{01}=87). Experiment 3 confirmed robust SNARC and ratio effects for Arabic digits in the same participant pool. Together, these findings indicate that although linguistic and numerical magnitudes share a similar psychophysical structure, spatial mapping in language may be task-dependent and lack the automaticity observed in numerical cognition.

Keywords: spatial-linguistic associations; SNARC; SLARC; magnitude processing automaticity;

Introduction

In everyday communication, speakers routinely convey perceptual estimates of continuous magnitudes to their interlocutors. Expressions such as *Paul is tall* or *There are many swans in the lake* encode approximate information about size or quantity through, respectively, gradable adjectives and vague quantifiers, two highly frequent linguistic devices across languages (e.g., Szymanik & Kieraś, 2022). Both adjectives and quantifiers map properties of the world onto continuous scales (e.g., height), allowing speakers to express imprecise yet informative magnitude representations without relying on exact measurements.

Building on this idea, the present study asks to what extent this mapping recruits the same magnitude representations that support perceptual estimation. If these linguistic expressions rely on magnitude-based representations, their processing should engage cognitive mechanisms analogous to those involved in non-linguistic magnitude judgments. To test this hypothesis, we examine whether judgments involving gradable adjectives and vague quantifiers give rise to the behavioral effects classically observed in magnitude processing.

Specifically, research has identified a set of robust behavioral effects that reflect the structure of underlying magnitude representations. According to the literature, quantities are mentally represented along a logarithmically compressed scale, such that differences between quantities become less distinguishable as absolute magnitude increases. This representational property gives rise to the distance effect, whereby discrimination between two quantities becomes easier as their distance increases (Moyer & Landauer, 1967; Pinel et al., 2004), and the ratio effect, which shows that discriminability depends on the ratio between quantities rather than on their absolute difference (Buckley & Gillman, 1974; Moyer & Landauer, 1967; Pica et al., 2004). Specifically, according to the distance effect, judgments become more difficult as the distance between two quantities decreases (e.g., 2 vs. 3 compared to 3 vs. 9). The ratio effect further refines this account by showing that quantities closer in ratio (i.e., approaching 1) are harder to discriminate, reflecting increased similarity between their underlying representations.

Importantly, evidence suggests that such distance and ratio effects are not limited to numerical stimuli, but also extend to linguistic representations of magnitude. Comparisons of social-status words, for example, are slower when the status levels are closer together, mirroring the numerical distance effect (Chiao et al., 2004). It has also been observed in other ordered semantic domains, such as time, quality, and temperature (Holyoak & Walker, 1976), as well as in uncertainty-related words (Jaffe-Katz et al., 1989). Moreover, research suggests that linguistic magnitudes display a non-linear organization across quantifiers (Pezzelle et al., 2018) and gradable adjectives (Varma et al., 2024), pointing to a ratio-based organization in the linguistic domain, more akin to the logarithmically compressed line that typically characterizes quantity representations.

Another characteristic feature of the mental organization of quantity is reflected in systematic spatial associations between magnitude and response codes. This phenomenon, known as the Spatial–Numerical Association of Response Codes (SNARC) effect (Dehaene et al., 1993), refers to the tendency to map numerical magnitudes onto a horizontal spatial continuum. Specifically, Western populations respond faster to small numbers with their left hand and to larger numbers with their right hand (e.g., Shaki et al., 2009). This pattern has been interpreted as evidence for a 'mental number line' (Galton, 1880), along which numerical magnitude increases from left to right.

Although originally described for symbolic numerical quantities, the SNARC effect has also been observed for other

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

types of magnitude information. Spatial associations have been reported for ordered sequences such as months and days (Gevers et al., 2003, 2004), as well as for perceptual dimensions including pitch frequency (Ariga & Saito, 2019; Rusconi et al., 2006) and luminance (Fumarola et al., 2014). Moreover, converging evidence suggests the existence of a horizontal mental timeline, with temporal concepts mapped from left (past) to right (future) (e.g., Ishihara et al., 2008; Santiago et al., 2007; Weger & Pratt, 2008).

While much of this work has focused on perceptual inputs, such as visual and auditory stimuli, the linguistic dimension also plays a crucial role. Research on quantity-related words has only recently begun to emerge. Indeed, the Spatial-Linguistic Association of Response Codes (SLARC) effect has been demonstrated. Participants respond faster to small quantifiers (e.g., few) with their left hand and to larger quantifiers (e.g., many) with their right hand (Abbondanza et al., 2021; Dooley, 2021). Specifically, using a magnitude comparison task, Abbondanza and colleagues (2021) investigated the spatial representation of both numerical (e.g. "five", "three") and non-numerical (e.g."some","many") quantifiers. Their findings revealed that non-numerical quantifiers are mentally ordered from left to right, as are numerical quantifiers. Furthermore, they observed a correlation between the performance in the two tasks: participants who showed a stronger left–right mapping for non-numerical quantifiers also exhibited a stronger SNARC effect for numerical quantifiers. Similarly, Dooley (2021) extended this work by demonstrating a comparable association effect between probability ratings (e.g., "never", "certainly") and response side, further supporting the robustness of spatial mappings in linguistic contexts.

However, most evidence for spatial–magnitude associations in language comes from tasks in which participants are required to process magnitude information, for instance by judging which of two stimuli denotes a larger quantity. A central debate regarding these associations concerns their automaticity. In numerical cognition, the SNARC effect is often characterized as an automatic byproduct of magnitude processing (Fias, 1996), suggesting that the mere perception of a number symbol activates its corresponding position on the mental number line.

By contrast, evidence in the linguistic domain remains mixed: while some studies report spatial associations for object names even in indirect tasks (Sellaro et al., 2015), others suggest that such effects fail to emerge when magnitude processing is not task-relevant, as observed, for example, in the temporal domain (Santiago et al., 2025). It therefore remains an open question whether linguistic magnitudes—which require more complex semantic decoding than digits—possess the same degree of automatic activation.

A useful framework for addressing this issue distinguishes between direct and indirect tasks, depending on whether magnitude is task-relevant (Prpic et al., 2016). Critically, indirect tasks provide a stringent test of automaticity, as any spatial–magnitude association observed under these conditions cannot be attributed to task-driven processing. This distinction is particularly informative in light of the contrast between numerical and linguistic evidence: whereas spatial–magnitude associations reliably emerge for numbers even in indirect tasks (e.g., Fias, 2001), it remains unclear whether linguistic stimuli exhibit the same degree of automatic activation.

In parallel, previous research has distinguished between explicit and implicit forms of spatial–magnitude associations based on whether spatial processing is required by the task instructions (Cipora et al., 2018). In explicit tasks, participants are directly instructed to represent or manipulate stimuli in space, whereas in implicit tasks no reference to space is made, and spatial–magnitude associations are inferred indirectly from behavioral measures such as response times. These two forms may rely on partially different mechanisms: implicit effects are often considered more automatic and potentially grounded in overlearned or biologically constrained representations, whereas explicit effects may be more flexible and shaped by cultural or task-specific contexts (Eccher et al., 2025).

To disentangle the contribution of task demands to linguistic spatial–magnitude associations, we adopted a tripartite experimental approach. In Experiment 1, we employed an explicit spatial arrangement task, in which participants were directly required to represent stimuli in space, to establish the presence of a "mental linguistic line." In Experiment 2, we used a lexical decision task, an implicit-indirect paradigm, in which neither spatial nor magnitude information was relevant to the task, thus providing a stringent test of whether spatial–magnitude associations are automatically activated during word processing. Finally, Experiment 3 served as a critical within-subjects validator, using a numerical magnitude classification task to confirm that the same participant pool exhibits a robust, automatic SNARC effect for symbolic numbers. By including quantifiers and scale-ordered Italian gradable adjectives (e.g., tiny, small, big, huge) in our set of stimuli, we tested whether a wider range of linguistic stimuli elicit additional magnitude-related phenomena, such as the distance and ratio effects, alongside spatial associations. We have chosen gradable adjectives as their semantic content is based on a scalar concept that supports orderings of the objects in their domains (e.g., Kennedy & McNally, 2005).

Methods

Participants

A total of 68 participants (58 females; Mage=20.9) completed the three experimental tasks on different days. Inclusion criteria were: age of at least 18 years, right-handedness, native proficiency in Italian, and normal or corrected-to-normal vision. Participants were unaware of the study's purpose. They were recruited through the SONA system (https://www.sona-systems.com) and received university credits as compensation. Written informed consent was obtained from all participants prior to participation.

Materials

Stimuli were 42 Italian adjectives and non-numerical quantifiers. In Italian, most adjectives can be conjugated in both masculine and feminine forms, depending on the suffix (e.g., "alt-o" is the masculine form of "tall," while "alt-a" is the feminine form). In this study, adjectives were used exclusively in their masculine form to avoid possible influences of grammatical gender. Word frequency has been calculated using SUBTLEX-it (https://osf.io/zg7sc/) (Mean=4.49, SD=0.84).

Stimuli contain Italian adjectives explicitly referring to magnitude (e.g., 'basso', which means short), to temporal dimension (e.g., 'lento', i.e., slow), to temperature (e.g., 'caldo', i.e., hot), non-numerical quantifiers (e.g., 'tutti', i.e., all), and emotional adjectives (e.g., 'calmo' and 'arrabbiato', quiet and angry). The reference to scale stimuli was defined in terms of ordered semantic scales, referring to emotions (calm, annoyed, angry, furious), temperature (freezing, cold, hot, boiling), magnitude (tiny, small, large, enormous), and quantifiers (none, few, some, many, all).

Design and Procedure

Participants completed three tasks in a random order on three different days. Each task lasted approximately 20 minutes. All the tasks were created with PsychoPy (Peirce, 2007, 2009; Peirce et al., 2019). Participants were tested online via the Pavlovia platform (https://pavlovia.org/).

Experiment 1: Card Sorting Task Participants were presented with two cards, containing words drawn from the same underlying continuum (e.g., tall–short), one above the other. They were instructed to arrange the words along a horizontal axis using the mouse and to press the spacebar to proceed to the next trial (Figure 1A). Each pair of adjectives was presented twice in each block, for a total of 81 experimental pairs per block. To avoid presentation effects, the card shown first and the hidden one were alternated in the next presentation. A brief break separated the two experimental blocks. The total experiment consisted of 162 trials.

Trials in which the response was omitted, i.e., when participants did not move the cards from the initial position (0.43% of the total trials), were excluded from the analysis.

Experiment 2: Lexical Decision Task In this task, participants made a lexical decision. They needed to decide whether a presented sequence was an Italian word by pressing a key with their right hand for words and a key with their left for non-words (or vice versa). Pseudowords were generated using the Wuggy algorithm for Italian (Keuleers & Brysbaert, 2010).

Each trial began with a central fixation cross (500 ms), followed by the stimulus (3000 ms), during which responses were recorded. A blank screen was then shown for 1000 ms. All stimuli were displayed in white on a black screen (Figure 1B). Half of the participants started by pressing the right key ('L') for words and the left key ('A') for nonwords. In the second part of the experiment, this assignment was reversed. This order was reversed for the other half of the participants. A brief break separated the two experimental blocks.

The same set of stimuli was randomly presented in both parts of the experiment, with each stimulus presented twice per block (336 trials in total).

Participants with overall task accuracy below the a priori defined cutoff of 65% were excluded (n=5), resulting in a final sample of 63 participants. In addition, trials with no response, incorrect responses, or reaction times below 150 ms were excluded (6.48% of all trials).

Experiment 3: Numerical Magnitude Classification Task In this experiment, the stimuli were the number digits from 1 to 29; digit 15 was removed because it served as the task's target. To control for visual length, each digit was presented within a five-character string including placeholder symbols (#). The position of the digit within the string was varied across trials (e.g., ###1#, ##1##, #1###), such that each number appeared in all possible positions while always being preceded and followed by at least one placeholder. At the beginning of the experiment, participants were instructed to classify whether the number shown was lower or higher than 15 using right vs. left keys (i.e., A and L on their keyboards). Each participant performed the task twice, in both congruent (i.e., A – lower than 15; L – higher than 15) and incongruent (A – higher than 15; L – lower than 15) conditions. A brief break separated conditions, and their order was counterbalanced across participants. Each trial began with a central fixation cross (500 ms), followed by a digit (up to 3000 ms). Then, after the participants' response or after 3000 ms, the trial moved to a black screen (1000 ms), ending the trial (Figure 1C).

Participants with overall task accuracy below 65% were excluded from the analyses (n=6). A total of 5.58% of trials were excluded due to reaction-time filtering or omitted responses.

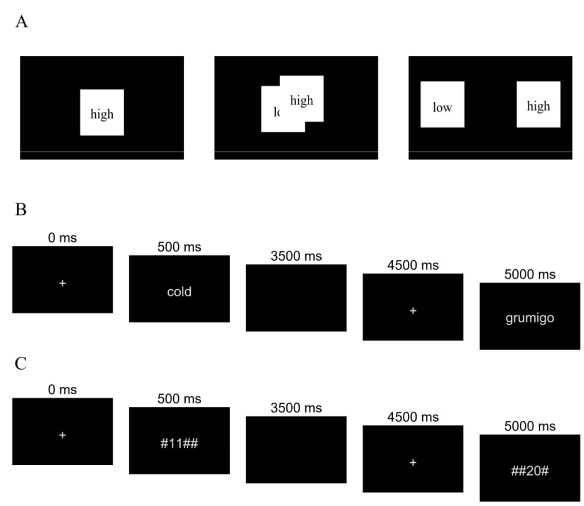

Figure 1: Example of a trial in Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C).

Results

Participants' data were analysed using RStudio (RStudio Team, 2016). Linear mixed-effects models were fitted using lme4 (Bates et al., 2015), and Bayesian models were fitted using brms (Bürkner, 2017).

Experiment 1: Explicit Spatial Mapping: Card Sorting Task

The mean percentage of participants who arranged word pairs in a left-to-right ascending order (small to the left and big to the right) was 77.5%. To test whether there was a specific effect of the word's quantity on its final position, we performed a linear mixed model analysis. First, we calculated the final horizontal position of the word, representing the location participants chose for a specific stimulus relative to the other word. It was computed by subtracting the horizontal value of the opposite word from its horizontal value. Positive values indicated placement on the right side of the screen (with respect to the other word), whereas negative values indicated placement on the left side. We examined whether the horizontal position value could be predicted by the "magnitude" of the word relative to its pair (e.g., "high" conveys a greater quantity than "low," so we expected "high" to be allocated to the right and "low" to the left). Random intercepts were included for participants, semantic dimension, and word length. In addition, random slopes of word quantity for participants were included to allow the effect of word quantity to vary across participants. The analysis has shown that the effect of word's quantity was significant, $t_{(67)}$=7.60, $p < .001$, b=.78 (marginal R^2=.29). This indicates that participants showed a horizontal linguistic quantity line in the task, i.e., they were more likely to position the word of the pair with the smaller quantity on the left compared to the word of the pair with the larger quantity (Figure 2).

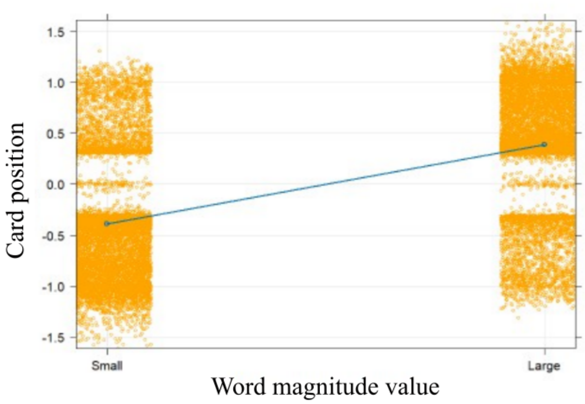

Figure 2: Participants place smaller quantities to the left and larger quantities to the right. The x-axis shows category (small vs. large), and the y-axis represents card position (negative values = leftward, positive = rightward).

Furthermore, we examined whether the spatial distance between words in the task reflected their distance along a gradable scale. To this end, we tested whether participants' computed spatial distance between word pairs was predicted by the ordinal distance between the corresponding terms on the scale, controlling for average word length and average word frequency, and including random slopes for subjects. In this analysis, we considered only words that are part of an underlying gradable scale (i.e., quantifiers, temperature, magnitude, and emotion terms). Following the procedure adopted by Varma and colleagues (2024), words within each scale were ordered from the lowest to the highest magnitude (i.e., from "smaller" to "larger") and assigned sequential numerical values reflecting their relative position along the scale. For instance, the quantifiers scale is: none, few, some, many, all; we assigned a number from 1 (none) to 5 (all) and calculated distance accordingly, i.e., between none and few the distance is 1, whereas between none and some the distance is 2. This effect was significant for all four scales considered ($t(58.91)$= 5.44, $p < .001$, b=.12) (Figure 3A), this indicates that quantitative distance along the scale predicted card placement, such that items that were farther apart were also positioned at greater spatial distances. This held for each scale: quantifiers ($t_{(70.67)}$=9.43, $p < .001$, b=.13), emotions ($t_{(95.88)}$=9.40, $p < .001$, b=.17), temperature ($t_{(111.73)}$=6.63, $p < .001$, b=.15), and magnitude ($t_{(69.43)}$=8.90, $p < .001$, b=.18.

Crucially, we further tested whether word positions were better accounted for by the ratio between word pairs, reflecting a logarithmically compressed representation of quantity. Under this account, quantities with larger ratios should be positioned closer together than those with smaller ratios. We computed ratios between stimuli using the same metrics calculated for distance. For instance, in the temperature scale (freezing–cold–hot–boiling), the ratio between boiling and

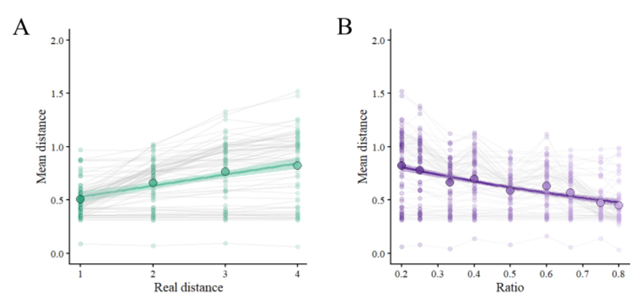

Figure 3: Distance and ratio effects in the card-sorting task. (A) Greater semantic distance between scale items was associated with greater spatial separation of the cards. Pale dots represent each participant's mean judged distance across trials at a given semantic distance; dark points represent the group mean. (B) Greater semantic ratios were associated with smaller card separation. Pale dots represent participant means; dark points represent the group mean.

hot (3/4) is higher than the ratio between cold and freezing (1/2) and therefore predicts a shorter spatial distance between the former pair. We therefore examined whether participants' spatial arrangements were predicted by these ratio-based distances. The mixed-effects model showed a significant effect of the ratio predictor on spatial distance, $t_{(52.15)}$=–4.82 , p < .001, b=.54, indicating that the relative ratio between magnitudes reliably predicted participants' spatial arrangements (Figure 3B).

Experiment 2: Implicit Spatial Mapping: Lexical Decision Task

Participants' overall accuracy was high (mean=96.23%; range=88.35%-99.43%).

We tested whether response-key congruency with the quantity evoked by the word predicted faster reaction times (log-transformed), using a linear mixed-effects model that included random intercepts for participants, word length, and word frequency in Italian. This analysis revealed no significant effect of congruency on reaction times, (p = .93), indicating that participants' reaction times did not differ between the congruent condition (i.e., when the magnitude evoked by the word corresponded to the response side in space, left key/small quantity words and right key/large quantity words) and the incongruent condition. Beyond traditional null-hypothesis testing, a Bayesian model comparison was conducted to quantify the evidence for the absence of a SLARC effect. The analysis yielded a Bayes factor of BF_{01}=87, providing very strong evidence in favor of the null hypothesis (Figure 4). This indicates that the data are approximately 87 times more likely under a model excluding congruency than under a model including it, supporting the conclusion that spatial associations do not emerge automatically in this implicit linguistic task.

Experiment 3: SNARC effect: magnitude classification task

To assess the presence of an SNARC effect, we fitted a linear mixed-effects model predicting the difference in reaction times (log-transformed) between left and right responses, with numerical magnitude as the fixed effect and random intercepts for

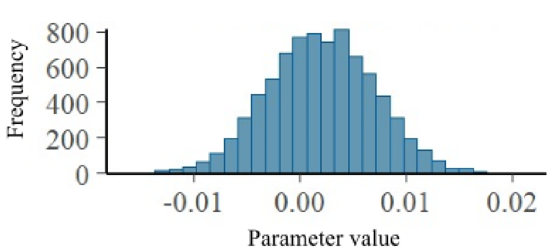

Figure 4: The posterior distribution of the congruency effect in Experiment 2 was centered around zero, indicating no effect of congruency on participants' reaction times.

participants. The analysis revealed a significant effect of numerical magnitude on reaction time differences, $t_{(1673)}$=9.96, p < .001, b=.00, indicating that larger numbers were associated with relatively faster right-hand responses compared to left-hand responses, consistent with an SNARC effect.

Moreover, we tested for the presence of distance and ratio effects. A strong distance effect was observed, with reaction times decreasing as the numerical distance from the reference value (15) increased ($t_{(58.52)}$=–21.55, p < .001 , b=.01). Reaction times also increased as the numerical ratio between the target number and the reference value approached 1, indicating a clear ratio effect consistent with Weber's law ($t_{(59.08)}$=17.30, p < .001, b=.2).

In addition, participant-specific slopes were extracted for each effect (SNARC, distance, and ratio) and compared with the corresponding slopes obtained in Experiment 1 in order to assess cross-task correlation of these effects. No significant associations were observed between the slopes across tasks.

Discussion

Across three experiments, we examined whether linguistic stimuli referring to approximate magnitudes give rise to hallmark behavioral signatures of magnitude cognition, including spatial mapping, distance effects, and ratio effects, in order to explore whether and to what extent magnitude representation is involved in processing these stimuli.

Experiment 1 provided clear evidence for an explicit left–right organization of linguistic magnitude. In a card-sorting task, participants reliably arranged words according to their relative magnitude, placing smaller-quantity words to the left and larger-quantity words to the right. Importantly, this spatial organization was systematically modulated by the relative distance between words. Greater differences in word magnitude were associated with greater spatial separation, and this pattern held across quantifiers, temperature terms, and emotion terms.

Moreover, participants' spatial arrangements were also sensitive to ratio-based differences between magnitudes. Spatial distances were better accounted for by the ratio between word magnitudes than by absolute differences alone, consistent with a logarithmically compressed representation of quantity. These findings align with proposals that magnitudes across different continuous domains are supported by a single, shared representational system, abstract in nature and independent of sensory modality, namely the general magnitude system (GMS) (Gallistel & Gelman, 2000; Lourenco & Longo, 2011; Walsh, 2003). Indeed, in this system, magnitudes are encoded as continuous and overlapping distributions along a shared metric, with increasing representational uncertainty for larger magnitudes. Together, the emergence of three characteristic effects of magnitude perception in this task (spatial organization, distance, and ratio effects) supports the involvement of magnitude processing in the representation of these linguistic stimuli.

In contrast, Experiment 2 did not reveal evidence for an implicit spatial congruency effect in reaction times. Participants' response times were not modulated by congruency between the response side and the quantity evoked by the word. Crucially, Bayesian model comparison yielded very strong evidence in favor of the null hypothesis, indicating that the data is nearly 90 times more likely under a model excluding congruency. This strongly suggests that, unlike explicit spatial judgments, linguistic magnitude may not automatically engage spatial response codes under the present task demands. These results stand in contrast to previous studies reporting a left-to-right spatial organization of similar terms even in implicit tasks (e.g., Abbondanza et al., 2021; Dooley, 2021). One possible explanation for this discrepancy is that, although spatial processing was not explicitly required in those studies, the quantitative dimension was nevertheless made explicitly task-relevant, for example, by asking participants to determine which of two vague quantifiers denoted a greater magnitude. Thus, the absence of direct magnitude-based categorization demands in the present task may have limited the engagement of a spatially organized mental representation for linguistic stimuli referring to approximate quantities. This suggests a difference in the degree of automaticity with which magnitude representations are engaged when processing numerical symbols as compared to linguistic expressions referring to approximate quantities, as numerical magnitude, despite being task-irrelevant, automatically activates spatial codes that can interfere with task-related responses (Fias, 2001; Fischer, 2003; Roth et al., 2025). Converging evidence for this asymmetry comes from size congruity paradigms with gradable adjectives, which reveal interference of physical size during semantic magnitude judgments, but not vice versa (Kochari & Schriefers, 2022). By contrast, Arabic digits reliably produce bidirectional size congruity effects, consistent with a more automatic involvement of magnitude representations (Henik & Tzelgov, 1982). This asymmetry may stem from the greater semantic complexity and context-dependency of linguistic stimuli; whereas a digit provides direct access to a precise quantity, adjectives and quantifiers may require deeper, more deliberative processing to 'release' their magnitude information, preventing it from interfering with speeded lexical decisions.

Finally, Experiment 3 replicated classic magnitude effects in the numerical domain. A robust SNARC effect was observed, along with clear distance and ratio effects in numerical magnitude comparison. However, no evidence was found for a systematic relationship between the magnitude effects observed in Experiment 1 and those observed in Experiment 3. In particular, individual sensitivity to the distance effect in the card-sorting task did not correlate with sensitivity to the distance effect in the numerical magnitude classification task, and the same pattern held for the other effects. This dissociation further strengthens the view that numerical processing and the processing of these linguistic expressions may not be equally grounded in a generalized magnitude system. Specifically, it suggests that while both domains may share a psychophysical architecture (i.e., ratio-dependency), the mapping of these representations onto space is not a singular, unified trait but is instead format-specific and task-dependent. Notably, Varma and colleagues (2025) similarly attempted to relate magnitude effects observed in a task involving adjectival quantifiers to corresponding effects in a numerical magnitude comparison task, but did not find significant associations. By contrast, other authors have reported correlations between SLARC and SNARC effects in the context of quantifier processing, indicating that under certain task conditions spatial–magnitude associations may generalize beyond the numerical domain (e.g., Abbondanza et al., 2021). In the present case, the absence of correlations across experiments may be attributable to differences in task demands and dependent measures: Experiment 1 primarily indexed spatial positioning, whereas Experiment 3 relied on reaction-time–based measures. Future work should aim to assess magnitude effects using more closely matched task structures. It would also be informative to examine whether correlations emerge when the card-sorting task employed in Experiment 1 is administered using numerical stimuli, or conversely, whether dissociations persist even when identical stimuli are tested across different task formats. Moreover, investigating these effects using number words (linguistic symbols for magnitude) rather than Arabic digits may provide a more appropriate comparison, as number words are likely to share greater processing similarities with adjectives and vague quantifiers, whereas Arabic digits afford more direct and automatic access to numerical magnitude (Damian, 2004; Kadosh et al., 2008).

Taken together, these results indicate that linguistic expressions referring to magnitude can engage spatial and representational mechanisms characteristic of magnitude cognition, particularly when magnitude relations are made explicit. At the same time, the absence of an implicit congruency effect suggests important limits on the automaticity with which linguistic magnitude maps onto spatial representations. This lack of correlation suggests that, although linguistic and numerical stimuli can give rise to superficially similar behavioral signatures of magnitude processing, these effects may not be driven by a single, unified mechanism operating identically across domains. Instead, magnitude representations may be flexibly recruited depending on task demands and representational format, with explicit spatial tasks promoting alignment between linguistic meaning and spatial structure, while numerical comparison engages more automatic, domain-specific processes.

Acknowledgments

AI tools were used to assist with writing and analysis code.

References

Abbondanza, M., Rinaldi, L., Foppolo, F., & Marelli, M. (2021). The mental representation of nonnumerical quantifiers: The spatial-linguistic association of response codes (slarc) effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(12).

Ariga, A., & Saito, S. (2019). Spatial–musical association of response codes without sound. *Quarterly Journal of Experimental Psychology*, *72*(9), 2288–2301.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*(6), 1131–1136.

Bürkner, P. C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28.

Chiao, J. Y., Bordeaux, A. R., & Ambady, N. (2004). Mental representations of social status. *Cognition*, *93*(2), B49–B57.

Cipora, K., Schroeder, P. A., Soltanlou, M., & Nuerk, H. C. (2018). More space, better mathematics: Is space a powerful tool or a cornerstone for understanding arithmetic? In *Visualizing mathematics: The role of spatial reasoning in mathematical thought* (pp. 77–116). Springer.

Damian, M. F. (2004). Asymmetries in the processing of arabic digits and number words. *Memory & Cognition*, *32*(1), 164–171.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.

Dooley, E. (2021). *Snarc effect for verbal probability quantifiers in economic decision making task* [Master's thesis, Humboldt University] [Master's thesis].

Eccher, E., Josserand, M., Caparos, S., Boissin, E., Buiatti, M., Piazza, M., & Vallortigara, G. (2025). A left-to-right bias in number-space mapping across ages and cultures. *Nature Communications*, *16*, 495.

Fias, W. (1996). The importance of magnitude information in numerical processing: Evidence from the snarc effect. *Mathematical Cognition*, *2*(1), 95–110.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the snarc effect. *Psychological Research*, *65*(4), 250–259.

Fischer, M. (2003). Spatial representations in number processing—evidence from a pointing task. *Visual Cognition*, *10*(4), 493–508.

Fumarola, A., Prpic, V., Da Pos, O., Murgia, M., Umiltà, C., & Agostini, T. (2014). Automatic spatial association for luminance. *Attention, Perception, & Psychophysics*, *76*(3), 759–765.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65.

Galton, F. (1880). Visualised numerals. *Nature*, *21*(533), 252–256.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87–B95.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex*.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395.

Holyoak, K. J., & Walker, J. H. (1976). Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior*, *15*(3), 287–299.

Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the stearc effect. *Cortex*, *44*(4), 454–461.

Jaffe-Katz, A., Budescu, D. V., & Wallsten, T. S. (1989). Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition*, *17*(3), 249–264.

Kadosh, R. C., Henik, A., & Rubinsten, O. (2008). Are arabic and verbal numbers processed in different ways? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1377–1391.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*(2), 345–381.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633.

Kochari, A. R., & Schriefers, H. (2022). Processing symbolic magnitude information conveyed by number words and by scalar adjectives. *Quarterly Journal of Experimental Psychology*, *75*(3), 422–449.

Lourenco, S. F., & Longo, M. R. (2011). Origins and development of generalized magnitude representation. In *Space, time and number in the brain* (pp. 225–244).

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520.

Peirce, J. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13.

Peirce, J. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, *2*, 343.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pezzelle, S., Bernardi, R., & Piazza, M. (2018). Probing the mental representation of quantifiers. *Cognition*, *181*, 117–126.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499–503.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*(6), 983–993.

Prpic, V., Fumarola, A., De Tommaso, M., Luccio, R., Murgia, M., & Agostini, T. (2016). Separate mechanisms for magnitude and order processing in the snarc effect. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1241–1251.

Roth, L., Caffier, J., Reips, U. D., Cipora, K., Braun, L., & Nuerk, H. C. (2025). True colors snarc: Semantic number processing is highly automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

RStudio Team. (2016). *Rstudio: Integrated development for r*. RStudio, Inc. Boston, MA. http://www.rstudio.com/

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The smarc effect. *Cognition*, *99*(2), 113–129.

Santiago, J., Beracci, A., Flumini, A., Sanjuan, E., Ouellet, M., & Solana, P. (2025). Can the lateral mental timeline be automatically activated in language comprehension? *Journal of Memory and Language*, *143*, 104644.

Santiago, J., Lupáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*(3), 512–516.

Sellaro, R., Treccani, B., Job, R., & Cubelli, R. (2015). Spatial coding of object typical size: Evidence for a snarc-like effect. *Psychological Research*, *79*(6), 950–962.

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the snarc effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331.

Szymanik, J., & Kieraś, W. (2022). The semantically annotated corpus of polish quantificational expressions. *Language Resources and Evaluation*, *56*(3), 1057–1074.

Varma, S., Sanford, E. M., Marupudi, V., Shaffer, O., & Lea, R. B. (2024). Recruitment of magnitude representations to understand graded words. *Cognitive Psychology*, *153*, 101673.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488.

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review*, *15*(2), 426–430.