Evidence of context-dependence in people's representations of negative numbers

Nicholas A. Vest[1] **& Vijay Marupudi**[2]
[1]School of Teaching and Learning, University of Florida
[2]Department of Psychology, University of California, Berkeley

Abstract

Negative integers lack direct perceptual referents for humans and require integrating magnitude with polarity to use effectively. Prior work on integer representation has produced inconsistent findings and multiple hypotheses. We propose that these inconsistencies reflect context-dependent recruitment of multiple coexisting representations. Adults completed a symbolic magnitude comparison task involving positive – positive, negative – negative, and mixed – sign pairs, with presentation format (simultaneous vs. sequential) and numerical context (grouped vs. ungrouped comparison type) manipulated. We evaluated the componential, extension, and reflection hypotheses using individual-level Bayesian models. Group-level analyses revealed no distance effect for mixed-sign pairs. Individual-level modeling unveiled participants flexibly recruiting different representations across contexts, with extension-based representations emerging primarily when task demands were minimized. Reflection-based representations were rare. These findings highlight the flexibility of integer representations and emphasize that group-level analyses can obscure the dynamics of individuals' internal representations.

Keywords: negative numbers; integer representation; mental number line; individual differences; Bayesian modeling

Introduction

Negative integers pose a fundamental challenge for cognition because they lack direct perceptual referents and require reasoning about magnitude in a relational, bidirectional space. Whereas positive numbers can be grounded in counting and accumulation, negative numbers denote relations to a reference point, such as debts, deficits, or direction relative to zero (Kilhamn, 2011; Vlassis, 2004). As a result, understanding negative numbers requires extending or reorganizing representational systems that arose to support reasoning about positive quantities. For this reason, negative integers serve as a valuable test case for understanding how cognitive systems adapt to support increasingly abstract mathematical concepts.

A central tool for reasoning about numerical magnitude is the mental number line (MNL), an analog spatial representation that supports comparison, estimation, and ordering (Dehaene, 1992; Hubbard et al., 2005). Extensive evidence suggests that positive numbers are represented as positions on a logarithmically compressed ordered continuum, exhibiting distance effects consistent with Weber-Fechner scaling (Dehaene, 1992; Moyer & Landauer, 1967). However, how negative numbers are represented within or alongside this system remains an open question. It is not trivial to extend the MNL to include values below zero. It requires coordinating magnitude with polarity and redefining zero as both a boundary and a reference point (Bofferding, 2014; Varma & Schwartz, 2011).

Three broad theoretical accounts have been proposed to explain how negative integers are mentally represented by people. Componential accounts posit that negative numbers are processed by decomposing them into separable symbolic elements: an absolute magnitude and a polarity indicator (Ganor-Stern & Tzelgov, 2008; Tzelgov et al., 2009). In this view, magnitude processing is restricted to positive values, and negative numbers are interpreted by applying learned rules (e.g., ''all positives are greater than negatives''). Extension accounts propose that negative numbers are integrated into the existing magnitude system by extending the mental number line leftward from zero, preserving ordinal and metric relations across the integer domain (Fischer & Rottmann, 2005). Reflection accounts argue for a deeper reorganization in which negative numbers are represented as reflections of positive numbers around zero, encoding additive inverse relations and symmetry in the magnitude system (Varma & Schwartz, 2011).

These accounts make distinct predictions for symbolic magnitude comparison tasks, particularly for comparisons involving both positive and negative numbers. Componential processing predicts little or no sensitivity to numerical distance for mixed-sign comparisons, as judgments can be made using polarity alone. Extension accounts predict a standard distance effect across all comparison types, while reflection accounts predict an inverse distance effect for mixed-sign pairs due to psychophysical scaling and symmetry around zero (Varma & Schwartz, 2011).

Empirical findings, however, have been inconsistent. Some studies report no distance effect for mixed-sign comparisons (suggesting componential processing), others report inverse distance effects (supporting reflection), and still others find standard distance effects under certain task conditions (Fischer & Rottmann, 2005; Ganor-Stern et al., 2010; Shaki & Petrusic, 2005; Varma & Schwartz, 2011). Rather than converging on a single representational account, the literature reveals substantial variability across tasks, presentation formats, and participant samples. Importantly, most prior work has relied on group-level analyses of reaction time, implicitly assuming that a single representation dominates within a given context.

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

An alternative interpretation is that multiple representations coexist and are flexibly recruited depending on task demands. From this perspective, variability across studies is not noise but evidence of adaptive representational use. Task features such as simultaneous versus sequential presentation, or grouped versus ungrouped comparison types, may differentially support rule-based versus integrated magnitude processing by altering working memory demands, strategy switching costs, and the salience of polarity cues (Ganor-Stern et al., 2010; Varma & Schwartz, 2011). If so, group-level averages may obscure meaningful individual differences in representational recruitment.

The present study adopts this flexibility perspective and tests it using an individual-level modeling approach. We examine adults' performance on a symbolic magnitude comparison task involving positive-positive, negative-negative, and mixed positive-negative pairs. We systematically manipulate presentation format (simultaneous vs. sequential) and numerical context (grouped vs. ungrouped) to vary cognitive demands and contextual support for different representations. Rather than inferring representations from aggregate distance effects alone, we formally compare competing computational models corresponding to componential, extension, and reflection accounts for each participant and condition.

Our primary research questions are:

1. Do people differ in the representations they use to process negative integers, particularly for mixed-sign comparisons?

2. How does task context—specifically presentation format and numerical grouping—shape the recruitment of componential versus integrated magnitude representations?

3. To what extent do group-level distance effects mask underlying heterogeneity in individual representational strategies?

By addressing these questions, this work aims to clarify how negative integers are represented and used in real time, and to demonstrate the value of model-based, individual-level analyses for understanding flexibility in abstract numerical cognition.

Methods

Participants

Participants were 60 undergraduate students enrolled in a course in the Department of Psychology at the University of Wisconsin-Madison. These students were recruited from the departmental subject pool. Six participants with an error rate exceeding 45% were excluded, resulting in a final sample of 54 participants. Participants' average age was 19 years and 7 months (SD = 9 months). There were 43 women and 11 men. Thirty-eight were White (of these, 3 indicating Hispanic ethnicity), 8 Asian, 2 Native American (of these, 2 indicating Hispanic ethnicity), 4 some other race, and 2 bi- or multiracial. Participants received extra credit for their course for completing the experiment.

Design and Procedure

This study was designed to test competing accounts of negative number representation: the components account, the extension account, and the reflection account. These accounts were evaluated by examining reaction time (RT) patterns during symbolic magnitude comparison, where participants judged which of two numbers was either greater or lesser. This factor was counterbalanced within participant. Participants first began the experiment by viewing the instructions and were encouraged to take breaks between blocks. Stimuli consisted of pairs of integers formatted to include a ''+'' sign for positive values (e.g., ''+4'', ''-3''), following the recommendation of Krajcsi and Igács (2010). Stimulus order (grouped vs ungrouped comparison type) was manipulated within subjects and presentation format (simultaneous vs. sequential) was assigned between subjects. In simultaneous trials, the two numbers to be compared were presented next to each other. In sequential trials, each number was presented individually in the center of the screen for 1000 ms, with no fixation cross or inter-stimulus interval between the two presentations. Participants were then asked to respond by using the ''A'' key to indicate the left or first number and the ''L'' key to indicate the right or second number. Sessions lasted approximately 45 minutes and were presented online using jsPsych (De Leeuw, 2015).

Stimuli

The stimuli included integers ranging from -9 to 9. Number pairs containing the culturally significant boundary numbers 0 and 5 were excluded. These were excluded to avoid potential conceptual and perceptual biases, consistent with prior studies that omit such values due to their special status in both mathematical conventions and cognitive processing (e.g. Lyons & Beilock, 2011; Varma & Schwartz, 2011). This resulted in three categories of comparisons: positive-positive (e.g., 2 vs. 4), negative-negative (e.g., -2 vs. -4), and mixed positive-negative (e.g., -2 vs. 4). Additionally, pairs with equal absolute magnitudes but opposite signs (e.g., -1 vs. 1) were excluded to discourage reliance on strategies that ignored magnitude entirely (see Vest, 2025 for list of stimuli).

The numerical distance between the two numbers in each pair was limited to range between 1 to 8 to prevent participants from associating mixed comparisons with larger differences between the numbers. Additionally, stimuli were constructed to balance size across distance levels to avoid confounding numerical distance with overall number size. For positive-positive and negative-negative comparisons, near-distance stimuli paired two small integers or two large integers, while far-distance stimuli paired a small integer with a large integer. For negative numbers, those closer to zero (e.g., -1) were treated as large, and those further from zero (e.g., -9) were treated as small. For mixed comparisons, near-distance stimuli paired a large negative integer with a small positive integer (e.g., -1 vs. 2), while far-distance stimuli paired a small negative integer with a large positive integer (e.g., -9 vs. 7), or

Table 1: Models embodying competing integer representations that were fit for each participant and condition (simultaneous vs. sequential and grouped vs. ungrouped). n_1 and n_2 represent the smaller and larger numbers being compared respectively.

Intercept only	Extension model	Components model	Reflection model
$\begin{aligned} \alpha &\sim \mathcal{N}(0,3) \\ \sigma &\sim \text{HalfNormal}(3) \\ \log(\text{RT}) &= \mathcal{N}(\alpha, \sigma) \end{aligned}$	$\begin{aligned} w &\sim \text{softplus}(\mathcal{N}(10,5)) \\ p &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{n_2}{\sqrt{2}w}\right)\right] - \frac{1}{2}\left[1 + \text{erf}\left(\frac{n_1}{\sqrt{2}w}\right)\right] \\ \alpha &\sim \mathcal{N}(0,3) \\ \beta &\sim \mathcal{N}(0,3) \\ \sigma &\sim \text{HalfNormal}(3) \\ \log(\text{RT}) &= \alpha + \beta p + \mathcal{N}(0,\sigma) \end{aligned}$	$\begin{aligned} a &\sim \mathcal{N}(0,1) \\ f &\sim \mathcal{N}(0,1) \\ w &\sim \text{HalfNormal}(1) + 0.001 \\ d &= \begin{cases} a & \text{if } n_2 > 0 \oplus n_1 > 0 \\ f + \frac{1}{2}\left[1 + \text{erf}\left(\frac{\frac{n_1}{n_2} - 1}{\sqrt{2}w\sqrt{1+\left(\frac{n_1}{n_2}\right)^2}}\right)\right] & \text{if } n_2 < 0 \wedge n_1 < 0 \\ \frac{1}{2}\left[1 + \text{erf}\left(\frac{\frac{n_2}{n_1} - 1}{\sqrt{2}w\sqrt{1+\left(\frac{n_2}{n_1}\right)^2}}\right)\right] & \text{else} \end{cases} \\ \alpha &\sim \mathcal{N}(0,3) \\ \beta &\sim \mathcal{N}(0,3) \\ \sigma &\sim \text{HalfNormal}(3) \\ \log(\text{RT}) &= \alpha + \beta d + \mathcal{N}(0,\sigma) \end{aligned}$	$\begin{aligned} m(x,b,q) &= \begin{cases} \log_b(x) & \text{if } x > 0 \\ q \log_b(-x) & \text{if } x \leq 0 \end{cases} \\ b &\sim \text{HalfNormal}(1.5) + 1 \\ q &\sim \text{HalfNormal}(1.5) + 1 \\ d &= \lvert m(n_2,b,q) - m(n_1,b,q) \rvert \\ \alpha &\sim \mathcal{N}(0,3) \\ \beta &\sim \mathcal{N}(0,3) \\ \sigma &\sim \text{HalfNormal}(3) \\ \log(\text{RT}) &= \alpha + \beta d + \mathcal{N}(0,\sigma) \end{aligned}$

vice versa (e.g., -7 vs. 9). This construction scheme equated the numerical distance of the near and far levels across comparison types and ensured that distance and size were not confounded, particularly in the mixed condition.

74 unique number pairs met the inclusion criteria. Each pair was presented in both left-right and right-left order (e.g., 1 vs. 9 and 9 vs. 1), yielding 148 trials per comparison prompt. Since adults experienced two stimulus order conditions (grouped and ungrouped, a within-subjects factor) and two comparison prompts (i.e., ''choose the greater number'' or ''choose the lesser number''), they completed a total of 592 trials across the entire experiment.

Modeling

A core contribution of this paper is the individual difference models of participants' behavior. Four models were specified to compare the hypothesized representations to each other. These were the intercept, component, extension, and reflection models. The mathematical definitions of these models are presented in Table 1 and the differences in model predictions are visualized in Figure 1. The intercept model was included as a conservative null hypothesis in order to avoid making erroneous conclusions on noisy data. If the intercept model is selected as the best model during model selection, we conclude that data was not variable enough to discriminate between the competing models.

The component model enacts the componential strategy of first comparing the sign and responding quickly if they differ. If the signs do not differ, it assumes participants compare the magnitudes of the numbers. If the numbers are both negative, then the participants spend some additional time reversing their judgment. For magnitude comparison, people are assumed to be sensitive to the ratio between the two numbers, as outlined by the Weber-Fechner law (Fechner, 1860).

The extension model assumes an extension of the conventional logarithmically compressed mental number line representation, with an increase in precision around zero. This was modeled using a novel psychophysical function that respects this property while retaining the sublinear compression of numbers with high absolute magnitude. The reflection model follows the specification by Varma and Schwartz (2011), represented by a mirrored negative number line representation. Instead of the original inverse power law, we use a similar logarithmic curve to maintain consistency with the competing models. As can be observed in Figure 1, the models make similar predictions between positive-positive and negative-negative comparisons, but differ for mixed comparisons.

By using logarithmically compressed representations, these models take both the distance and size effects into account simultaneously compared to the separate linear models approach used in past work. They serve as a stronger test of the hypothesized internal representations, as the same parameters have to account for the distance effect, size effect, and predict RT across all comparison types. This is particularly important for the extension and reflection hypotheses, which assume that we use the same magnitude comparison machinery across comparison types. We fit all four models to every participant's data for each of the four conditions. This helps us identify if participants are dynamically recruiting differences representations based on the context.

These Bayesian models were specified and fit using numpyro (Phan et al., 2019). For each participant, we determined the best fitting model, i.e., the most likely integer representation, by extracting the MAP estimates of the posterior distributions of the parameters and using each models' AICc score to select between different models per participant. Since the AICc metric includes a penalty for each additional parameter in a model, a complex model with more parameters is only selected if the additional parameters increase fit beyond the penalty.

Results

First, we present the results of analyzing group-level distance effects similar to past studies. Then, we present the results of the fitting Bayesian models to participants data and using AICc to select between them.

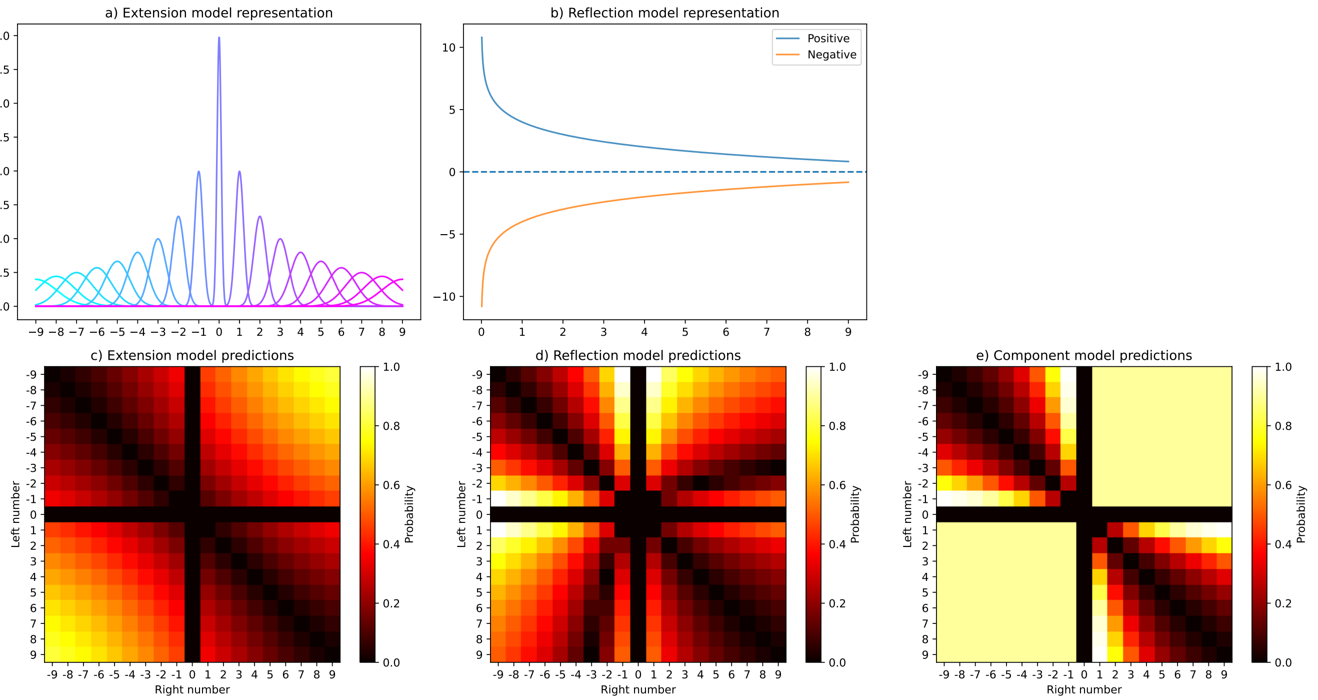

Figure 1: Depictions of the competing integer representations. a) A representation of the extension hypothesis, where integers with smaller absolute magnitudes can be indexed with higher precision compared to those with larger absolute magnitudes. b) A representation of the reflection hypothesis, where the vertical distance between points on the lines indicates perceived psychological distance. Plots c), d), and e) visualize model predictions, with darker colored number pairs predicted to take more time to compare compared to brighter pairs. Model predictions for mixed pairs vary drastically.

Group-Level Distance Effects

To assess distance effects at the group level, we fit linear mixed-effects models (LMEMs) predicting RTs. Each model included numerical distance (1-8) as a continuous predictor, as well as order (grouped vs. ungrouped), presentation format (simultaneous vs. sequential), and their three-way interaction, along with all lower-order terms. Predicate was included as a covariate. Random intercepts and random slopes for within-subject predictors were included to account for individual variability. Separate models were fit for positive, negative, and mixed comparisons. F values and omnibus tests were calculated using the lmerTest package, which applies Satterthwaite's approximation for degrees of freedom.

As expected, a standard distance effect was observed in positive comparisons: participants responded more slowly to pairs that were closer in magnitude, $\beta = -8.45$, $F(1, 53.18) = 15.13$, $p < .001$. Note that because the beta reflects change in RT, a negative beta indicates faster responses as numerical distance increases. A similar standard distance effect emerged in negative comparisons, $\beta = -12.72$, $F(1, 202.65) = 51.55$, $p < .001$. However, a standard distance effect could arise from either a components representation that transforms negatives into positive magnitudes, or from a representation that integrates negative numbers onto the mental number line (extension or reflection). Thus, distance effects alone are not diagnostic for negative comparisons.

In contrast, mixed comparisons (those involving both a positive and a negative number) did not show a significant distance effect at the group level, $\beta = -1.82$, $F(1, 6320.5) = 0.50$, $p = .479$; see Table 2 for RT means. This finding initially suggests that participants might be relying on a components representation when comparing mixed-sign pairs—that is, rather than processing magnitude, they use a quick heuristic or rule such as ''all positive numbers are greater than negatives.'' This finding is inconsistent with findings in past work that revealed inverse distance effects for mixed pairs (Varma & Schwartz, 2011; Vest & Alibali, 2021). However, the lack of a significant distance effect at the group level could also reflect an averaging across participants employing different representations. Because these representations produce opposing distance effects (no effect for components, standard effect for extension, inverse effect for reflection), they may cancel each other out when combined in aggregate data. Examining the individual-level variability in how participants represent negative numbers can help disambiguate these representations.

Individual-Level Modeling Results

For each participant in each condition, we selected the model that with the lower AICc score for the median posterior estimates (i.e. MAP) as the best descriptor of their performance.

Table 2: Distance Effects Reported in Milliseconds

Condition	Overall	Far	Near	$t(53)$	p	Direction
Positive	742.06 (4.34)	734.43 (7.01)	770.62 (5.50)	−3.89	< .001	Standard
Negative	803.39 (4.60)	770.01 (7.46)	822.23 (5.82)	−7.18	< .001	Standard
Mixed	742.06 (5.41)	735.07 (7.14)	750.84 (8.28)	−0.71	.479	None
Mixed Simultaneous–Ungrouped						
	1098.45 (8.70)	1092.09 (11.34)	1106.37 (13.51)	-0.67^a	$.502^a$	None
Mixed Simultaneous–Grouped						
	733.46 (5.04)	723.63 (6.44)	745.91 (7.97)	-0.50^a	$.621^a$	None
Mixed Sequential–Ungrouped						
	505.31 (11.30)	499.17 (15.00)	512.93 (17.18)	0.16^a	$.873^a$	None
Mixed Sequential–Grouped						
	507.68 (11.78)	504.95 (15.61)	511.13 (17.97)	-0.40^a	$.692^a$	None

Note. Values are reported as M (SE). t-values and p-values reflect the main or simple effect of numerical distance (1–8), estimated from linear mixed-effects models including condition and predicate. p-values were obtained using Satterthwaite's method via the `lmerTest` package. [a] Simple effects.

The proportion of participants for whom each model was selected is depicted in Figure 2. On a macro scale, we see that performance in the Simultaneous (grouped) condition differed from all other conditions, as the null model was selected only for a small proportion of participants (6.7%) in contrast with the other groups (> 56.7%). This result underscores the importance of these ordering and grouping manipulations and how they change the strategies used by participants.

For the Simultaneous (grouped) condition, an equal proportion of participants (43.3%) were best described by the components model and the extension model. This suggests that large differences exist in participants' recruitment of internal representations. Models performed well at predicting participants' responses as the median R^2 value for participants for whom the intercept model was not selected was 0.12. Performance in this condition is particularly interesting as grouped comparisons reduce the need for strategy flexibility between trials and simultaneous comparisons remove the need to store internal representations in memory. Due to these contextual factors, participants may have felt comfortable using an extension representation. The difference between this condition and the others raises the possibility that working memory may play an important role in moderating strategy flexibility. Nevertheless, there were also a large number of participants who performed most consistently with the component model. When the comparison type is mixed, it can be optimal to solely focus on the difference in the signs to minimize the amount of time taken, and it is possible that participants were relying on this strategy.

The reflection model only best described 6.7% of the participants, i.e. 2 out of 30 participants, in the Simultaneous (grouped) condition. This stands in contrast with prior work, that has largely used the presence of a reverse distance effect to justify the reflection model. The differences might stem due to our modeling paradigm, which is more comprehensive by including all posited competing models, and conservative by penalizing extra parameters using AICc and including a null model. The poor performance may also be due some unintuitive predictions made by the reflection model that were not found in the empirical data. For example, the multiplicative scaling factor of the negative numbers, instead of an added offset, would imply that distances between negative numbers would be perceived as larger than equal distances between positive numbers, as shown in Figure 1. The absence of such a empirical trend would make the model less competitive relative to other models. Similarly, for the Simultaneous (ungrouped), only one participant's performance was most consistent with the reflection model.

The proportions of the selected models for all conditions other than Simultaneous (ungrouped) were similar to each other. We found that for a large proportion of participants (> 56.7%), the intercept model was the selected model. For these participants, the data was too variable to determine the most appropriate model. The median R^2 of selected non-null models was 0.062. Among these models, the component model was the most commonly selected model followed by the extension model and then the reflection model. Interestingly, participants in the sequential condition were less likely to behave most like the extension model compared to the simultaneous condition. This may be due to the increased working memory load in the sequential condition leading participants to store the symbolic components in working memory to compare later rather than their integrated magnitudes.

To summarize, we find that the component model fits the most participants across the different conditions. While the extension model explains some participants performance, it performs better in the simultaneous conditions, especially when the stimuli are grouped by comparison type. In contrast, only a few participants were best explained by the reflection model.

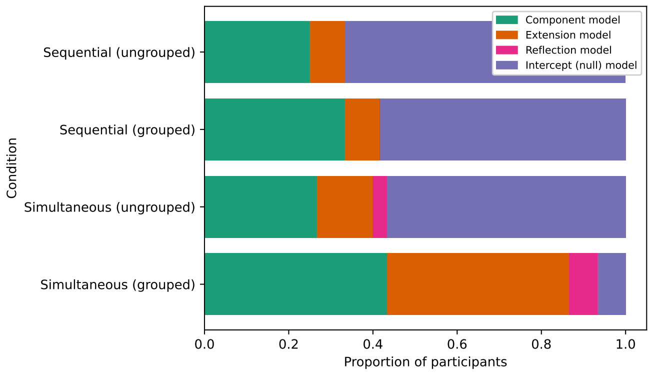

Figure 2: Proportions of selected models for individual-differences model fits for all conditions

Discussion

The present study examined how adults represent negative integers across task contexts, using the symbolic magnitude comparison task and Bayesian models to evaluate competing theoretical accounts. The results indicate that representations of negative integers are not uniform or fixed, but vary across individuals and contexts. Critically, these conclusions are not apparent from group-level reaction time analyses alone. Instead, they emerge when the hypotheses are evaluated at the individual level.

At the group level, we observed standard distance effects for positive-positive and negative-negative comparisons, alongside no reliable distance effect for mixed-sign comparisons. Similar aggregate patterns have been reported repeatedly in the literature and are often interpreted as evidence for componential processing, in which polarity rules dominate mixed-sign judgments (Ganor-Stern & Tzelgov, 2008; Ganor-Stern et al., 2010; Tzelgov et al., 2009). However, individual-level Bayesian model comparison revealed that these null effects reflect averaging across participants who recruited qualitatively different representations. Importantly, they overlook participants' recruitment of an extension representation in certain task contexts. Thus, the absence of a group-level distance effect does not necessarily imply the absence of magnitude processing, but rather the coexistence of multiple representational strategies whose behavioral signatures cancel out in aggregate (Fischer & Rottmann, 2005; Varma & Schwartz, 2011).

We present a framework called Representational Dynamics of Integer Cognition (RDIC) to help interpret this phenomenon. RDIC characterizes integer cognition as involving multiple layers of representation that coexist and are differentially recruited depending on task context. At a symbolic layer, integers can be represented componentially, with sign and magnitude treated as separable elements and combined via learned rules. At a magnitude layer, integers can be represented on an analog spatial continuum, yielding extension-based representations that preserve ordinal relations across zero. At a relational-structural layer, integers can be organized in terms of symmetry and additive inverse relations, corresponding to reflection-based representations. Importantly, RDIC does not posit that these layers replace one another over development or expertise. Instead, they remain available and are adaptively coordinated, with task affordances determining which representational layer is most strongly engaged at a given moment. Future analyses which focus on representational shifting within individuals will help shed further light on this flexibility.

Task context played a central role in shaping representational recruitment. When cognitive and strategic demands were minimized, as in the simultaneous, grouped condition, a substantial proportion of participants were best described by the extension model, indicating use of an integrated magnitude representation spanning zero (Fischer, 2003; Fischer & Rottmann, 2005). At the same time, many participants continued to rely on componential processing, consistent with evidence that rule-based strategies remain efficient and frequently used even in adults with extensive experience with integers (Ganor-Stern et al., 2010; Varma & Schwartz, 2011). Across conditions, few participants were best described by the reflection model. This contrasts with prior work that has inferred reflection-based representations primarily from inverse distance effects in mixed-sign comparisons (Varma & Schwartz, 2011). The present results suggest that reflection representations are either rare or highly context-dependent. More broadly, this finding highlights the limits of inferring representational structure from single behavioral signatures, a concern echoed in related debates about numerical representation and task-dependent strategy use (Barth & Paladino, 2011; Siegler, 2016).

A central contribution of this work is methodological. Rather than relying on aggregate behavioral effects, we implemented a model comparison approach that directly instantiated competing representational hypotheses and evaluated their explanatory power at the level of individual participants and task contexts. This approach is novel in the integer cognition literature and responds to long-standing calls to more directly link formal theory and behavioral data (Siegler, 2016; Varma & Schwartz, 2011). By including an explicit intercept model, we also avoid over-interpreting noisy data, which is particularly important in tasks known to elicit heterogeneous strategies. Critically, this approach reveals heterogeneity in representations which may explain long-standing inconsistencies in the literature as consequences of representational dynamics rather than theoretical incompatibility.

In sum, the present findings provide evidence that representations of negative integers are flexible, context-sensitive, and heterogeneous, even in adulthood. By introducing the RDIC framework and applying formal modeling, this work demonstrates that understanding cognition requires moving beyond group-level effects toward models that capture when and how different representational layers are recruited. Negative integers thus offer a powerful domain for investigating the dynamics of abstract thought.

References

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

Bofferding, L. (2014). Negative integer understanding: Characterizing first graders' mental models. *Journal for Research in Mathematics Education*, *45*, 194–245. https://doi.org/10.5951/jresematheduc.45.2.0194

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*(1), 1–12.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1-2), 1–42.

Fechner, G. T. (1860). Elements of psychophysics. *1*. https://doi.org/10.1037/11304-026

Fischer, M. H. (2003). Cognitive representation of negative numbers. *Psychology Science*, *15*, 30–31. https://doi.org/10.1111/j.1467-9639.1993.tb00256.x

Fischer, M. H., & Rottmann, J. (2005). Do negative numbers have a place on the mental number line? *Psychology Science*, *47*, 22–32.

Ganor-Stern, D., Pinhas, M., Kallai, A., & Tzelgov, J. (2010). Holistic representation of negative numbers is formed when needed for the task. *Quarterly Journal of Experimental Psychology*, *63*, 1969–1981. https://doi.org/10.1080/17470211003721667

Ganor-Stern, D., & Tzelgov, J. (2008). Negative numbers are generated in the mind. *Experimental Psychology*, *55*, 157–163. https://doi.org/10.1027/1618-3169.55.3.157

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature reviews neuroscience*, *6*(6), 435–448.

Kilhamn, C. (2011). Making sense of negative numbers, 291. https://gupea.ub.gu.se/handle/2077/24151

Krajcsi, A., & Igács, J. (2010). Processing negative numbers by transforming negatives to positive range and by sign shortcut. *European Journal of Cognitive Psychology*, *22*(7), 1021–1038. https://doi.org/10.1080/09541440903211113

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261.

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

Phan, D., Pradhan, N., & Jankowiak, M. (2019). Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. *Program Transformations for ML Workshop at NeurIPS 2019*.

Shaki, S., & Petrusic, W. M. (2005). On the mental representation of negative numbers: Context-dependent snarc effects with comparative judgments. *Psychonomic Bulletin and Review*, *12*, 931–937. https://doi.org/10.3758/BF03196788

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, *19*, 341–361. https://doi.org/10.1111/desc.12395

Tzelgov, J., Ganor-Stern, D., & Maymon-Schreiber, K. (2009). The representation of negative numbers: Exploring the effects of mode of processing and notation. *Quarterly Journal of Experimental Psychology*, *62*, 605–624. https://doi.org/10.1080/17470210802034751

Varma, S., & Schwartz, D. L. (2011). The mental representation of integers: An abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition*, *121*(3), 363–385. https://doi.org/10.1016/j.cognition.2011.08.005

Vest, N., & Alibali, M. W. (2021). The mental representation of integers: Further evidence for the negative number line as a reflection of the natural number line. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Vest, N. A. (2025). *The mental representation of negative integer magnitudes over development and across contexts*. The University of Wisconsin-Madison.

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in 'negativity' [The Conceptual Change Approach to Mathematics Learning and Teaching]. *Learning and Instruction*, *14*(5), 469–484. https://doi.org/https://doi.org/10.1016/j.learninstruc.2004.06.012