Supplemental Analyses for *Spatial Mapping of Fraction Magnitudes*

Experiment 1: Reaction Time on Magnitude Comparison Task

Method

As reported in the main text, only reaction times (RTs) from correct responses and those within three standard deviations of the individual's average RT were included in the analyses. Furthermore, adults who performed at or below chance (4/8 or below) or who had fewer than three included trials (i.e., that were correct and within three SDs of their average RT) were excluded, resulting in a final sample of 38 (out of 50) participants having complete and useable RT data. At the group level, values more than three standard deviations away from the group mean were considered outliers and were replaced with the next highest value within the acceptable range (3 SDs of the mean). This resulted in the replacement of less than 1% of the data.

Results and Discussion

When looking at RT (with only n = 38), the pattern of results was identical to that reported in the main text for accuracy. There was a main effect of notation $F(2, 74) = 13.1$, $p < 0.001$, *partial* $\eta^2 = 0.26$, with RT on FvF trials, $M = 2013$ms, taking significantly longer than both NvF, $M = 1702$ms, $t(37) = 4.9$, $p<0.001$, and DvF trials, $M = 1804$ms, $t(37) = 4.1$, $p<0.001$, which were not significantly different from each other, $t(37) = 1.5$, $p<0.15$. Again, there was also a main effect of ratio, $F(1, 37) = 19.7$, $p < 0.001$, *partial* $\eta^2 = 0.35$, with the small ratio, $M =$ 1941ms, taking significantly longer than the large ratio, $M = 1738$ms. Again, there was not a significant interaction between notation and ratio, $F(2, 74) = 1.14$, $p = 0.3$, *partial* $\eta^2 = 0.03$. These patterns replicate the pattern reported in the main text for accuracy, suggesting that the

overall pattern of ratio-dependent responding and differences in overall performance across notation are robust to the particular DV used.

Experiment 2: Adult Sample

Method

Participants. Fifty adults were randomly assigned to each condition: Number Line Condition (M_{age} = 18.6 years, Range: 17 to 21 years, 40 females), or the Pie Chart Condition (M_{age} = 19.0 years, Range: 18 to 23 years, 36 females). Adults were tested in our laboratory and received partial course credit for participation.

Procedure. The method was identical to that reported in the main text for children in Experiment 2, except that both accuracy and RT are reported for adults (using the same inclusion criteria as Experiment 1, resulting in replacing ~ 1.7% of the RT data and a final sample of n = 46 in the PC condition and n = 43 in the NL condition for the RT analyses).

Results and Discussion

Magnitude Comparison Task. We used two separate ANOVAs on proportion correct and on RT, with notation (3: FvF, DvF, and NvF) and ratio (2: small and large) as within-subject factors and condition (2: Number Line and Pie Chart) as a between subject factor.

When analyzing proportion correct, data revealed a main effect of notation $F(2, 196) = 9.2$, $p < 0.001$, *partial* $\eta^2 = 0.09$, a main effect of ratio, $F(1, 98) = 57.5$, $p < 0.001$, *partial* $\eta^2 = 0.37$, and again, a ratio by notation interaction, $F(2, 196) = 8.9$, $p < 0.001$, *partial* $\eta^2 = 0.08$. Paired t-tests indicated there were significant ratio effects in each notation type, replicating Experiment 1 and prior work (e.g., Hurst & Cordes, 2016): FvF $M_{small} = 0.83$, $M_{large} = 0.95$, $t(99) = 6.4$, $p < 0.001$, Cohen's d = 0.8; NvF $M_{small} = 0.91$, $M_{large} = 0.95$, $t(99) = 3.1$, $p = 0.003$, Cohen's d = 0.42; and DvF $M_{small} = 0.9$, $M_{large} = 0.95$, $t(99) = 4.2$, $p < 0.001$, Cohen's d = 0.46.

However, accuracy on the FvF trials ($M = 0.89$) was significantly lower overall and resulted in significantly higher ratio effects than on the NvF trials ($M = 0.93$; performance difference: $p < 0.001$, ratio effect difference: $p = 0.001$) and the DvF trials ($M = 0.92$; performance difference: $p = 0.004$, ratio effect difference: $p = 0.005$). There were no significant differences between DvF and NvF on overall performance ($p = 0.35$) or ratio effects ($p = 0.38$). For the adults, there were no main or interaction effects involving condition ($ps > 0.1$).

The same analyses involving adults' RT showed a very similar pattern: there was a main effect of notation $F(2, 174) = 19.2$, $p < 0.001$, *partial* $\eta^2 = 0.18$, a main effect of ratio, $F(1, 87) = 72.0$, $p < 0.001$, *partial* $\eta^2 = 0.5$, and a ratio by notation interaction, $F(2, 174) = 12.1$, $p < 0.001$, *partial* $\eta^2 = 0.1$. The largest ratio effects were on the DvF trials, $M_{small} = 2592$ms, $M_{large} = 2132$ms, $t(88) = 7.0$, $p < 0.001$, Cohen's d = 0.36, followed by the FvF trials, $M_{small} = 2603$ms, $M_{large} = 2367$ms, $t(88) = 3.9$, $p < 0.001$, Cohen's d = 0.24, and NvF trials, $M_{small} = 2081$ms, $M_{large} = 2016$ms, $t(88) = 1.9$, $p = 0.06$, Cohen's d = 0.1 (which were only marginally significant) and all three ratio-effects were significantly different from each other ($ps < 0.05$). Furthermore, FvF took the longest on average, followed by DvF (only marginally significantly different from FvF, $p = 0.06$), and NvF (significantly different from both, $ps < 0.001$). However, there were no main effects or interactions involving condition ($ps > 0.4$). Thus, unlike the children, practice with the number line or the pie charts did not appear to differentially impact performance on the subsequent symbolic comparison task.

Visual representation task. Adults did not show a significant difference in PAE on the Number-to-Position task (Pie Chart condition: $M = 3.0\%$; Number Line condition: $M = 3.4\%$), $t(86.7) = 0.96$, $p = 0.3$, Cohen's d = 0.19. Adults were significantly more accurate in the Number Line condition than in the Pie Chart condition on the Position-to-Number trials, $M_{NL} = 0.99$, M_{PC}

= 0.93, $t(69.5) = 4.16$, $p < 0.001$, Cohen's d = 0.83; however, it should be noted that performance in both conditions was very accurate and approaching ceiling on this task, providing little variability.

Overall, adults tended to not use overt partitioning strategies regardless of the format of the representation, however pie charts were more likely to have evidence of partitioning strategies than number lines. In the Number Line condition, none of the 50 adults ever showed evidence of written partitioning on the number lines. In contrast, in the Pie Chart condition only 29 adults consistently did not show evidence of overt partitioning, 7 adults consistently used a partitioning strategy, and 14 adults were inconsistent in their strategy use. Furthermore, when looking at just those participants using a consistent strategy, the types of strategy employed differed across conditions, $\chi^2 = 10.6$, $p = 0.001$, Fisher's Exact Test $p = 0.002$ (given the low expected values in some cells).

Overall, this is broadly consistent with the pattern shown in children, such that adults were more likely to consistently partition pie charts than number lines. However, unlike the children, adults were overall unlikely to partition, regardless of the representation.

Experiment 3: Adult Sample

Method

Participants. A smaller sample of 48 adults participated in the same procedure used for children in Experiment 3, and were assigned to either the Extended Number Line (N = 23, M_{age} = 19.4 years, Range: 18 to 22 years, 15 females) or Extended Pie Chart (N = 25, M_{age} = 19.6 years, Range: 18 to 22 years, 29 females) condition. Adults were tested in our laboratory and received partial course credit for participation.

Procedure. The procedure was identical to that used for children in Experiment 3. Performance data on the magnitude comparison task was treated the same way as Experiments 1 and 2. Adult RT analyses included a final sample of 20 in the Extended Number Line condition and 23 in the Extended Pie Chart condition and ~2.3% of the data were considered outliers and replaced. Three adults in the Extended Pie Chart condition responded to the booklet in an atypical fashion, making it impossible to score the accuracy of these participants in a way that is comparable to the others (e.g., coloring in the pie charts as you might a rectangle), and are thus excluded from analyses involving accuracy on the spatial task.

Results

Magnitude Comparison Task. As in Experiment 2, we used a ratio (2: small, large), notation (3: FvF, DvF, NvF), by condition (2: Extended NL, Extended PC) ANOVA on both proportion correct and RT separately.

The ANOVA on proportion correct revealed a significant main effect of ratio, $F(1, 46) = 21.1$, $p < 0.001$, *partial* $\eta^2 = 0.3$, with adults performing better on the larger ratio (M=0.96) than the small ratio (M=0.9). However, there were no other significant main effects or interactions, including those involving condition (all *ps* > 0.05). Given the near ceiling performance of adults on this task, we also looked at RT.

For RT, there was a main effect of notation $F(2, 82) = 22.5$, $p < 0.001$, *partial* $\eta^2 = 0.4$, a main effect of ratio, $F(1, 41) = 31.0$, $p < 0.001$, *partial* $\eta^2 = 0.4$, and a ratio by notation interaction $F(2, 82) = 5.8$, $p = 0.004$, *partial* $\eta^2 = 0.1$. Additional analyses revealed there was a significant ratio effect in each of the three notation types, with slower performance on the small ratio compared to the large ratio (paired t-tests in each notation separately: NvF: M_{Small} = 1977ms, M_{large} = 1811ms, $p < 0.001$; DvF: M_{Small} = 2591ms, M_{large} = 2070ms, $p < 0.001$; FvF:

M_{Small} = 2631ms, M_{large} = 2398ms p = 0.002). However, this ratio effect was largest for DvF (ps < 0.05) and the NvF and FvF ratio effects were not significantly different from each other (p = 0.4). In addition, there was a small and marginal ratio by notation by condition interaction, $F(2, 82) = 2.7$, $p = 0.07$, *partial* $\eta^2 = 0.06$.

Thus, in contrast to the child sample, adults did show some small evidence that the number line tasks may have impacted performance on the magnitude comparison task when comparing decimals and fractions in particular. However, this was only evident in the fact that RT ratio effects were higher in the number line, compared to the pie chart, condition. Given that ratio effects are generally taken as evidence for attending to magnitude during the task (e.g.,Moyer & Landauer, 1967, 1973), it may be that the higher ratio effects in decimal vs. fraction comparisons are reflective of the use of a more approximate, magnitude-based strategy.

Visual representation task. First, we compared adults' performance on the spatial mapping task for fractions and decimals separately, in both the number line and pie chart conditions. Adults' performance showed significantly less error when mapping to the number line (N = 23) than the pie chart (N=22) for decimals (M_{NL} = 1.1, M_{PC} = 2.7, $t(34.4) = 4.2$, $p <$ 0.001, Cohen's d = 1.3), but did not show a significant difference on fractions (M_{NL} = 3.6, M_{PC} = 5.1, $t(43) = 0.9$, $p = 0.4$).

Next, we more closely analyzed adults' use of overt partitioning strategies (see Table S1). Similarly to Experiment 2, most adults in both conditions, tended to provide the answer without an overt partitioning strategy. However, this pattern is slightly less strong for decimals, where a substantial number of adults were inconsistent in their strategy use. When looking at just those adults who consistently used a single strategy (either always partitioning or never partitioning, ignoring inconsistent adults), there was not a significant difference in strategy use across

Extended Number Line and Extended Pie Chart for decimals ($\chi^2 = 2.5$, $p = 0.11$, Fisher's Exact Test: $p = 0.20$) or for fractions (the tests cannot be calculated because the number of Partitioning is 0 for both conditions).

Table S1: Number of adults categorized in each strategy category

		NL condition		PC Condition	
		Fractions	Decimals	Fractions	Decimals
Adults	Never Partitioning	21	23	16	17
	Always Partitioning	0	0	0	2
	Inconsistent	2	0	9	6