The Emergence of Task-Relevant Representations in a Nonlinear Decision-Making Task

Menghi N.[1,2,*], Silvestrin F.[1], Pascolini L.[1] and Penny W.[1]

1 University East Anglia, School of Psychology, UK.
2 Max Planck for Human Cognitive and Brain Sciences, Department of Psychology, Germany
** Corresponding author: menghi@cbs.mpg.de*

March 2023

Supplementary Material

Summary of Behavioural Results

Feature Value	Overall Accuracy	Stats
0	$\mu = .85$, $\sigma_M = .09$	t(22) = 39.97, $p < 0.001$
1	$\mu = .60$, $\sigma_M = .13$	t(22) = 20.36, $p < 0.001$
2	$\mu = .59$, $\sigma_M = .17$	t(22) = 14.46, $p < 0.001$
3	$\mu = .67$, $\sigma_M = .17$	t(22) = 16.90, $p < 0.001$
4	$\mu = .73$, $\sigma_M = .17$	t(22) = 18.78, $p < 0.001$

Table 1: **Subtraction Task: Overall accuracy by Feature Value** The table show the overall mean accuracies, μ, the standard error of the mean, σ_M, and the results of the t-tests.

Feature Value	Good Performers	Stats
0	$\mu = .90$, $\sigma_M = .07$	t(10) = 18.21, $p < 0.001$
1	$\mu = .69$, $\sigma_M = .09$	t(10) = 6.77, $p < 0.001$
2	$\mu = .70$, $\sigma_M = .17$	t(10) = 3.62, $p = 0.004$
3	$\mu = .77$, $\sigma_M = .19$	t(10) = 4.32, $p = 0.001$
4	$\mu = .83$, $\sigma_M = .14$	t(10) = 7.69, $p < 0.001$

Table 2: **Subtraction Task: Good performers accuracy by Feature Value** The table show the overall mean accuracies, μ, the standard error of the mean, σ_M, and the results of the t-tests.

Feature Value	Bad Performers	Stats
0	μ = .81, σ_M = .09	t(11) = 10.86, p < 0.001
1	μ = .52, σ_M = .09	t(11) = 0.68, p = 0.509
2	μ = .50, σ_M = .11	t(11) = -0.05, p = 0.958
3	μ = .59, σ_M = .09	t(11) = 3.46, p = 0.005
4	μ = .63, σ_M = .13	t(11) = 3.29, p = 0.007

Table 3: **Subtraction Task: Bad performers accuracy by Feature Value** The table show the overall mean accuracies, μ, the standard error of the mean, σ_M, and the results of the t-tests.

Feature Value	Overall Accuracy	Stats
0	μ = .64, σ_M = .19	t(22) = 3.38, p = 0.002
1	μ = .55, σ_M = .13	t(22) = 1.86, p = 0.075
2	μ = .61, σ_M = .13	t(22) = 3.97, p < 0.001
3	μ = .70, σ_M = .17	t(22) = 5.53, p < 0.001
4	μ = .77, σ_M = .13	t(22) = 9.42, p < 0.001

Table 4: **Addition Task: Overall accuracy by Feature Value** The table show the overall mean accuracies, μ, the standard error of the mean, σ_M, and the results of the t-tests.

Feature Value	Good Performers	Stats
0	μ = .78, σ_M = .18	t(10) = 4.96, p < 0.001
1	μ = .63, σ_M = .14	t(10) = 3.06, p = 0.012
2	μ = .71, σ_M = .11	t(10) = 5.90, p < 0.001
3	μ = .85, σ_M = .09	t(10) = 12.11, p < 0.001
4	μ = .85, σ_M = .08	t(10) = 12.96, p < 0.001

Table 5: **Addition Task: Good performers accuracy by Feature Value** The table show the overall mean accuracies, μ, the standard error of the mean, σ_M, and the results of the t-tests.

Feature Value	Bad Performers	Stats
0	μ = .50, σ_M = .06	t(11) = 0.32, p = 0.749
1	μ = .48, σ_M = .07	t(11) = -0.96, p = 0.356
2	μ = .52, σ_M = .06	t(11) = 1.07, p = 0.305
3	μ = .57, σ_M = .11	t(11) = 2.14, p = 0.053
4	μ = .69, σ_M = .12	t(11) = 5.21, p < 0.001

Table 6: **Addition Task: Bad performers accuracy by Feature Value** The table show the overall mean accuracies, μ, the standard error of the mean, σ_M, and the results of the t-tests.

Learning Curve

In this supplementary section, we provide additional insights into the learning accuracy across trials. Here we present a description of learning accuracy and reaction time across different trials. This figure shows accuracy over trials computed as a moving average over 10 trials.

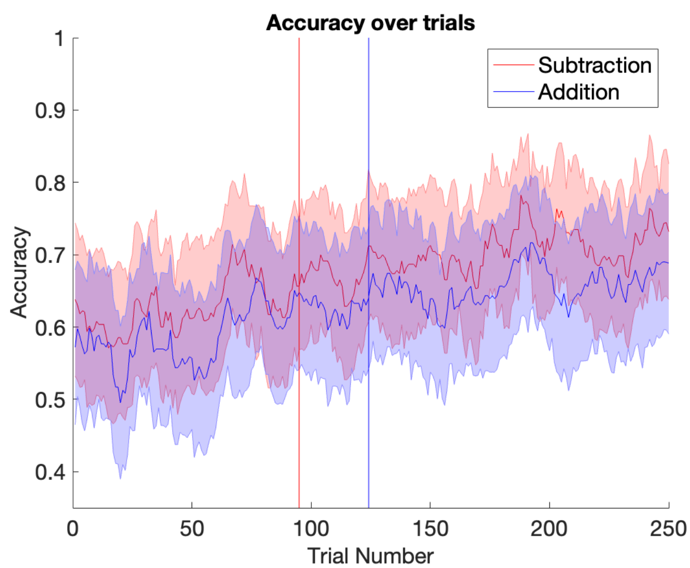

Figure 1: **Accuracy Over trials in Subtraction and Addition tasks.** *Learning accuracy is computed with a moving average with a window of 10 trials. The error bars in the figure correspond to the standard error of the mean. Additionally, a line on the graph indicates the average point in time at which participants declared their respective strategies.*

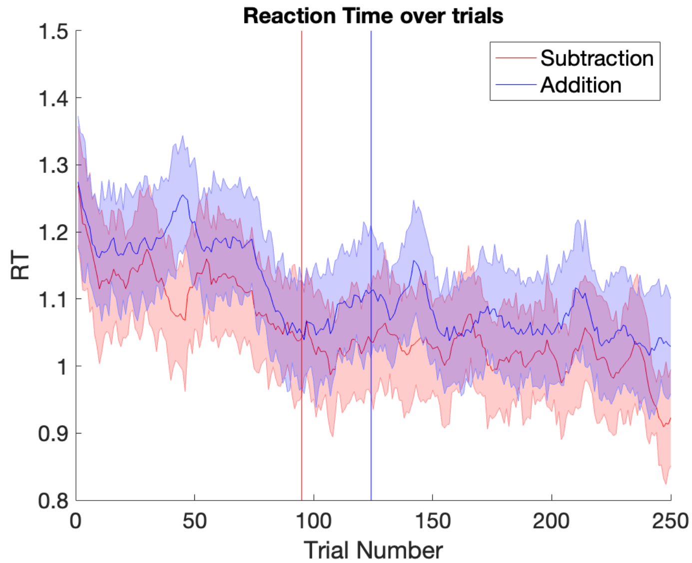

Figure 2: **Accuracy Over trials in Subtraction and Addition tasks.** *Learning accuracy is computed with a moving average with a window of 10 trials. The error bars in the figure correspond to the standard error of the mean. Additionally, a line on the graph indicates the average point in time at which participants declared their respective strategies.*

Good and bad performers

Here we divide participants based on their performance, providing additional insights on the learning pattern of these two groups.

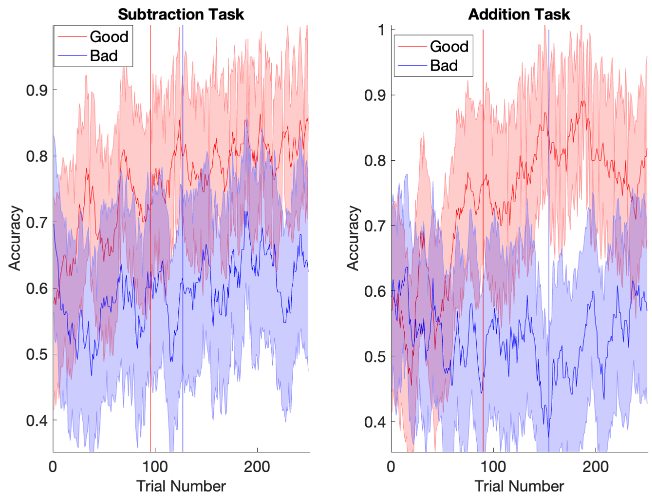

Figure 3: **Accuracy Over trial for Good and Bad performers in Subtraction and Addition tasks.** *Learning accuracy is computed with a moving average with a window of 10 trials. The error bars in the figure correspond to the standard error of the mean. Additionally, a line on the graph indicates the average point in time at which participants declared their respective strategies.*

Declarative Knowledge

Here we divide participants into declarative and non-declarative based on their correct declaration of the task rule, providing additional insights on the learning pattern of these two groups.

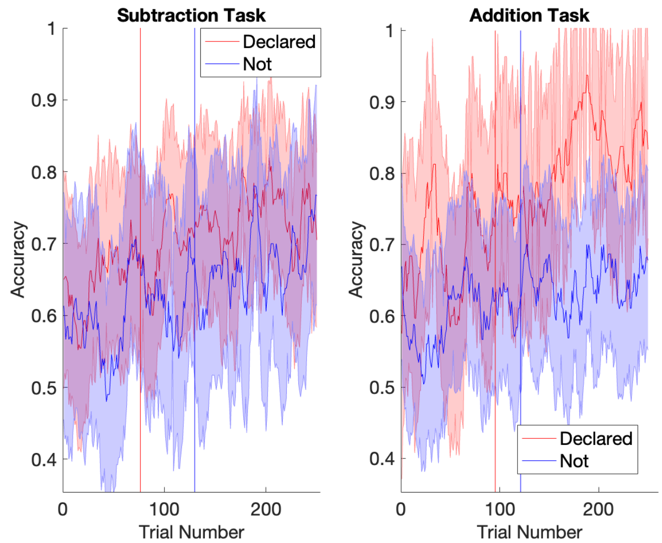

Figure 4: **Accuracy Over trials for Declarative and Non-declarative participants in Subtraction and Addition tasks.** *Learning accuracy is computed with a moving average with a window of 10 trials. The error bars in the figure correspond to the standard error of the mean. Additionally, a line on the graph indicates the average point in time at which participants declared their respective strategies.*

Multivariate EEG Results

Decoding of configurations

Figure **??** shows the decoding accuracy averaged across all 25 stimulus configurations. Before and just after stimulus presentation, grand average decoding accuracy fluctuated around the chance level. In the subtraction task, classification reached significance at 190 ms (190–320 ms), followed by a cluster at 340 ms (340–490 ms) and the last cluster at 510 ms (510-750 ms). In the addition task classification reached significance at 210 ms (210–340 ms), followed by a cluster at 480 ms (480–570 ms), a cluster at 1150 ms (1150-1250 ms) and a last one at 1270 ms (1270-1440 ms). Thus, multi-variate analysis of EEG data revealed the temporal dynamics of the visual processing of the different configurations in the brain.

Multidimensional Scaling

Because it is difficult to directly make sense of the $25 \times 25 \times 161$ EEG decoding matrix, we used multidimensional scaling (MDS) to project the data into a two-dimensional space of the first two dimensions of the solution, such that similar representation are grouped together and dissimilar ones far apart. MDS is a method to visualize the level of similarity of individual objects contained in a distance matrix (here the decoding matrix), whereby objects are automatically assigned coordinates in space so that distances between objects are preserved. For the purpose of MDS we averaged the EEG decoding matrix over those time points shown to be significant using the non-parametric permutation tests.

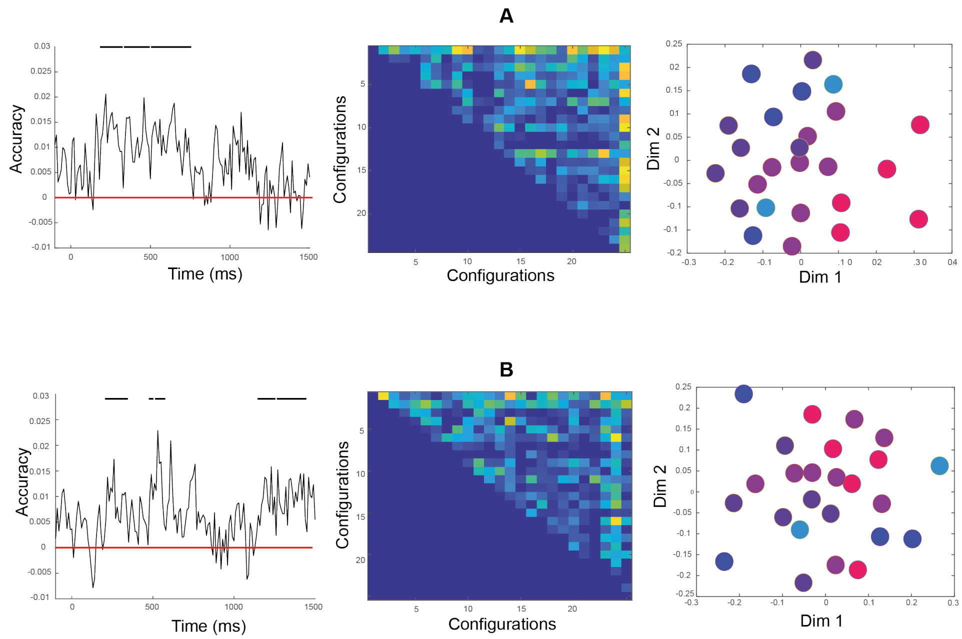

Figure 5: **Timecourse of decoding accuracy among configurations, the structure of decoding matrices and MDS spaces in (A) Subtraction and (B) Addition tasks.** *The left panel illustrates the time course of overall decoding. The horizontal bars above represent the significant clusters. The* $[i, j]$*th entry in the EEG Decoding Matrices (central panels) correspond to the cross-validated accuracies with which stimulus configuration* i *and can be discriminated from configuration* j *(with yellow denoting highest accuracy). These accuracies have been averaged over time points containing significant effects (see left panels). The right panel illustrates the first two dimensions of the MDS in the EEG decoding matrix, according to the feature value (see Fig 2).*