A Hidden semi-Markov model classifier for strategy detection in multiplication problem solving

Ernö Groeneweg (ernogroeneweg@gmail.com)
Utrecht University

Kim Archambeau (k.c.archambeau@uva.nl)
University of Amsterdam & Université Libre de Bruxelles

Leendert Van Maanen (l.vanmaanen@uu.nl)
Utrecht University

Abstract

Self-report as a tool to understand different cognitive processing strategies has been criticised for decades, but to date there have not been many alternatives. To remedy this hiatus, we propose to apply a recently developed method for processing stage analysis (Hidden semi-Markov Model Multivariate Pattern Analysis, HsMM-MVPA) to a cognitive strategy prediction task. HsMM-MVPA uses specific patterns in EEG data to determine the most likely number of sequential processing stages. Under the assumption that cognitive processing strategies differ in the number of stages, we constructed a classifier using fitted HsMM-MVPA to try and differentiate between two cognitive strategies in unseen data. The method is applied to data from a multiplication verification task, in which participants are asked to verify the truth of a solution to a multiplication problem (3×9). We asked participants to indicate via self-report whether they knew the answer by heart (Strategy 1, Retrieval) or needed to compute the answer (Strategy 2, Procedural). The classifier could predict the self report labels above chance, suggesting that the number of processing stages identified using EEG can be used to track the cognitive processing strategy that are in use throughout a task.

Keywords: cognitive strategies; cognitive processing stages; classification; HsMM-MVPA; EEG

Introduction

When different people apply different cognitive strategies to solve the same problem, a question arises: if people use different strategies, how do these strategies differ? It seems that we could differentiate between cognitive strategies by way of the underlying processing stages. A robust method for decomposing a trial into processing stages should then be able to do a differentiation by strategy. To test this hypothesis, we constructed an EEG classifier based on a novel modelling method with which we will try to predict what strategy is used in unseen EEG data.

Processing Stage Decomposition

A recently proposed framework for detecting cognitive processing stages by Anderson et al. (2016), called Hidden semi-Markov Model multivariate pattern analysis (HsMM-MVPA), has been suggested to give nuanced insight into processing stages that might be present in EEG, MEG, or fMRI data. Within this framework, processing stages are modelled as a hidden Markov chain using distributed peaks in activity. By additionally modelling the cognitive processing stages as a semi-Markov chain, we can get insight into the temporal on- and offset as well as the durations of processing stages. Such nuanced insight into the characteristics of processing stages allows for differentiation between strategies, under the assumption that these characteristics differ between strategies. The HsMM-MVPA framework has been shown to be a versatile method for detecting processing stages in a variety of conditions and tasks (Anderson et al., 2018; Berberyan et al., 2021; Borst & Anderson, 2015; Portoles et al., 2018; van Maanen et al., 2021; Walsh et al., 2017; Zhang, van Vugt, et al., 2018; Zhang, Walsh, & Anderson, 2017, 2018; Zhang, Borst, et al., 2017).

In many implementations, HsMM-MVPA has fitted models that explain EEG data from multiple participants very well. This suggests that there could be some commonality between participants in how the HsMM-MVPA method represents these processing stages. Therefore, it stands to reason that there is overlap in cognitive processing stages between participants employing the same cognitive strategy for the same task. In the current paper, we aim to understand whether an HsMM-MVPA model can be used to distinguish between cognitive strategies in EEG data from unseen participants. Concretely, we collected EEG data from people performing a multiplication task, while we also collected self-reports of the strategies that people use during the task. Then, we estimated the optimal HsMM-MVPA model for the self-reported strategies. We hypothesise that a classifier based on these HsMM-MVPA models predicts which strategy unseen participants used on a particular multiplication problem. If this prediction is above chance, this will support the hypothesis that processing stages can be used to differentiate between strategies.

Hidden semi-Markov Model Multivariate Pattern Analysis Standard Hidden Markov Models consist of two stochastic finite-time chains. One is a hidden Markov chain X and the other is an observable chain Y whose behaviour depends on X. For every pair (x,y) where $x \in X$, $y \in Y$ there is a probability that x happens when y is observed (Visser et al., 2009). In a HMM, the duration of a state corresponds to the duration of a single observation. In contrast, in a hidden *semi*-Markov model, it is possible to have multiple observations per hidden state, which allows for variable state durations (Yu, 2010). Since processing stages are not assumed to have the same duration, HsMMs are best suited for this analysis (Anderson et al., 2016). In this study, extracted components from

EEG data are the observations Y, with the underlying processing stages being modelled by the most likely sequence of hidden states X.

To discover different processing stages, HsMM-MVPA relies on the assumption that a processing stage onset is signified by a cognitive event that can be discovered in the EEG signal by looking for positive or negative peaks, distributed across different brain regions. This assumption is shared by two main theories explaining the generation of event-related potential (ERP); the classical theory (Shah et al., 2004) and the synchronised oscillations theory (Makeig et al., 2002). These theories agree that a cognitive event is signified by a positive or negative peak in the EEG signal, although they disagree on the exact origin of this peak (see Anderson et al. (2016) for a more extensive discussion).

HsMM-MVPA searches for positive or negative peaks in the EEG signal (called *bumps*), with the subsequent *flats* denoting a processing stage. The HsMM-MVPA model consists of a number of bumps, as well as a set of gamma distributions of stage durations across trials. The algorithm first attempts to find bumps that represent the onset of a cognitive stage and the flats that separate these bumps. The goal of HsMM-MVPA is to identify the topography and temporal location of each bump on each trial. The method allows for variability within the duration of cognitive processes for each trial, so bumps can occur at different time points per trial. The trials are analysed individually, but all trials of all participants are taken into account simultaneously. A model is then fitted to various participants simultaneously.

Classification

A challenge with classifying EEG data is that there is a high degree of variability from participant to participant (Saha & Baumert, 2020). Implementations of HsMM-MVPA as described above are able to discover good-fitting models across multiple participants, as long as the participants' EEG signal is collected under the same conditions. This suggests that the number of processing stages is equal across participants using the same strategy, yielding the hypothesis that the EEG signal of different participants performing a task under the same condition can be modelled with the same HsMM-MVPA model. These models will be used to distinguish between the same strategies in unseen EEG data, making classification possible.

We will fit models to EEG data collected from participants verifying single-digit multiplication problems, differentiating between two self-reported strategies: *retrieval* for memory retrieval and *procedural* for procedural strategies. Using these models, we will attempt to classify these same strategies in unseen participants. We determine the sensitivity and specificity of the models to self-reported trials using an receiver operating characteristic (ROC) curve based on the likelihood of the Retrieval strategy. If the area under that curve is larger than 0.5, the model is correctly identifying self-reported Retrieval-trials higher than chance.

We chose to focus our analysis on the retrieval strategy, since previous work suggested that what people interpret as a simple memory retrieval is much more homogenous (Archambeau et al., unpublished). In fact, there are many different solution strategies that all can be described as a procedural strategy (LeFevre et al., 1996; Ashcraft, 1992). For example, to calculate that 6×4 is equal to 24, one can retrieve from memory the related fact that $6 \times 5 = 30$, and then subtract 6. An alternative procedural strategy involves consecutively adding 6 while simultaneously counting the number of additions. When this number reaches 4, you have arrived at the answer.

Method

We collected EEG, accuracy, and response times from individuals verifying single-digit multiplication problems (Archambeau et al., 2019). Participants were shown a single-digit multiplication problem with an answer and asked to verify whether the given answer was correct. Next, they were asked to self-report whether they knew the answer from memory (i.e., the retrieval strategy) or computed the answer another way (Procedural).

Design & Procedure

Forty-two undergraduate students from the Université Libre de Bruxelles (ULB) between the ages of 17 and 52 ($m = 22.24$) took part in the multiplication experiment. The study was approved by the local Ethical Review Board of the ULB, Faculty of Psychological and Educational Sciences. All participants provided informed consent and received course credit for their participation. Each trial started with the presentation of a fixation point of 500 ms. A multiplication problem containing two operands in Arabic format and the multiplication operator "x" (e.g., 6×4) was displayed in the centre of the screen for 200 ms, followed by a blank screen of 120 ms. Then, an answer was shown (24) until a response was provided. The participants were asked to indicate via button press whether the proposed solution of the problem was correct or not. Participants were asked to be as fast and as accurate as possible. When a response is given, a 300 ms interval occurred, after which participants were prompted to report what strategy they used to verify the multiplication problem; "memory" or "calculation strategy". Then, the next trial was initiated with an inter-trial interval of 1000ms. The task consisted of 4 blocks of 248 trials, for a total of 992 trials. There were three trial types. Besides trials where the given solution was correct (positive or P), there were trials two trial types where the given solution was incorrect: interfering solutions (I) and non-interfering solutions (NI). With an interfering solution, the answer given is table-related to one of the operands. The given interfering solution of a problem $a \times b$ could be the correct solution of $(a \pm 1) \times b$ or $a \times (b \pm 1)$. Multiplication problems with a single digit as the correct solution were removed, as well as 9×9. Half of both I and NI were smaller and half were larger than the correct solution. Although the I and NI split is not relevant for the current study, the impact interference has on ERP is accounted for in this analysis. There were 496 positive trials and 496 negative trials with 248 interfering solution trials and 248 unrelated solution trials. Each multiplication problem was repeated 32 times. On 16 of these repetitions the provided solution was correct, on 8 it was an I solution, and on 8 it was an NI solution. The multiplication problems were presented in a pseudo-randomised order, ensuring that successive problems never had the same operands. The multiplication task was run on a 17-inch laptop computer, using the Psychophysics Toolbox extension (Brainard, 1997) in MATLAB (version R2013a, The Mathworks Inc., Natick, Massachusetts, USA). EEG data was collected using a BioSemi interface with 72 channels, at a sampling rate of 2048 Hz. EMG activity was also recorded for other purposes, beyond the scope of the current study.

Behavioural Analysis

Four data subsets were created. Three of these based on the three experimental conditions positive (P), related (I), and unrelated (NI). For the fourth subset, we collapsed all data under the assumption that the two self-report strategies would be shared across the three experimental conditions, increasing the sample size. Participants 32-42 were split off as a test set, making participants 1-31 the training set. For this analysis, incorrect responses were removed (6.5% of trials). Additionally, we removed response time outliers from all subsets of the data to remove any trials where the participant may have been confused or distracted. When matching the behavioural data to the EEG data for epoching, four participants from the training set were removed due to incorrect event numberings in the EEG data. In total, 15% of the data was removed in this step.

After cleaning our data sets, class imbalance was computed. As seen in Table 1, in all subsets the majority of trials is labelled as 'retrieval'. These numbers will be considered as a baseline for classification accuracy. Although the standard deviation in RT in the test data subsets is much lower (potentially due to some participants having generally slower RTs in the training set), HsMM-MVPA can account for variations in the temporal offset and stage durations (Anderson et al., 2016).

Condition	% Retrieval	Mean RTs (ms)	SD RTs (ms)
Training data			
All	87.1%	1398	2211
P	88.5%	1312	2051
I	83.4%	1628	2101
NI	87.9%	1339	2581
Test data			
All	88.4%	1110	874
P	91.9%	1040	819
I	81.3%	1280	998
NI	84.0%	1099	837

Table 1: Overview of trials labelled 'retrieval' after error and outlier removal. Standard deviation is computed within every subset (Positive (P), Interfering (I), and non-interfering (NI), across participants.

Data Preprocessing

The data was processed in MATLAB (version 2020a, The Mathworks Inc., Natick, Massachusetts, USA) using the open source EEGLab plugin (Delorme & Makeig, 2004). First, the data was re-referenced to all-electrode average and high-pass filtered at 1 Hz and low-pass filtered at 40 Hz, as oscillations outside of this range are not commonly associated with brain activity (Henry, 2006). The data was resampled to 512 Hz and flatlines and overly noisy sections of data were removed automatically using built-in EEGlab functions, before applying Independent Component Analysis (ICA) using the FastICA algorithm (Hyvarinen, 1999). Next, the ICLabel plugin was used to automatically flag non-brain related components from the data (Pion-Tonachini et al., 2019). These flagged components were then removed from the data. In total about 10% of the data was removed in this step.

Fitting HsMM-MVPA

To fit the HsMM-MVPA models, the data were resampled to 100 Hz and then epoched between stimulus onset and the participants' response. Spatial principal component analysis (PCA) was applied to all datasets to extract the 10 principal components from the data channels. In all subsets, the 10 principal components account for more than 97% of variance in the data.

We consider our bumps to have a duration of 50 ms, as this duration produces robust results even if the actual durations are slightly longer or shorter (Anderson et al., 2016). The duration of the subsequent flats was modeled with a gamma distribution with a shape parameter of 2. The results are not sensitive to the exact choice of shape parameter, except that it simplifies the estimation of flat distributions (Anderson et al., 2016). In a model, n bumps results in $n+1$ flats (or $n+1$ processing stages), since the first stage starts with a flat when the stimulus is applied.

We constructed different HsMM-MVPA models for every subset of the data and for both self-reported strategies, resulting in eight different models. Model estimation begins with a 1-bump model and creates models for an increasing number of bumps until a number of bumps n_{max} is reached, with n_{max} being the maximum number of 50 ms bumps that fit in the duration of the shortest epoch. During estimation, two parameters of each hidden state are obtained: (1) the amplitudes of the bumps that mark the onsets of the processing stages and (2) the scale parameter of a gamma distribution describing the stage durations. Data from all trials and all participants in a training set were taken into account simultaneously. The match between the EEG data and the model was maximised using a standard expectation-maximisation (EM) algorithm (Moon, 1996).

The fitting process begins by defining initial amplitudes for both the bumps and the gamma distributions for stage durations. Since the convergence of the EM algorithm can be sensitive to the choice of starting point, ending up in a local maximum (Wu, 1983), we used a process based on work by Zhang, Walsh, & Anderson (2018). Per subset, we first fit separate HsMM-MVPA models for each condition on n_{max} bumps, obtaining bump amplitudes and gamma distributions. Next, we used those parameters for models with $n_{max}-1$ bumps, iteratively leaving out each of the bumps in n_{max}, selecting the model with best fit. These bumps become the new n_{max} before the above process is repeated until only a one-bump model n_1 is left. This way, we can find all potential bump topologies while avoiding local maxima.

We used a leave-one-out cross-validation (LOOCV) procedure to prevent overfitting. For every training subset, we estimate an HsMM-MVPA model on all participants minus one and then test the fit of this model on the omitted participant. This process is repeated for all participants. Finally, we used a sign test to test for how many participants the log-likelihoods of the models with $n+1$ bumps increased significantly compared to an n-bump model. If a model with one additional bump outperforms the previous model for a sufficiently large number of participants, we can say that the additional complexity of that model is warranted. This step is crucial for fitting models that generalise well across participants (Anderson & Fincham, 2014). To select the best models, sign tests were used on every n-bump model to see for how many participants its fit improved compared to the $n-1$ bump model. The best model is one that improves significantly for more than half of the training participants. For a more detailed mathematical description and code for HsMM-MVPA we refer to Anderson et al. (2016) and Berberyan et al. (2021).

Classification

After estimating the most likely parameters for all models, we used our preprocessed test data to estimate the likelihood of every trial per subset under the models. We also estimated the likelihoods of all test trials per subset under models of different subsets to further test how well the models generalise. As this is a binary classification task, we compute true positive rates (TPR) and false positive rates (FPR) to plot ROC curves. Then, we determine the area under the curve (AUC), where 0.5 denotes the model classifying according to chance. We also report classification accuracy (the proportion of trials classified as their corresponding self-reports) and F1-values (computed as $F_1 = \frac{TPR}{TPR+(FPR+FNR)/2}$) for all test trials classified under all models.

Results

Model Selection

As can be seen in Figure 1, Retrieval models are consistent. In all four subsets, the models fit to the Retrieval strategy show

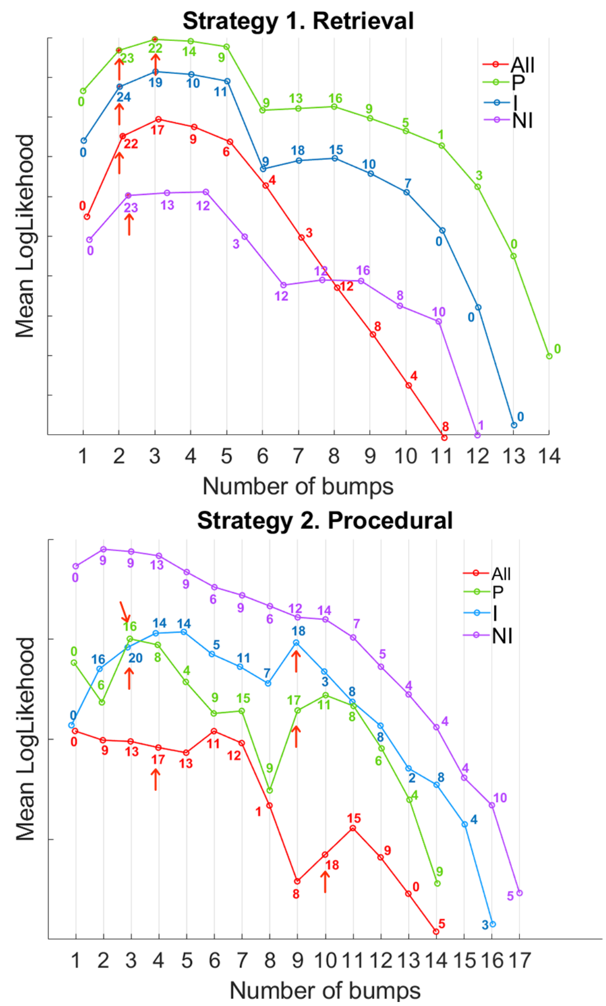

Figure 1: Model selection curves for both strategies. The numbers beside the points denote for how many out of 27 participants the log-likelihood of an n-bump model significantly increases over $n-1$ bumps. The arrows highlight the optimal models according to a sign-test for each subset of the data (see the section Fitting HsMM-MVPA for details).

that the 2-bumps model improves significantly over the 1-bump model in more than half of participants. As this result is consistent across all four subsets, the assumption seems justified that the Retrieval-strategy is generally well described by a 2-bump, 3-stage Hidden semi-Markov model. As there is no significant improvement going to a 3-bump, 4-stage model, 2-bump 3-stage models were selected. In the Procedural strategy, results are far less consistent. As a different number of bumps seems to perform best in all four subsets, we classify using the Retrieval-model only.

Classification Results

After estimating log-likelihood of every trial in the test set under all four Retrieval-models, we constructed Receiver Operating Characteristic-curves with an increasing discrimination

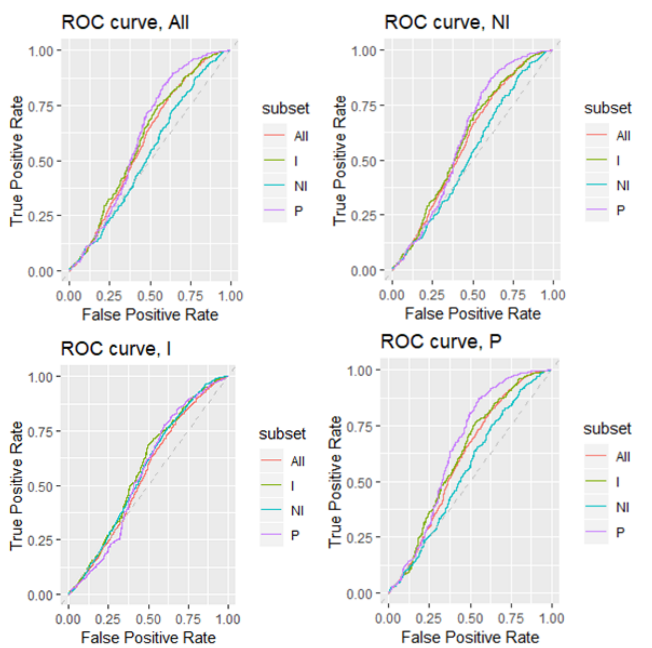

Figure 2: ROC curves for all test sets predictions under all models

threshold (Figure 2. Accuracies and F1 values can be seen in Table 2. The highest accuracy and F1 score per subset were selected. We see that all classifiers perform better than random (curve above the grey dashed line, AUC > 0.5), with P-trials in the test set performing best overall, averaging at 0.61. AUC's were computed per test participant for significance testing, showing that under all models All and P-trials are significantly higher than chance. We also see that F1-scores tend to be higher than accuracies, denoting that the classifiers are better at correctly identifying Retrieval-trials as Retrieval than they are at identifying non-Retrieval trials as non-Retrieval.

Discussion

The goal of the current study was to discover whether cognitive strategies can be differentiated between based on the number of processing stages. In general we can say with confidence that HsMM-MVPA when applied to EEG data can do so in a way that finds processing stages across participants. In other words, an HsMM-MVPA model fit to a specific cognitive strategy can recognise most unseen trials of that same strategy, even when that unseen trial is from a participant whose EEG data the model has not seen at all. This implies that, when different people report using the same cognitive strategy under the same experimental condition, their EEG patterns and consequently the processing stages are similar as well. There is also consistency between the models fit to the four experimental conditions with respect to their classification on the test data subsets, but almost everywhere this consistency is proportional to the variation in the class imbalance of our data. F1 scores everywhere tend to be higher than classification accuracy, which is closer to random performance. This means that a retrieval-based classifier is accurate at identifying trials that were self-reported as retrieval, but less accurate with respect to trials that were self-reported as procedural.

Subset	AUC	Max F1	Max acc
Model trained on All			
All	0.584‡	0.943	89.3%
P	0.612‡	0.974	94.9%
I	0.597†	0.902	82.4%
NI	0.527	0.916	84.5%
Model trained on P			
All	0.603‡	0.944	89.4%
P	0.653‡	0.974	95.0%
I	0.618	0.902	82.6%
NI	0.550	0.917	84.7%
Model trained on I			
All	0.548‡	0.942	89.1%
P	0.557‡	0.974	94.9%
I	0.577†	0.902	82.2%
NI	0.568	0.943	89.3%
Model trained on NI			
All	0.586†	0.943	89.3%
P	0.610‡	0.974	95.0%
I	0.598†	0.902	82.4%
NI	0.529	0.916	84.7%

Table 2: Area Under Curve (AUC) of the ROC curves, as well as the highest F1 scores and accuracies for every test set.
†: significant below p=0.05
‡: significant below p=0.01

In our data, the retrieval strategy is the more consistent one. All 2-bumps, 3-stage models were well fitting on all four data subsets. This means that it is highly likely that participants used 3 processing stages when using the retrieval strategy. In contrast, the procedural-strategy seems to be much less cohesive. A first possible explanation is that participants can calculate the answer to a problem in different ways, which could lead to a different number of processing stages. This means that ”procedural” encompasses a number of strategies (LeFevre et al., 1996; Ashcraft, 1992). Some of these strategies might be closer in number of stages to our 3 memory retrieval stages, which could partially account for the number of false positives our retrieval-classifier finds. For instance, a procedural strategy to verify 6×7 could involve memory retrieval of 6×6 as part of the strategy, leading this hypothetical trial to fit well with our retrieval models.

To complicate matters further, there is the potential of noise or biases to be introduced when using self-reports as a tool for setting our ground-truth (Kirk & Ashcraft, 2001). This means that wrongly labeled trials might be present. Alternative ways of ascertaining cognitive strategies, like a mixture modelling approach could be considered for this purpose (Archambeau et al., unpublished; Thevenot et al., 2007; van Maanen et al., 2014, 2016). There is also a discrepancy in the means and standard deviations of our response times between our training and test sets. HsMM-MVPA is able to account for variance in temporal onset and duration of processing stages (Anderson et al., 2016), so in theory this is no problem. However, this is a variable that could be corrected for in future.

Future work might encompass cross-validating this analysis with different train-test data splits. In addition, a comparison between scalp topographies of the best models and both correctly-labeled and mislabeled test set trials could give insight into trials that might have been incorrectly self-reported. Another avenue would be to apply this analysis to a different experimental task, such as a division task instead of multiplication. Finally, further insight into using model-based classifiers instead of more traditional machine learning approaches could improve the explainability of cognitive data classifiers. HsMM-MVPA provides a nuanced understanding of processing stages that other machine learning methods often do not.

In conclusion, this first investigation into using HsMM-MVPA as a tool for classification of cognitive strategies shows promise. The next step would be to investigate how far this promise can lead.

References

Anderson, J. R., Borst, J. P., Fincham, J. M., Ghuman, A. S., Tenison, C., & Zhang, Q. (2018). The common time course of memory processes revealed. *Psychological science*, *29*(9), 1463–1474.

Anderson, J. R., & Fincham, J. M. (2014). Extending problem-solving procedures through reflection. *Cognitive psychology*, *74*, 1–34.

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of sternberg's method. *Psychological review*, *123*(5), 481.

Archambeau, K., De Visscher, A., Noël, M.-P., & Gevers, W. (2019). Impact of ageing on problem size and proactive interference in arithmetic facts solving. *Quarterly Journal of Experimental Psychology*, *72*(3), 446–456.

Archambeau, K., Molenaar, D., Forstmann, B., Noël, M.-P., Gevers, W., & Van Maanen, L. (unpublished). Age-related differences in the resolution of interference in simple arithmetic depends on the strategy used.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1-2), 75–106.

Berberyan, H. S., van Maanen, L., van Rijn, H., & Borst, J. (2021). Eeg-based identification of evidence accumulation stages in decision-making. *Journal of Cognitive Neuroscience*, *33*(3), 510–527.

Borst, J. P., & Anderson, J. R. (2015). The discovery of processing stages: Analyzing eeg data with hidden semi-markov models. *NeuroImage*, *108*, 60–73.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433–436.

Delorme, A., & Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9–21.

Henry, J. C. (2006). Electroencephalography: basic principles, clinical applications, and related fields. *Neurology*, *67*(11), 2092–2092.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, *10*(3), 626–634.

Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 157.

LeFevre, J.-A., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, *125*(3), 284.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science*, *295*(5555), 690–694.

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, *13*(6), 47–60.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197.

Portoles, O., Borst, J. P., & van Vugt, M. K. (2018). Characterizing synchrony patterns across cognitive task stages of associative recognition memory. *European Journal of Neuroscience*, *48*(8), 2759–2769.

Saha, S., & Baumert, M. (2020). Intra- and inter-subject variability in eeg-based sensorimotor brain computer interface: A review. *Frontiers in Computational Neuroscience*, *13*, 87. Retrieved from `https://www.frontiersin.org/article/10.3389/fncom.2019.00087` doi: 10.3389/fncom.2019.00087

Shah, A. S., Bressler, S. L., Knuth, K. H., Ding, M., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral cortex*, *14*(5), 476–483.

Thevenot, C., Fanget, M., & Fayol, M. (2007). Retrieval or nonretrieval strategies in mental arithmetic? an operand recognition paradigm. *Memory & Cognition*, *35*(6), 1344–1352.

van Maanen, L., Couto, J., & Lebreton, M. (2016). Three boundary conditions for computing the fixed-point property in binary mixture data. *Plos one*, *11*(11), e0167377.

van Maanen, L., de Jong, R., & van Rijn, H. (2014). How to assess the existence of competing strategies in cognitive tasks: a primer on the fixed-point property. *PloS one*, *9*(8), e106113.

van Maanen, L., Portoles, O., & Borst, J. P. (2021). The discovery and interpretation of evidence accumulation stages. *Computational brain & behavior*.

Visser, I., Raijmakers, M. E., & van der Maas, H. L. (2009). Hidden markov models for individual time series. In *Dynamic process methodology in the social and developmental sciences* (pp. 269–289). Springer.

Walsh, M. M., Gunzelmann, G., & Anderson, J. R. (2017). Relationship of p3b single-trial latencies and response times in one, two, and three-stimulus oddball tasks. *Biological psychology*, *123*, 47–61.

Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, 95–103.

Yu, S.-Z. (2010). Hidden semi-markov models. *Artificial intelligence*, *174*(2), 215–243.

Zhang, Q., Borst, J. P., Kass, R. E., & Anderson, J. R. (2017). *Inter-subject alignment of meg datasets in a common representational space* (Tech. Rep.). Wiley Online Library.

Zhang, Q., van Vugt, M., Borst, J. P., & Anderson, J. R. (2018). Mapping working memory retrieval in space and in time: A combined electroencephalography and electrocorticography approach. *NeuroImage*, *174*, 472–484.

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2017). The effects of probe similarity on retrieval and comparison processes in associative recognition. *Journal of Cognitive Neuroscience*, *29*(2), 352–367.

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2018). The impact of inserting an additional mental process. *Computational Brain & Behavior*, *1*(1), 22–35.