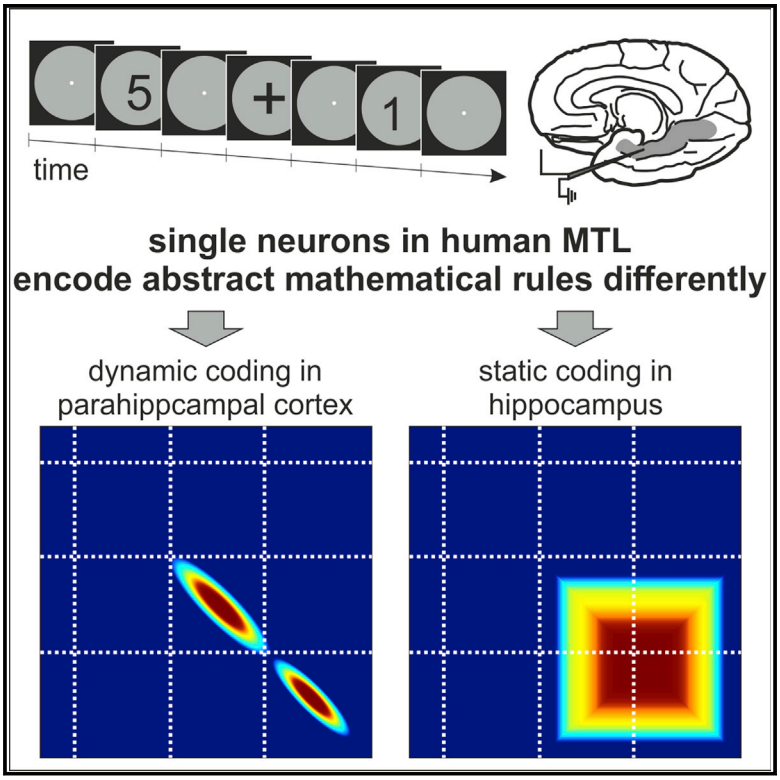
Article

Current Biology

Neuronal codes for arithmetic rule processing in the human brain

Graphical abstract



Authors

Esther F. Kutter, Jan Boström, Christian E. Elger, Andreas Nieder, Florian Mormann

Correspondence

andreas.nieder@uni-tuebingen.de (A.N.), florian.mormann@ukbonn.de (F.M.)

In brief

Kutter et al. demonstrate abstract and notation-independent codes for addition and subtraction in neuronal populations in the human medial temporal lobe (MTL). A dynamic code in the parahippocampal cortex contrasts with a static code in the hippocampus, suggesting different cognitive functions for these MTL regions in arithmetic.

Highlights

- Single neurons in the human MTL show abstract codes for addition and subtraction
- Time-resolved decoding analysis shows a dynamic code in the parahippocampal cortex
- The hippocampus shows a static code based on persistently rule-selective neurons
- Different codes suggest different cognitive functions of MTL regions in arithmetic

Kutter et al., 2022, Current Biology *32*, 1275–1284
March 28, 2022
https://doi.org/10.1016/j.cub.2022.01.054

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Neuronal codes for arithmetic rule processing in the human brain

Esther F. Kutter,[1,2] Jan Boström,[3] Christian E. Elger,[1] Andreas Nieder,[2,4,*] and Florian Mormann[1,4,5,6,*]
[1]Department of Epileptology, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany
[2]Animal Physiology, Institute of Neurobiology, University of Tübingen, 72076 Tübingen, Germany
[3]Department of Neurosurgery, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany
[4]These authors contributed equally
[5]Twitter: @humansingleunit
[6]Lead contact
*Correspondence: andreas.nieder@uni-tuebingen.de (A.N.), florian.mormann@ukbonn.de (F.M.)
https://doi.org/10.1016/j.cub.2022.01.054

SUMMARY

Arithmetic is a cornerstone of scientifically and technologically advanced human culture, but its neuronal mechanisms are poorly understood. Calculating with numbers requires temporary maintenance and manipulation of numerical information according to arithmetic rules. We explored the brain mechanisms involved in simple arithmetic operations by recording single-neuron activity from the medial temporal lobe of human subjects performing additions and subtractions. We found abstract and notation-independent codes for addition and subtraction in neuronal populations. The neuronal codes of arithmetic in different brain areas differed drastically. Decoders applied to time-resolved recordings demonstrate a static code in hippocampus based on persistently rule-selective neurons, in contrast to a dynamic code in parahippocampal cortex originating from neurons carrying rapidly changing rule information. The implementation of abstract arithmetic codes suggests different cognitive functions for medial temporal lobe regions in arithmetic.

INTRODUCTION

Mental arithmetic is an intricate skill and a hallmark of our scientifically advanced culture. Calculating with numbers requires semantic knowledge about numbers, online maintenance of numerical values, and their goal-directed transformation according to calculation rules. Therefore, mental arithmetic engages multiple brain systems, including those for the semantic representation of numeric values, the learning and memory of mathematical principles, and the cognitive control of mental operations.[1–4]

Studies in humans[1,2,5] and nonhuman primates[6–9] have indicated parts of the parietal and prefrontal cortices as core number representation and manipulation system. In particular, arithmetically selective brain areas have been identified in the parietal cortex of patients using intracranial electrocorticography (ECoG) recordings that measure summed and synchronized postsynaptic potentials (bulk tissue mass potentials).[10,11] Moreover, direct electrical stimulation studies in human patients have shown a specific arrest of counting and calculation performance during transient perturbation of parietal and frontal regions.[12–14] The latter investigations, in particular, suggest a causal involvement of parietal and frontal cortical regions in mental arithmetic.

However, recent findings implicate a wider cortical number network beyond parietal and frontal association cortices, also integrating the temporal lobe. Direct evidence resulted from ECoG studies in human patients; these recordings reconfirmed the presence of addition-selective locations not only in the posterior parietal cortex but also in the ventral temporal cortex.[10,11] In addition, functional neuroimaging implicated medial temporal lobe regions in the development of arithmetic fact knowledge,[3,15–17] including knowledge about arithmetic operators.[18] Moreover, performance enhancements in arithmetic fact retrieval are related to functional connectivity in hippocampal-neocortical circuits, including hippocampal-frontal[16,19] and hippocampal-parietal[16] connectivity. Hippocampal volume and functional connectivity of the hippocampus with dorsolateral and ventrolateral prefrontal cortices predict math tutoring success in children,[20] and reduced parahippocampal gray matter is associated with math learning disabilities (“developmental dyscalculia”).[21] Finally, we have recently shown directly by intracranial single-neuron recordings that the human medial temporal lobe (MTL) contains neurons that selectively respond to numerical values of different (symbolic and nonsymbolic) visual formats.[22] Here, we explored how single neurons in the human MTL represent the arithmetic addition and subtraction rules presented in different symbolic notations.

As a neuronal representation of the abstract rules applied to perceptual categories, rule-selective neurons have been identified in nonhuman primates.[23] They increase firing rates when a subject follows a specific rule but remain silent for alternative rules.[24,25] To bridge longer working memory delays necessary for mental arithmetic, two fundamentally different neuronal codes are conceivable: neurons might be tuned to a specific calculation rule and maintain this representation over long time periods via persistent firing. In this case, a decoder (a statistical classifier) trained on neuronal activity during a brief moment can successfully generalize across different time points. This

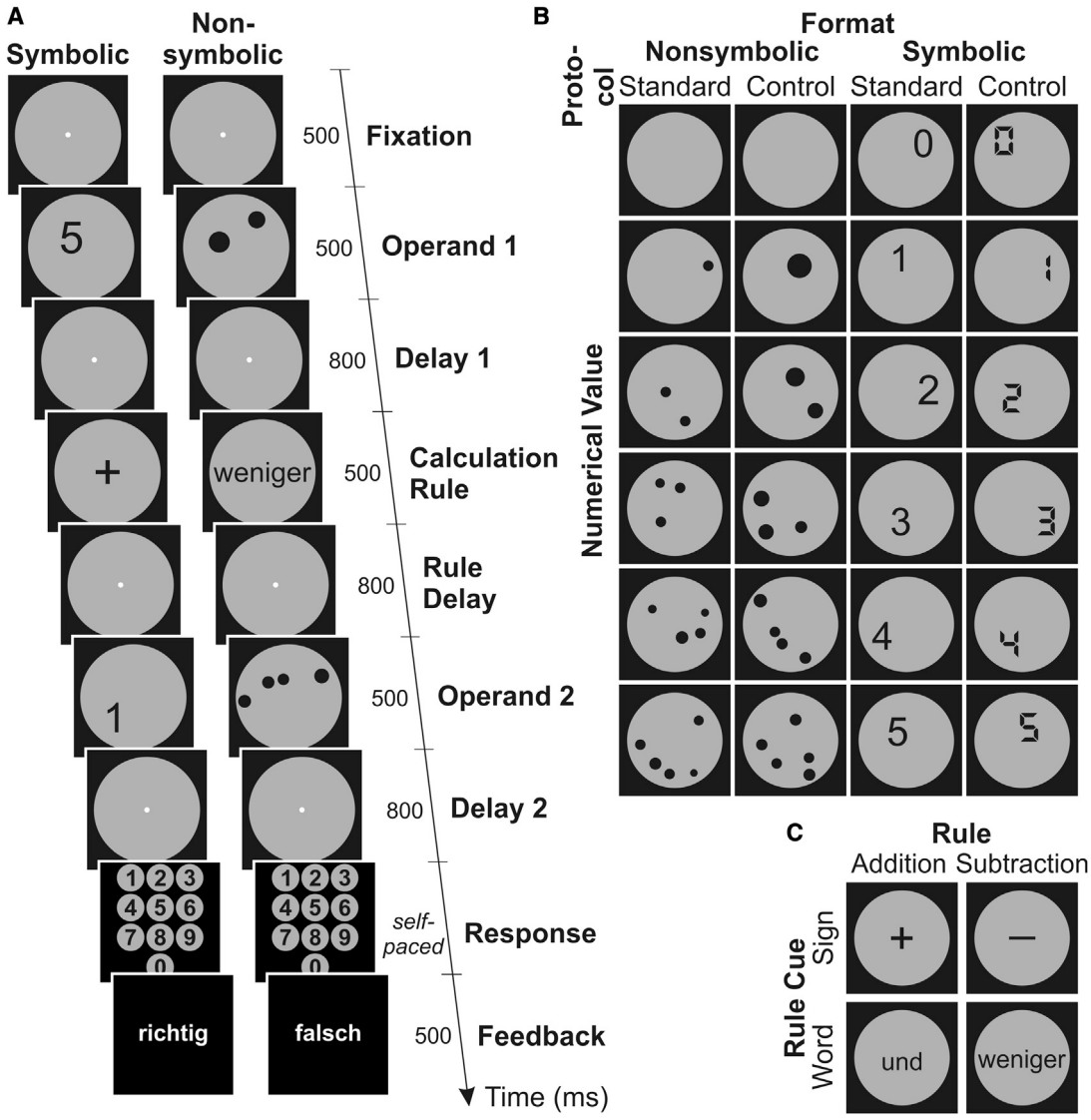


Figure 1. Behavioral task and example stimuli

(A) Experimental design of the calculation task. After visual fixation on the screen, stimuli were presented sequentially in the order operand 1—operator—operand 2. Each stimulus phase was followed by a brief delay. Afterward, the subjects were required to indicate the result of the calculation (ranging from 0 to 9) on an Arabic numeral panel and subsequently received feedback indicating whether the result was correct ("richtig") or false ("falsch").

(B) Example number stimuli for the nonsymbolic (numerosity) and symbolic format (numeral) for standard and control protocols. Numerical values of operand 1 ranged from 1 to 5; those of operand 2 ranged from 0 to 5.

(C) Example stimuli for the different mathematical rules, indicated by arithmetic signs ("+," "–") and written words ("und" [add], "weniger" [subtract]).

type of coding is known as static coding. Alternatively, neurons may fire sparsely and rapidly change tuning to calculation rules over time.[26] Under this scenario, a decoder trained on neuronal activity during one time point cannot generalize to the next. This has become known as dynamic coding.[27] By applying decoders to time-resolved recordings, we probe the codes for abstract arithmetic rules in the human MTL.

RESULTS

We asked nine human participants to perform simple addition and subtraction tasks on a computer display with operand values ranging from 0 to 5 (Figure 1A, see STAR Methods). Both operands were displayed with equal probability either as dot numerosities (nonsymbolic) or Arabic numerals (symbolic). Numerosities were shown in standard (variable dot size and arrangement) and control (constant total dot area and dot density) displays to control for non-numerical visual parameters; Arabic numerals were shown in two (standard and control) font types to ensure the generalization of symbols across visual shapes (Figure 1B). Addition and subtraction rules were instructed by two different notations, either as arithmetic signs (+, –) or written words (German "und" and "weniger," indicating "add" and "subtract"). The two rule notations allowed us to later

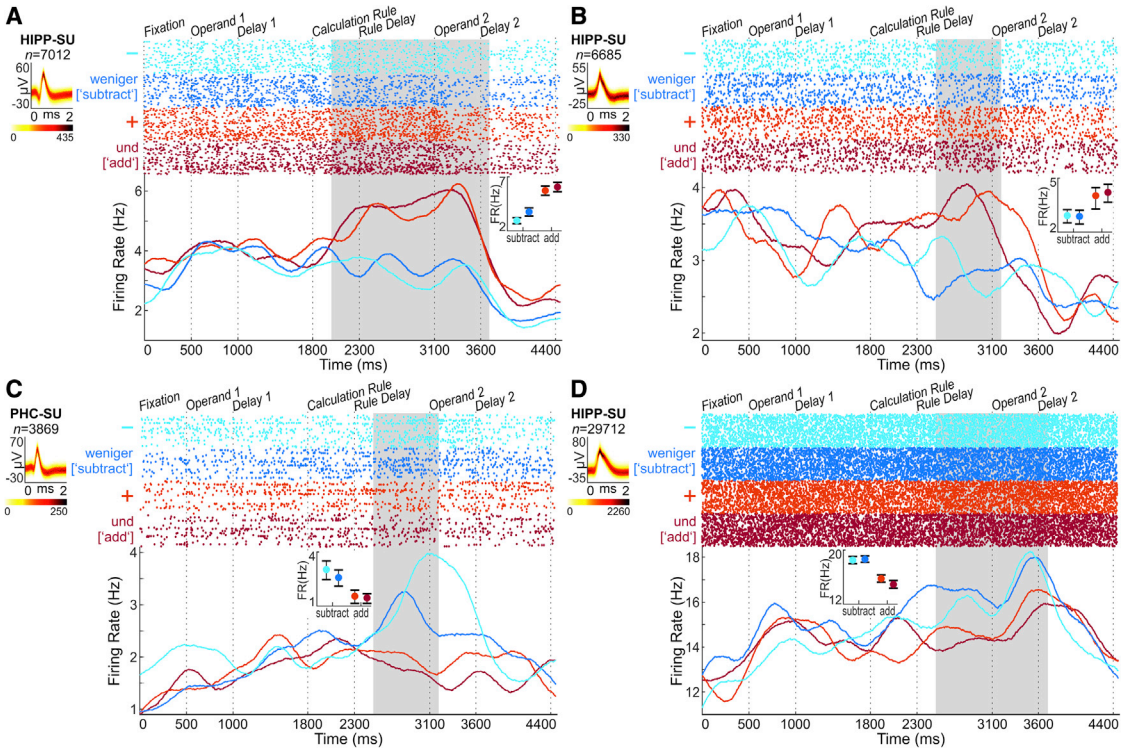


Figure 2. Neural responses of rule-selective neurons

(A–D) Across-trial responses of four example neurons responding with increased firing rate to the "addition" rule (A, B) and the "subtraction" rule (C, D), regardless of the concrete cue indicating the rule. The left small panels depict a density plot of the recorded action potentials (color darkness indicates the number of overlapping wave forms according to color scale at the bottom). Top panels show single-cell dot-raster plots for many repetitions of the rule cue (each line represents a trial and each dot represents an action potential, color coded according to the two rules and the two rule cues), and averaged instantaneous firing rates (spike-density histograms) are shown below. Blueish colors depict subtraction (for two different rule cues); reddish colors correspond to addition. Insets show average activity per rule cue during the rule discrimination period (gray shaded area), as defined by statistical significance in the ANOVA. Error bars denote SEM.

dissociate neural activity related to the visual properties of the cue (sign or word) from the abstract rule that it represented (addition versus subtraction) (Figure 1C). The participants' average performance was close to perfect (98.5% ± 0.6%) and comparable between addition and subtraction ($p = 0.97$; t test).

Single neurons respond to calculation rules

We recorded the action potentials of a total of 585 single neurons in the MTLs of the participants performing the calculation tasks: 126 neurons in parahippocampal cortex (PHC), 199 neurons in hippocampus (HIPP), 107 neurons in entorhinal cortex (EC), and 153 neurons in amygdala (AMY). As an obligatory but not sufficient prerequisite for mental calculation, MTL neurons were previously shown to represent numerical cardinality of the first operand.[22] We predicted that single neurons and neuronal populations also encode mentally performed additions and subtractions.

Using a multi-factor analysis of variance (ANOVA) (see STAR Methods), we first identified rule-selective neurons that selectively increased firing rates to either the addition or subtraction rule after the instruction of the calculation ("calculation rule") (Figure 1A). After the presentation of the calculation rule, addition-selective neurons enhanced firing whenever an addition was instructed (Figures 2A and 2B; reddish colors); whereas subtraction-selective neurons showed a specific increase in activity whenever a subtraction was cued (Figures 2C and 2D; blueish colors).

The proportion of neurons selectively tuned to calculation rules differed for different task periods and MTL areas (Figure 3). In the "calculation rule" period, a small but significant proportion of MTL neurons (4.8%; 28/585) was modulated by the arithmetic rule ($p < 0.001$; binomial test with $p_{chance} = 0.01$). Most of these neurons (3.5%; 20/585) showed activity that varied exclusively with the arithmetic rule ($p < 0.001$; binomial test), irrespective of the cue indicating that rule (i.e., no significant main effect for the factor cue, or any other factor) (Figure 3A, first column). Only PHC (7%; 9/126) and HIPP (4%; 7/199), but not EC and AMY, showed proportions of such exclusively rule-selective neurons higher than expected by chance ($p < 0.05$; binomial test with $p_{chance} = 0.01$, Bonferroni-corrected for multiple tests across areas, $n = 4$; Figures 3B and 3D). The overall proportion of rule-selective neurons increased in the "rule delay" period in which the value of operand 1 and the calculation instruction needed to be held in working memory to solve the task. Here, 6.0% (35/585) of MTL neurons were modulated by the arithmetic rule, with 5.3% (31/585) being exclusively rule-selective (both

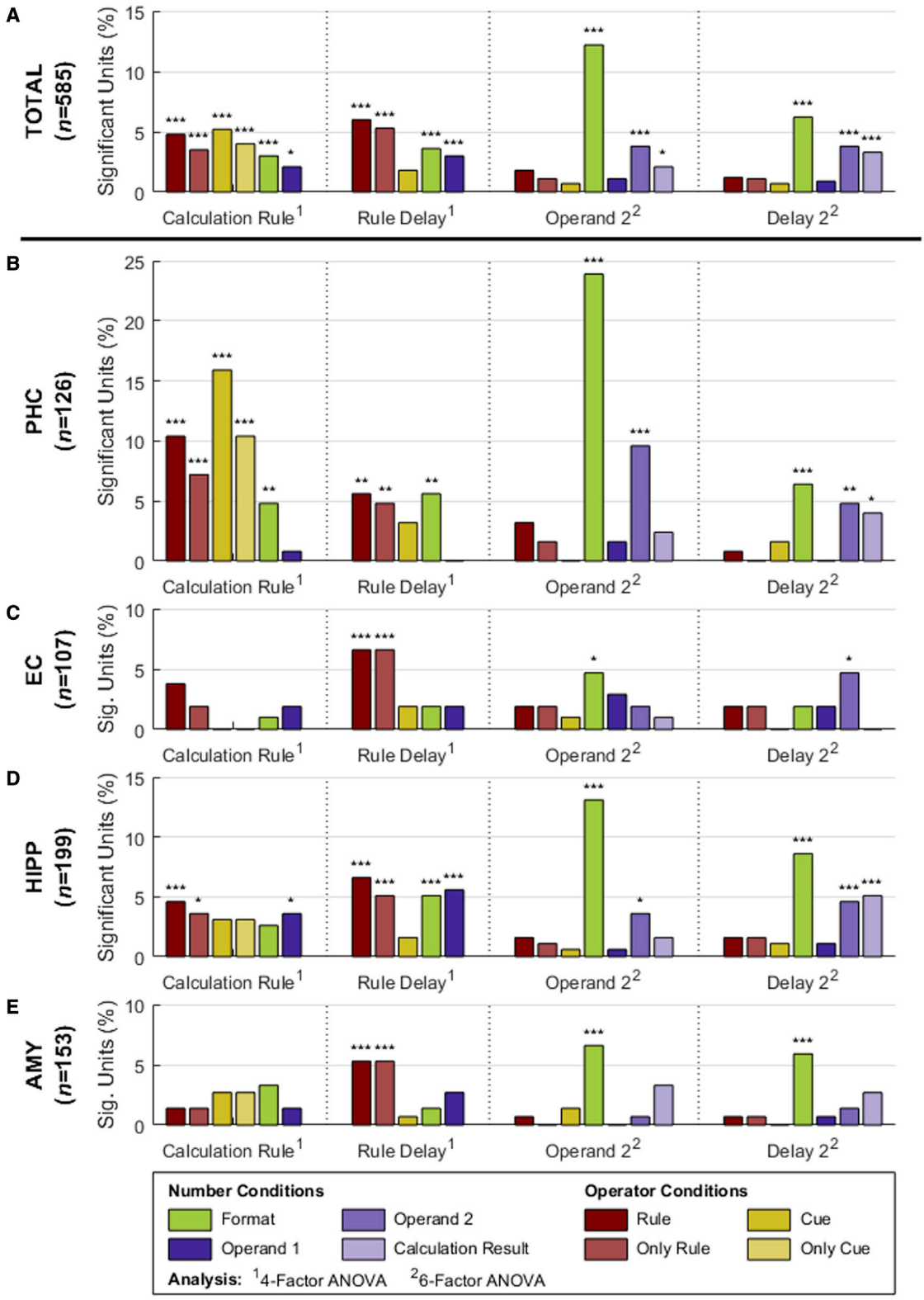


Figure 3. Neuronal selectivity of MTL single units

(A–D) Proportions of single units significant to different task factors for different MTL regions: (A) total population, (B) parahippocampal cortex, (C) entorhinal cortex, (D) hippocampus, and (E) amygdala. ANOVAs for the different task phases were evaluated at $\alpha = 0.01$. Neurons with an effect for "arithmetic rule," but no concurrent other main effects are termed "exclusively rule-selective" ("Only Rule"); same applies for factor "rule cue." Numbers of significant neurons were subjected to a Bonferroni-corrected (n = 4) binomial test; asterisks indicate significance (*p < 0.05, **p < 0.01, ***p < 0.001). See also Figures S1 and S2.

$p < 0.001$, binomial test; Figure 3A, second column). During this period, all MTL areas contained a significant proportion of exclusively calculation-rule-selective neurons (EC: 7%, 7/107; HIPP: 5%, 10/199; PHC: 5%, 6/126; and AMY: 5%, 8/153; all $p < 0.01$, Bonferroni-corrected binomial test) in addition to neurons coding other task-relevant factors (Figures 3B–3E). The proportion of rule-selective neurons dropped to chance level in the subsequent task periods ("operand 2" and "delay 2") (Figure 3A). Instead, the neurons represented the task factors additionally introduced with trial progression (for instance, the numerical value of operand 2) (Figure 3, third and fourth columns). In sum, a small but significant proportion of neurons encoded arithmetic rules after the presentation and memorization of the calculation rule prior to the presentation of operand 2. A separate analysis of the neurons in the left and right hemisphere qualitatively confirmed these results for each hemisphere in all regions (Figure S1). Note that the proportions of neurons showing significant selectivity during different trial periods do not generally represent the same neuronal populations (see Figure S2).

Cue-independent representation of addition and subtraction rules

MTL neurons showed variation in the time point and duration of rule selectivity across the task period. With increasing task complexity later in the trial, they also exhibited selectivity to several task factors, seemingly multiplexing the different information required to solve the task. Therefore, we focused our attention on the collective properties of groups of neurons, or "neuronal populations." This allowed us to read out (or "decode") information not only from an individual neuron but from a population of neurons that the subjects can base their decisions on. In combination with decoding methods, such as statistical classifiers, this allows to predict the accuracy and abstractness of arithmetic rule representations on a trial-by-trial basis.

To explore whether this variable activity yielded a reliable read-out of arithmetic rules, we adopted a machine-learning approach. We trained support vector machine (SVM) classifiers to discriminate between addition trials ("und" [add] and "+" cues combined) and subtraction trials ("weniger" [subtract] and "–" cues combined) across trial periods based on firing rates (see STAR Methods). The classifiers were then tested on novel data from the same neurons to explore how well they could predict the rules based on the information extracted from trials used during training. Cluster permutation tests ($p < 0.05$) were used to identify the trial intervals of classification accuracies significantly above chance level (50%).

We found long time intervals for which rule information could be successfully decoded in all MTL regions (black horizontal bars in Figures 4A–4D). Consistent with the single-cell analysis, we observed strong and long-lasting effects in PHC (Figure 4A; two selective periods interrupted by a non-selective period) and HIPP (Figure 4C). In HIPP particularly, the calculation rule was continuously decoded with high accuracy from rule onset until the end of the trial (Figure 4C). By contrast, rather weak and short-lived significant classification performance was observed in EC (Figure 4B) and AMY (Figure 4D). Separate analysis of left and right hemisphere qualitatively confirmed these results for PHC and HIPP (Figure S3).

We wanted to find out whether the calculation rules could be decoded irrespective of the rule cues, as would be expected for abstract rule coding. Therefore, we explored classification accuracies for individual rule cues during the previously found significant time intervals shown in Figures 4A–4D. We trained an SVM classifier using the firing rates combined for both rule cues per calculation rule in the individual significant windows for each MTL area. We then tested whether the SVM could predict the correct calculation rule from novel trials based on either one of the two applied cues per calculation rule (i.e., signs and words).

We found that classifier performance for addition and subtraction across word and sign rule cues was significant in all MTL areas ($p < 0.05$; permutation test compared with shuffled data labels). Highest classification accuracies were found in HIPP (addition: 74% and subtraction: 73%; Figure 4G) and PHC (addition: 65% and subtraction: 68%; averaged across both significant time windows; Figure 4E) followed by EC (addition: 65% and subtraction: 61%; Figure 4F) and AMY (addition: 58% and subtraction: 63%; Figure 4H). In AMY, performance was mainly due to accurate encoding of one specific cue (the "weniger" [subtract] cue), whereas classification accuracies were relatively low for the other three cues. Overall, information about the calculation rules was encoded irrespective of the rule cues prompting addition and subtraction, respectively.

Cross-temporal calculation rule decoding (static-dynamic)

Next, we explored the neuronal codes of arithmetic rules. By applying decoders to time-resolved recordings, we asked whether the code remained stable across trial time or rather changed dynamically (Figures 4I–4L). To this end, we performed a temporal cross-training analysis. We trained SVM classifiers on the firing rates from a given time point and tested them during other time points of the trial. This analysis was again performed separately for the four MTL areas, and the accuracy results were plotted in a confusion matrix spanning the trial times of classifier training against the trial times of classifier testing.

In PHC, we observed that classifiers trained during a specific time interval after rule cue presentation were only able to decode the arithmetic rule in the same time interval (Figure 4I). This resulted in high classification accuracies only along the main diagonal of the confusion matrix. The classifiers' inability to generalize the calculation rules across trial time periods indicates a dynamic neuronal code in PHC based on neurons that rapidly change their tuning properties with time.

A rather different picture emerged for HIPP (Figure 4K). Significant cross-temporal generalization from the end of the rule cue period all the way up to the end of the trial was present. A classifier trained on firing rates observed, e.g., during the rule delay, was still able to decode the calculation rule when tested on activity recorded during the operand 2 or even delay 2 phases. This resulted in a square-like accuracy pattern in the cross-temporal confusion matrix. This pattern argues for a static neuronal code in HIPP based on tuned neurons that remained stable across time throughout the trial.

In EC, cross-temporal generalization was weak and observed only to a small extent (Figure 4J). A square-like accuracy pattern emerged only around the rule delay. This suggests relatively

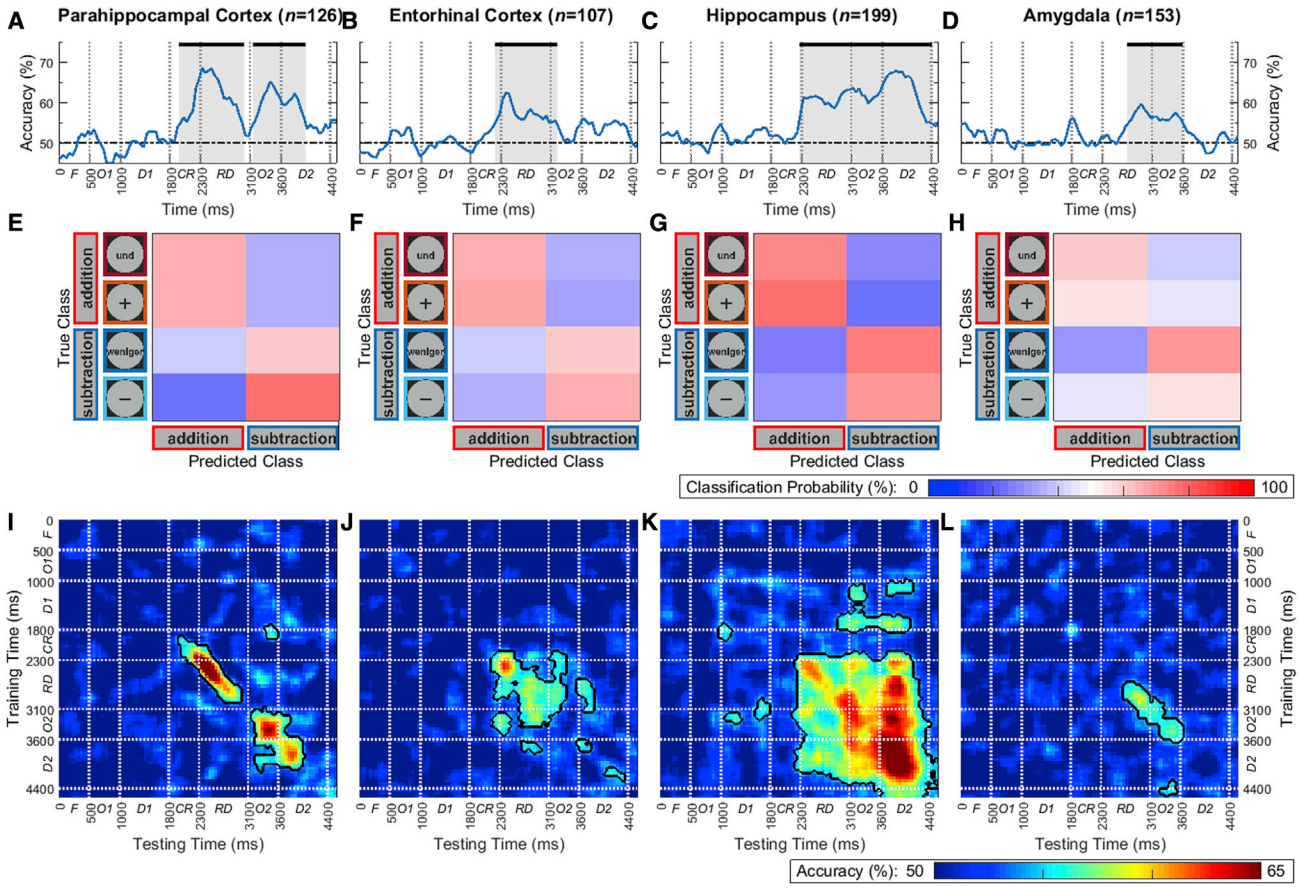


Figure 4. Rule decoding using a support vector machine (SVM) classifier. Decoding performance for the four different MTL regions (columns)

(A–D) Classification accuracy for decoding arithmetic rule information when training an SVM classifier on the instantaneous firing rates across the trial period. The dashed line represents chance level (50% for two classes). Black bars above the data and areas shaded in gray indicate significance ($p < 0.05$) when testing against performance for SVMs trained on shuffled data in a permutation test. Abbreviations at the axes indicate task phases: *F*, fixation; *O1*, operand 1; *D1*, delay 1; *CR*, calculation rule; *RD*, rule delay; *O2*, operand 2; and *D2*, delay 2.

(E–H) Confusion matrices derived from training an SVM classifier on firing rates averaged across the significant time windows in (A–D), respectively. (E) shows the average of the confusion matrices obtained for each significant window (depicted in A).

(I–L) Accuracy when training an SVM classifier at a given time point of the trial and testing on another one (the main diagonal of the matrix corresponds to the curve in [A–D]). Black contours indicate significance ($p < 0.05$) in a permutation test. See also Figures S3–S5.

stable calculation rule coding for as long as only the calculation rule was kept in working memory and before the second operand was presented. In AMY, neurons did not encode arithmetic rules abstractly (see Figure 4H). In addition, the cross-temporal classifier analysis showed only a mild accuracy diagonal during the rule delay and the operand 2 periods (Figure 4L). Both findings preclude statements about coding dynamics in AMY.

The observed response patterns, in particular a dynamic code in PHC contrasted by a static code in HIPP, were still present after equalizing the numbers of neurons for each MTL area (Figure S4; see STAR Methods). Information about the rule cues were only encoded during cue presentation in PHC (Figure S5), which further confirmed the abstractness of the calculation rule coding.

Cross-notation decoding of addition and subtraction

The previous analyses showed that the population of neurons differentiated between addition and subtraction rules and indicated that calculation rules were encoded irrespective of the rule cues. As a final step, we put this observation to the test and explored whether MTL neurons generalize calculation rules across rule notations.

We performed a time-resolved sliding-window decoding analysis and trained a SVM classifier on the trials of one rule notation and tested it on the other rule notation for the same calculation rule. First, we used all the word trials (“und” [add] and “weniger” [subtract]) as training data and all the sign trials [“+” and “–”] as test data [henceforth called “word → sign”]) and performed the same analyses as before (i.e., temporal cross-training classification and verification via fixed-window analysis). Then, we analyzed the generalization in the opposite direction, i.e., using all sign trials as training data and all word trials as test data (in the following called “sign → word”) and repeating the whole procedure. Generalization was judged successful if (1) synchronous intervals of significant classification for both directions of generalization were found (Figures 5A–5D)

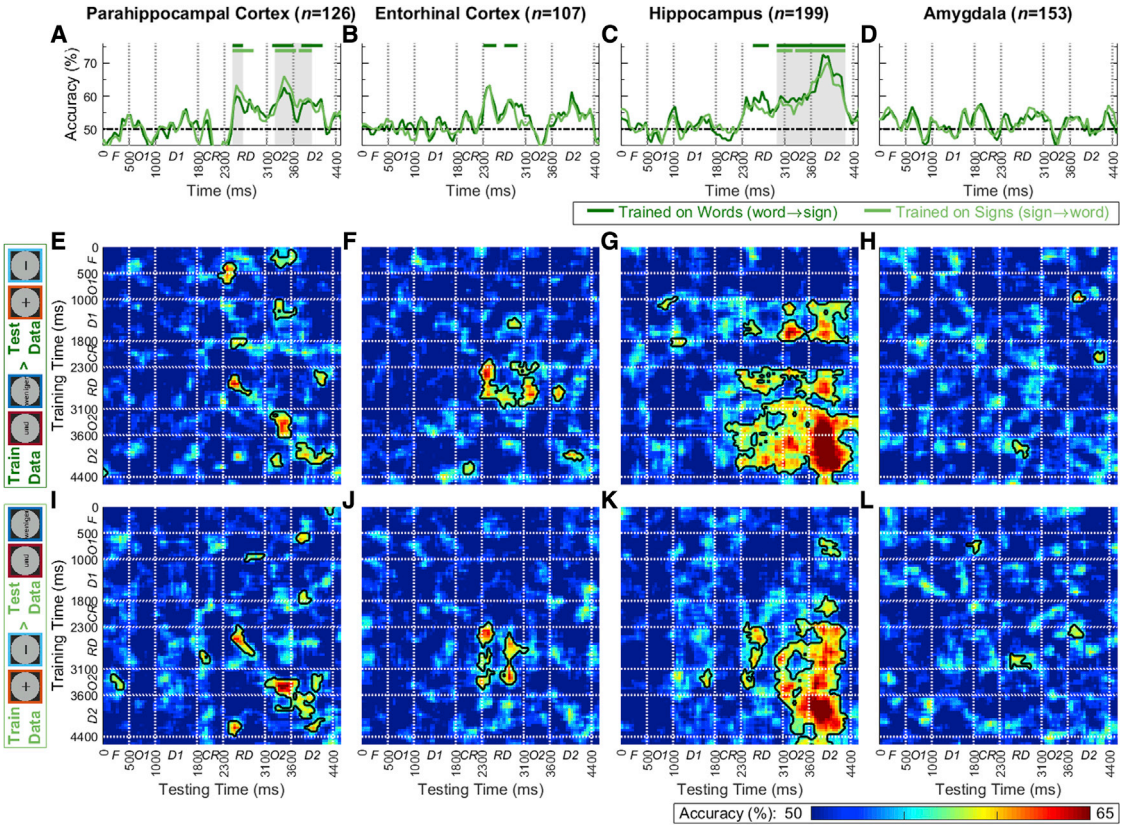


Figure 5. Generalization between arithmetic rule notations using an SVM classifier. Classifier performance for the four different MTL regions (columns)

(A–D) Classification accuracy when training an SVM on the instantaneous firing rates across the trial period for both directions of generalization. The dashed line represents chance level (50% for two classes). Light and dark green bars above the data indicate significance ($p < 0.05$) in a permutation test for both test directions ("word → sign" and "sign → word," respectively). The areas shaded in gray indicate the synchronous time windows used for the fixed-window classification analysis. Abbreviations at the axes indicate task phases: *F*, fixation; *O1*, operand 1; *D1*, delay 1; *CR*, calculation rule; *RD*, rule delay; *O2*, operand 2; and *D2*, delay 2.

(E–H) Accuracy in temporal cross-training analysis when a classifier is trained on trials showing word rule cues and tested on trials showing sign rule cues (the main diagonals of the matrices correspond to the dark green curves in [A–D]). Black contours indicate significance ($p < 0.05$) when testing against performance for SVMs trained on shuffled data in a permutation test.

(I–L) Accuracy in temporal cross-training analysis when a classifier is trained on trials showing sign cues and tested on trials showing word cues (the main diagonals of the matrices correspond to the light green curves in [A–D]). Same conventions as in (E–H).

and (2) performance in these synchronous intervals was significant (permutation test, $p < 0.05$) for each arithmetic rule after averaging classification accuracies across both directions of generalization.

Significant cross-notation rule decoding was present in PHC (Figure 5A) and HIPP (Figure 5C). In both areas, extended and overlapping intervals of significant rule classifications for both test directions ("word → sign" and "sign → word") emerged after rule cue offset and up to the end of the trial (significant phases are indicated by light and dark green horizontal bars in Figures 5A and 5C, with synchronous intervals indicated by gray shaded areas). In PHC, the accuracy of transfer was 57% for addition and 64% for subtraction (Figure 5A). In HIPP, transfer was even better and reached an average accuracy of 73% for addition and 69% for subtraction (Figure 5C). In both PHC and HIPP, the transfer for both calculation rules and both test directions were individually significant and thus pooled.

By contrast, the cross-notation decoding of calculation rules failed in AMY and EC. In AMY, classification accuracy remained at chance level throughout the whole trial for both directions of generalization (Figure 5D). In EC, cross-notation decoding briefly transferred for the direction "word → sign," but not for the direction "sign → word," and therefore failed our generalization criterion (Figure 5B). Thus, activity in PHC and HIPP did generalize arithmetic rule information across notations, whereas AMY and EC did not.

To explore the dynamics of rule codes during cross-notation generalization, we again employed a temporal cross-training analysis separately for the two generalization directions. The resulting confusion matrices confirmed the earlier findings (albeit with weaker effect size due to the reduced data dimensionality). In HIPP, they showed a static rule code for both notation generalization directions from rule cue offset to the end of the trial (indicated by the square-like significant classification pattern) (Figures 5G and 5K). By contrast, a dynamic code

emerged for both notation generalization directions in PHC, as evidenced by significant classification performance along the main diagonal of the matrices (Figures 5E and 5I). The findings for EC and AMY were unreliable due to the lack of rule-notation generalization found in these areas (Figures 5B and 5D). In summary, neuron populations in both PHC and HIPP generalized between calculation rule notation but exhibited fundamentally different rule codes.

DISCUSSION

Our findings demonstrate that the activity of neurons in the MTL carries sufficient information to allow a statistical classifier to discriminate between addition and subtraction instructions during mental calculation. After having been trained on activity during mental processing of one rule notation, decoding generalized to another notation cueing the same arithmetic operation. This generalization observed for arithmetic signs and words implies an abstract and notation-independent representation of arithmetic rules. Therefore, our research unveils a neuronal correlate for mental arithmetic, which generalizes between calculation tasks involving learned mathematical symbols.

Number and calculation recruit the MTL

Our discovery of arithmetic rule selectivity in MTL is in agreement with a growing body of studies that suggests that number and calculation recruit an interconnected network of cortical areas, including parts of the temporal lobe. We have recently shown directly by intracranial single-neuron recordings that the human MTL contains a significant percentage of neurons selectively tuned to numerical values.[22] This finding in humans concurs with the numerical tuning of hippocampal neurons in nonhuman primates.[28] Moreover, using functional imaging, the MTL has been shown to be involved in arithmetic skill acquisition and memory-based problem-solving strategies during childhood.[15,16,20] Similarly, ECoG studies in human patients reported addition-selective locations not only in the posterior parietal cortex but also in the ventral temporal cortex.[10,11] Together, these data implicate a wider cortical number network beyond parietal and frontal association cortices, also integrating the temporal lobe.

The MTL is suited to transform and manipulate representations of incoming numerical information.[22] It is highly interconnected with the frontal and parietal areas[29,30] that constitute the core number system involved in perceiving the number of sensory stimuli.[2,7,31] The prefrontal lobe, in particular, is associated with representing abstract rules and concepts, information that can be directly accessed by MTL.[23] Therefore, MTL could mediate the transformation of perceived numerical information in a working memory buffer. Interestingly, as an associative brain area, the MTL also contains sensorimotor neurons that are activated by hand-grasping observation and execution.[32] This opens the possibility that MTL may play a role in the sensorimotor translation of perceived and produced number,[33] a speculation yet to be explored.

Numerical rule-selective neurons

Although the observed rule-selective neurons encode arithmetic rules, this is not to say that some of these neurons may not also become engaged in encoding other types of rules that we have not explored. However, recordings show that rule-selective neurons in the nonhuman primate exhibit a substantial degree of specialization and preferentially respond only to quantitative rules applied to a specific magnitude type.[34] Therefore, we think that the majority of rule-selective neurons specifically and genuinely encode arithmetic rules.

So far, the neuronal correlates of addition and subtraction have not been studied in monkeys. However, what has been investigated is how single neurons respond to more basic mathematical operations, namely, "greater than" (>) and "less than" (<) operations.[6,24,34] In each trial, monkeys had to flexibly switch between these two rules according to rule cues and had to choose either a larger (in the case of the "greater than" rule) or smaller numerosity (in the case of the "smaller than" rule) than the one they had seen in the beginning of a trial. Recording from frontal and parietal areas, we found single neurons that responded selectively (by increased firing rates) to one of the two rules. Rule selectivity was stronger and more abundant in the frontal lobe than in the parietal lobe.[24] In monkeys, the frontal association areas are thus more important when it comes to nonsymbolic mathematical rules. This is consistent with imaging results in humans, where areas specific to calculation rather than simple number comparison are primarily found in the frontal lobe.[2]

The numerical coding capacities of such neurons do not seem to be fixed. Although neurons selective to numerical rules were recorded in brain areas in which previous studies had shown a relatively high proportion of numerosity-selective neurons,[7–9] the number of neurons representing pure numerical values were reduced when monkeys applied numerical rules.[24,34] This indicates that the neurons in the fronto-parietal number network dynamically encode different types of numerical information as a function of task demands. In the same vein, flexible numerical coding may apply to the MTL during a (top-down) interplay with the frontal lobe, depending on the precise mathematical task at hand.

Static and dynamic calculation codes

Mental calculation is a classic working memory task, and although working memory has traditionally been attributed to the prefrontal cortex,[35] more recent data suggest that the MTL may also be important in working memory tasks[36–38] and that it is part of a brain-wide network subserving working memory.[39]

Previous intracranial recording studies show that the delay activity of a selection of MTL neurons correlated with memory load and predict the successful retrieval of working memory contents.[39,40] These neurons' persistent activation maintained the same stimulus preference throughout several seconds of temporal gaps. This type of activity with robust across-time generalization is characteristic of a static code that we also observed in the hippocampus during mental calculation.

In contrast to the static code in hippocampus, we observed a clear dynamic code in PHC when processing calculation rules. Such a dynamic code based on sparsely bursting neurons supports the theory of activity-silent working memory.[41] It proposes that working memory can also be supported by short-term changes in synaptic weights. Synaptic weight changes are involved in episodic memory, which is why activity-silent working memory might be reminiscent of—or

even part of—episodic memory.[42] This fits with the finding that areas of the MTL are not only critically implicated in episodic memory but also important during working memory tasks.[36–40,43] Direct observation of neuronal reactivation after complete activity silence has recently been reported in a different working memory task in the human MTL.[44] Static and dynamic codes are not incompatible. Stable persistent activation with robust across-time generalization can exist in the presence of dynamically changing neuronal representations.[45,46]

Neuron recordings in human[39,40,43] and nonhuman primates,[47–49] as well as computational modeling,[50–52] suggest different cognitive functions for these two codes for working memory: although a dynamic code seems to suffice for short maintenance of more implicit information in memory, the intense mental manipulation of the attended working memory contents may require a static code. Following this logic, parahippocampal cortex may represent a short-term memory of the arithmetic rule, whereas downstream hippocampus may "do the math" and process numbers according to the arithmetic rule at hand. More fine-grained analyses, ideally combined with perturbation approaches,[53] will help to decipher the individual roles of brain areas and neuronal codes in mental arithmetic.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Experimental Task and Stimuli
  - Neurophysiological Recording
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Neuronal Analysis of Variance (ANOVA)
  - Support Vector Machine (SVM) Classification
  - Generalization of SVM Classification across Rule Cue Notations

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.01.054.

ACKNOWLEDGMENTS

We thank all patients for their participation. This research was supported by the Volkswagen Foundation and the German Research Council (SPP 2205, SFB 1089, Mo 930/4-2, Ni 618/10-1, Ni 618/11-1).

AUTHOR CONTRIBUTIONS

A.N. and F.M. designed the study. C.E.E. and F.M. recruited patients. J.B. and F.M. implanted the electrodes. E.F.K. and F.M. collected the data. E.F.K. analyzed the data with contributions from A.N. and F.M.. A.N., E.F.K., and F.M. wrote the paper. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 4, 2020
Revised: December 20, 2021
Accepted: January 20, 2022
Published: February 14, 2022

REFERENCES

1. Amalric, M., and Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. Proc. Natl. Acad. Sci. USA *113*, 4909–4917.
2. Arsalidou, M., and Taylor, M.J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. NeuroImage *54*, 2382–2393.
3. Menon, V. (2016). Memory and cognitive control circuits in mathematical cognition and learning. Prog. Brain Res. *227*, 159–186.
4. Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. Annu. Rev. Neurosci. *32*, 185–208.
5. Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. Neuron *53*, 293–305.
6. Bongard, S., and Nieder, A. (2010). Basic mathematical rules are encoded by primate prefrontal cortex neurons. Proc. Natl. Acad. Sci. USA *107*, 2277–2282.
7. Nieder, A. (2016). The neuronal code for number. Nat. Rev. Neurosci. *17*, 366–382.
8. Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. Curr. Biol. *26*, 1285–1294.
9. Viswanathan, P., and Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. *25*, 1259–1269.
10. Daitch, A.L., Foster, B.L., Schrouff, J., Rangarajan, V., Kaşikçi, I., Gattas, S., and Parvizi, J. (2016). Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. Proc. Natl. Acad. Sci. USA *113*, E7277–E7286.
11. Pinheiro-Chagas, P., Daitch, A., Parvizi, J., and Dehaene, S. (2018). Brain mechanisms of arithmetic: A crucial role for ventral temporal cortex. J. Cogn. Neurosci. *30*, 1757–1772.
12. Della Puppa, A., De Pellegrin, S., d'Avella, E., Gioffrè, G., Munari, M., Saladini, M., Salillas, E., Scienza, R., and Semenza, C. (2013). Right parietal cortex and calculation processing: intraoperative functional mapping of multiplication and addition in patients affected by a brain tumor. J. Neurosurg. *119*, 1107–1111.
13. Roux, F.-E., Boukhatem, L., Draper, L., Sacko, O., and Démonet, J.-F. (2009). Cortical calculation localization using electrostimulation. J. Neurosurg. *110*, 1291–1299.
14. Semenza, C., Salillas, E., De Pallegrin, S., and Puppa, A.D. (2017). Balancing the 2 hemispheres in simple calculation: evidence from direct cortical electrostimulation. Cereb. Cortex N. Y. N 1991 *27*, 4806–4814.
15. De Smedt, B., Holloway, I.D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. NeuroImage *57*, 771–781.
16. Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D.C., and Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. Nat. Neurosci. *17*, 1263–1269.
17. Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: a review of brain imaging studies. Dev. Cogn. Neurosci. *30*, 265–279.
18. Mathieu, R., Epinat-Duclos, J., Léone, J., Fayol, M., Thevenot, C., and Prado, J. (2018). Hippocampal spatial mechanisms relate to the development of arithmetic symbol processing in children. Dev. Cogn. Neurosci. *30*, 324–332.

19. Cho, S., Metcalfe, A.W., Young, C.B., Ryali, S., Geary, D.C., and Menon, V. (2012). Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. J. Cogn. Neurosci. *24*, 1849–1866.
20. Supekar, K., Swigart, A.G., Tenison, C., Jolles, D.D., Rosenberg-Lee, M., Fuchs, L., and Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. Proc. Natl. Acad. Sci. USA *110*, 8230–8235.
21. Rykhlevskaia, E., Uddin, L.Q., Kondos, L., and Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. Front. Hum. Neurosci. *3*, 51.
22. Kutter, E.F., Bostroem, J., Elger, C.E., Mormann, F., and Nieder, A. (2018). Single neurons in the human brain encode numbers. Neuron *100*, 753–761.e4.
23. Mansouri, F.A., Freedman, D.J., and Buckley, M.J. (2020). Emergence of abstract rules in the primate brain. Nat. Rev. Neurosci. *21*, 595–610.
24. Vallentin, D., Bongard, S., and Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. J. Neurosci. *32*, 6621–6630.
25. Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.
26. Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. Neuron *78*, 364–375.
27. King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. *18*, 203–210.
28. Opris, I., Santos, L.M., Gerhardt, G.A., Song, D., Berger, T.W., Hampson, R.E., and Deadwyler, S.A. (2015). Distributed encoding of spatial and object categories in primate hippocampal microcircuits. Front. Neurosci. *9*, 317.
29. Goldman-Rakic, P.S., Selemon, L.D., and Schwartz, M.L. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. Neuroscience *12*, 719–743.
30. Suzuki, W.A. (2009). Comparative analysis of the cortical afferents, intrinsic projections, and interconnections of the parahippocampal region in monkeys and rats. In The Cognitive Neurosciences, Fourth Edition (Massachusetts Institute of Technology), pp. 659–674.
31. Nieder, A. (2004). The number domain- can we count on parietal cortex? Neuron *44*, 407–409.
32. Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. Curr. Biol. *20*, 750–756.
33. Anobile, G., Arrighi, R., Castaldi, E., and Burr, D.C. (2021). A sensorimotor numerosity system. Trends Cogn. Sci. *25*, 24–36.
34. Eiselt, A.-K., and Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. J. Neurosci. *33*, 7526–7534.
35. Goldman-Rakic, P.S. (1995). Cellular basis of working memory. Neuron *14*, 477–485.
36. Goodrich, R.I., Baer, T.L., Quent, J.A., and Yonelinas, A.P. (2019). Visual working memory impairments for single items following medial temporal lobe damage. Neuropsychologia *134*, 107227.
37. Koen, J.D., Borders, A.A., Petzold, M.T., and Yonelinas, A.P. (2017). Visual short-term memory for high resolution associations is impaired in patients with medial temporal lobe damage. Hippocampus *27*, 184–193.
38. Olson, I.R., Moore, K.S., Stark, M., and Chatterjee, A. (2006). Visual working memory is impaired when the medial temporal lobe is damaged. J. Cogn. Neurosci. *18*, 1087–1097.
39. Kornblith, S., Quian Quiroga, R., Koch, C., Fried, I., and Mormann, F. (2017). Persistent single-neuron activity during working memory in the human medial temporal lobe. Curr. Biol. *27*, 1026–1032.
40. Kamiński, J., Sullivan, S., Chung, J.M., Ross, I.B., Mamelak, A.N., and Rutishauser, U. (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nat. Neurosci. *20*, 590–601.
41. Stokes, M.G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn. Sci. *19*, 394–405.
42. Beukers, A.O., Buschman, T.J., Cohen, J.D., and Norman, K.A. (2021). Is activity silent working memory simply episodic memory? Trends Cogn. Sci. *25*, 284–293.
43. Boran, E., Fedele, T., Klaver, P., Hilfiker, P., Stieglitz, L., Grunwald, T., and Sarnthein, J. (2019). Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. Sci. Adv. *5*, eaav3687.
44. Bausch, M., Niediek, J., Reber, T.P., Mackay, S., Boström, J., Elger, C.E., and Mormann, F. (2021). Concept neurons in the human medial temporal lobe flexibly represent abstract relations between concepts. Nat. Commun. *12*, 6164.
45. Murray, J.D., Bernacchia, A., Roy, N.A., Constantinidis, C., Romo, R., and Wang, X.J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proc. Natl. Acad. Sci. USA *114*, 394–399.
46. Spaak, E., Watanabe, K., Funahashi, S., and Stokes, M.G. (2017). Stable and dynamic coding for working memory in primate prefrontal cortex. J. Neurosci. *37*, 6503–6516.
47. Buschman, T.J., Siegel, M., Roy, J.E., and Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci. USA *108*, 11252–11255.
48. Mendoza-Halliday, D., Torres, S., and Martinez-Trujillo, J.C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. Nat. Neurosci. *17*, 1255–1262.
49. Sarma, A., Masse, N.Y., Wang, X.-J., and Freedman, D.J. (2016). Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. Nat. Neurosci. *19*, 143–149.
50. Masse, N.Y., Yang, G.R., Song, H.F., Wang, X.-J., and Freedman, D.J. (2019). Circuit mechanisms for the maintenance and manipulation of information in working memory. Nat. Neurosci. *22*, 1159–1167.
51. Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. Science *319*, 1543–1546.
52. Wolff, M.J., Jochim, J., Akyürek, E.G., and Stokes, M.G. (2017). Dynamic hidden states underlying working-memory-guided behavior. Nat. Neurosci. *20*, 864–871.
53. Lee, D.K., Fedorenko, E., Simon, M.V., Curry, W.T., Nahed, B.V., Cahill, D.P., and Williams, Z.M. (2018). Neural encoding and production of functional morphemes in the posterior temporal lobe. Nat. Commun. *9*, 1877.
54. Niediek, J., Boström, J., Elger, C.E., and Mormann, F. (2016). Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours. PLoS ONE *11*, e0166598.
55. Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., and Koch, C. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. J. Neurosci. *28*, 8865–8872.
56. Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 1–27.
57. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Cheetah software	Neuralynx Inc.	https://neuralynx.com/software/cheetah
Combinato spike sorting software	Niediek et al.[54]	https://github.com/jniediek/combinato
MATLAB R2017a	MathWorks	https://de.mathworks.com/
Psychtoolbox	Brainard (1997)	http://psychtoolbox.org/
LIBSVM	Chang and Lin[2]	https://www.csie.ntu.edu.tw/~cjlin/libsvm/
Other		
Behnke-Fried depth electrodes	AD-TECH Medical Instrument Corp.	https://adtechmedical.com/depth-electrodes
ATLAS neurophysiology system	Neuralynx Inc.	https://neuralynx.com/news/techtips/atlas-neurophysiology-system-for-cogneuro-applications

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Florian Mormann (florian.mormann@ukbonn.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data and custom-built MATLAB code can be found in a GitHub repository (https://github.com/EstherKutter/Neuronal-Codes-For-Arithmetic-Rule-Processing-In-The-Human-Brain).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Nine neurosurgical patients (4 male, all right-handed, mean age 43.3 years) undergoing treatment for pharmacologically intractable epilepsy participated in the study. Informed written consent was obtained from each patient. All studies conformed to the guidelines of the Medical Institutional Review Board at the University of Bonn, Germany. Other parts of the current data set were published in a previous publication.[22]

METHOD DETAILS

Experimental Task and Stimuli

Subjects performed a calculation task that required them to calculate the result of a simple arithmetic problem (Figure 1A). During experimental sessions, subjects sat in bed, facing a touch-screen laptop (display diagonal 11.7 in, resolution 1366x768 px) on which stimuli were presented at a distance of approximately 50 cm. They were not informed about hypotheses or purposes of the experiment, in order to avoid any bias.

Each trial began after a 500 ms fixation phase. Stimuli were presented successively in the order operand 1 – operator – operand 2, for 500 ms each, followed each by a 800 ms delay phase. Afterwards, subjects responded in a self-paced manner by touching the number matching the result of the calculation on a number pad showing the arabic numerals 0 to 9 that was presented on the screen. After a 500 ms feedback display ('richtig' [correct] or 'falsch' [false]) the next trial was started automatically.

All stimuli were presented within a filled gray circle (diameter approx. 6° of visual angle) on a black background. During fixation and delay phases, we presented a white fixation spot in the center of the gray area. During stimulus presentation, the fixation spot disappeared to avoid confusion with nonsymbolic stimuli and to distinguish it clearly from the nonsymbolic 'zero'-stimulus that was included as a potential operand 2-stimulus for control purposes.

Numerical values of operand 1 ranged from 1 to 5, and were in two visual 'formats', either 'nonsymbolic' arrays of randomly placed black dots of varying sizes with the number of dots corresponding to the respective numerical value ('numerosities'), or 'symbolic' black Arabic digits at randomized locations ('numerals'). Number stimuli of operand 2 ranged from 0 to 5, and were the same as for operand 1. The nonsymbolic 'zero'-stimulus was presented as the empty gray circle without fixation spot.

We used two 'protocols', standard and control displays, for both nonsymbolic and symbolic number formats (Figure 1B) in order to control for low-level visual features. The standard nonsymbolic numerosity displays consisted of dots at randomized locations and of pseudo-randomly varied sizes (diameter 0.3° to 0.8° of visual angle); in the control displays, we equated the overall screen area and density of the dots across numerosities. For the Arabic numerals, different fonts were used as standard (Helvetica, 34 pt) and control (DS-Digital, 34 pt) displays. A session comprised 50% nonsymbolic and 50% symbolic stimuli. Within each format, standard and control protocols were shown with equal probability of 50%.

We applied two different mathematical rules, i.e., addition and subtraction (Figure 1C). Two distinct cues, i.e., the mathematical sign ('+' or '–') or a verbal analogue ('und' [add] and 'weniger' [subtract]), were used for each rule (all Helvetica, 34 pt, presented in the center), in order to dissociate neuronal activity related purely to visual properties of the operator from the rule that it represented.

Overall, the task comprised seven factors. Five of these factors were varied systematically: Format (symbolic vs. nonsymbolic), protocol (standard vs. control) and numerical value of operand 1 (1–5), as well as mathematical rule (addition vs. subtraction) and rule cue (sign vs. word). Operand 2 was always of the same format and protocol as operand 1, but with random numerical values 0–5, albeit guaranteeing calculation results between 0 and 9. Due to this constraint, it was impossible to balance the other two factors 'numerical value of operand 2' (e.g., '5' is less likely to appear than '1', given that 'X–5' is only valid for X = {5}, but 'X-1' is valid for X = {1,2,3,4,5}), and 'numerical value of calculation result' (e.g., '4' is more likely to appear than '9', given the possible combinations of operands to obtain this result).

Each session consisted of a total of 320 trials and was divided into four blocks of 80 trials each, comprising the different conditions in pseudo-random order. To familiarize subjects with the task, sessions started with 10 rehearsal trials that were excluded from further analysis.

Neurophysiological Recording

To localize the epileptic focus for possible clinical resection, each subject was implanted bilaterally with chronic intracerebral depth electrodes in the medial temporal lobe (MTL). The exact electrode locations and numbers were defined exclusively by clinical criteria and varied across subjects. We used 9–10 clinical Behnke-Fried depth electrodes (AD-Tech Medical Instrument Corp., Racine, WI) to record neuronal signals. Each depth electrode contained a bundle of nine platinum-iridium micro-electrodes protruding from its tip by approximately 4 mm. Each bundle consisted of eight high-impedance active recording channels and one low-impedance reference electrode. A 256-channel ATLAS neurophysiology system (Neuralynx Inc., Bozeman, MT) was used to filter (bandwidth 0.1–9,000 Hz), amplify and digitize (sampling rate 32768 Hz) the differential neuronal signals (recording range ±3200 μV). The Cheetah software (Neuralynx Inc., Bozeman, MT) was used to synchronize the behavioral data with the recorded spikes via 8-bit timestamps.

Neuronal signals were band-pass filtered (bandwidth 300–3,000 Hz), then spikes were detected and pre-sorted automatically using the Combinato software.[54] Manual verification and classification as artifact, multi- or single unit was based on spike shape and its variance, inter-spike interval distribution per cluster and the presence of a plausible refractory period. Only units that responded with an average discharge rate of >1 Hz during stimulus presentation (fixation onset to delay 2 offset) were included in the analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Neuronal Analysis of Variance (ANOVA)

Only single units (n = 585) were included in the following analyses. All analyses were performed separately for each MTL area to identify regional differences (PHC: 126 units; EC: 107 units; HIPP: 199 units; AMY: 153 units). As all participants performed the task with high proficiency (98.5% ± 0.6%, range 90.3%–99.8%), we did not exclude the negligible number of error trials from the analyses.

For each unit, activity was analyzed separately for the different task phases involving rule processing. For each stimulus phase (calculation rule and operand 2 phase), discharge rates were measured in a 400 ms window starting 200 ms after stimulus onset. For each delay phase (rule delay and delay 2 phase), activity was assessed in a 700 ms window starting 200 ms after delay onset (latency chosen based on Mormann et al.[55]) In total, six factors were analyzed: 'mathematical rule' (addition/ subtraction) and 'rule cue' (word/ sign), as well as 'format' (nonsymbolic/ symbolic), 'numerical value of operand 1' (numbers 1–5), 'numerical value of operand 2' (numbers 0–5), and 'numerical value of calculation result' (numbers 0–9). We pooled over the factor 'protocol' given its incomparability for the different formats[4] and its irrelevance for the processing of the rule cues. For each task phase, we performed an ANOVA considering only those factors relevant for that phase. That is, for the calculation rule and rule delay phase, a 4-way ANOVA with the factors 'mathematical rule', 'rule cue', 'format' and 'numerical value of operand 1' was performed. For the operand 2 and delay 2 phase, we calculated a 6-way ANOVA with the factors 'mathematical rule', 'rule cue', 'format', 'numerical value of operand 1', 'numerical value of operand 2' and 'numerical value of calculation result'. All ANOVAs were evaluated at α = 0.01. A unit was counted as exclusively rule-selective ("Only Rule") if a significant main effect was observed for the factor 'mathematical rule', and there was no significant main effect for any other factor. Exclusive cue-selectivity ("Only Cue") was defined analogously.

To evaluate the significance of unit proportions, we subjected the number of significant neurons to a binomial test with an *a priori* probability of $p = 0.01$ corresponding to the alpha level for neurons to be regarded as significant, Bonferroni-corrected for multiple comparisons across different areas ($n = 4$).

Support Vector Machine (SVM) Classification

All single units were included in the following population analyses, irrespective of any selectivities found in the ANOVA. For each unit, data were divided into two classes, assigning the label 'addition' to trials with the cues 'und' [add] and '+', and the label 'subtraction' to the trials with the cues 'weniger' [subtract] and '–'. With 80 trials per cue, each class comprised 160 trials. For temporal cross-training classification, spike trains of each unit were smoothed (Gaussian kernel, $\sigma = 150$ ms, window size 300 ms) trial-wise within the trial window of 0–4500 ms (i.e., from fixation onset to 100 ms after delay 2 offset). An SVM classifier with a default linear SVM kernel[56] was then trained on the instantaneous firing rates at a certain time point, and tested on firing rates at different time points (sampling interval 50 ms).

We applied 10-fold cross-validation, i.e., we created 10 equal-size complementary splits of our dataset, balancing conditions within each split. Then, 9 splits were used as training set (comprising 288 trials), the remaining split was used as test set (comprising 32 trials). All firing rates were normalized by z-scoring (mean and standard deviation obtained from training data only), then we fitted the classifier to the training data and assessed the predictive accuracy by counting the instances that a certain activity pattern of the test data was labeled correctly. This process was repeated 10 times, using each of the 10 splits exactly once as the validation set. The results were then averaged across all splits.

To identify temporal clusters during which accuracy differed significantly from chance level (50% for two classes), the analysis was repeated with randomly shuffled trial labels ($n_{perm} = 1000$), and a cluster permutation test[57] was performed. In short, we identified temporal clusters of statistical significance by comparing the true accuracy values against the distribution of random ones ($\alpha_{clus} = 0.05$). The significance of these 'candidate clusters' was then evaluated by comparison with the clusters of the random data ($p_{rank} < 5\%$), using cluster size as a test statistic (i.e., number of connected significant 'pixels' in the cross-temporal accuracy matrix, or cluster length for the 'diagonal curve' when training and test time points were identical).

High accuracy values do not imply *per se* that the classifier has learned to encode abstract rule information; comparable accuracies might also be achieved if the SVM had learned to encode only one specific cue perfectly, but remained at chance level for the other three cues. To account for this, an SVM (with the same settings as above) was trained and tested on the firing rates obtained by averaging across the significant interval when training and testing at the same time point. In this fixed-window analysis we used the following time windows: PHC: 1950–3000 ms and 3150–4000 ms; EC: 2200–3200 ms; HIPP: 2250–4400 ms; AMY: 2700–3600 ms. We generated a confusion matrix which counted the frequency at which a trial of a certain rule cue was assigned different labels by the classifier, and calculated the accuracy per mathematical rule by averaging classification probabilities across the corresponding cues. In PHC, we trained a classifier and assembled confusion matrix and classification probabilities separately for each of the two significant time windows. Then, we averaged across both models to obtain one overall confusion matrix and overall average accuracies per rule. To evaluate significance, we repeated the analysis with shuffled labels ($n_{perm} = 1000$) and applied a permutation test ($\alpha = 0.05$).

As control, we equalized population sizes by drawing a random subset of units per area ($n = 107$) and re-calculated all analyses. This process was repeated 10 times, and the overall statistic was taken to be the mean of the stratified populations.

Finally, we assessed the units' ability to distinguish the two cue types (as opposed to the arithmetic rule information) by assigning the label 'word' to the 'und' [add] and 'weniger' [subtract] trials, and the label 'sign' to the '+' and '–' trials. We then repeated the temporal cross-training classification analysis and trained an SVM classifier on the window significant in the permutation test to generate the confusion matrix and average accuracy per cue type. The same procedures and settings as above (except for the labeling) were used for this control analysis.

Generalization of SVM Classification across Rule Cue Notations

To assess how well the results of the SVM classification might generalize to a different cue type, spike trains of all units were again trial-wise smoothed within the trial window (parameters as above), and labeled as before. Data were then divided into a training and a test set according to the rule cue.

First, all word trials (i.e., 'und' [add] trials labelled 'addition' and 'weniger' [subtract] trials labelled 'subtraction') served as training dataset. We applied 10-fold cross-validation, i.e., we split the training data into 10 balanced subsamples and used 9 splits as training dataset (comprising 144 trials). All sign trials (i.e., '+' trials labelled 'addition' and '–' trials labelled 'subtraction') served as test dataset (comprising 160 trials). Temporal cross-training classification was then performed using the same parameters and procedures as before. This process was repeated 10 times, leaving out each of the 10 subsamples exactly once. The results were then averaged across all splits. Again, significant temporal intervals were identified using a cluster permutation test ($n_{perm} = 1000$; $\alpha_{clus} = 0.05$; $p_{rank} < 0.05$). Generalization was analyzed also in the opposite direction, i.e., using sign trials as training dataset and word trials as test dataset, following the same procedures and settings as above.

Next, we identified synchronous intervals, i.e., time windows for which significant classification was observed for both directions of generalization. Intervals in which significance breaks of at most 150 ms for either one of the test directions occurred were

considered synchronous. Based on this criterion, we identified the following time windows: PHC: 2450–2650 ms and 3250–3950 ms; HIPP: 2950–4250 ms. For each direction, an SVM classifier (with the same settings as above) was trained on the firing rates obtained by averaging the training data across these synchronous intervals. Then, we tested the models on the firing rates obtained by averaging the test data across the same time window, and generated the confusion matrix. As before, in PHC, a classifier was trained for each of the two time windows. We then averaged the confusion matrices obtained for each interval to get one overall confusion matrix.

Generalization was then judged successful if (a) we found synchronous intervals of significant classification in the temporal cross-training analysis, and (b) performance in the fixed-window analysis was significant in a permutation test (n_{perm} = 1000; α = 0.05) for each arithmetic rule after averaging classification accuracies across both directions of generalization.

Current Biology, Volume *32*

Supplemental Information

Neuronal codes for arithmetic rule processing in the human brain

Esther F. Kutter, Jan Boström, Christian E. Elger, Andreas Nieder, and Florian Mormann

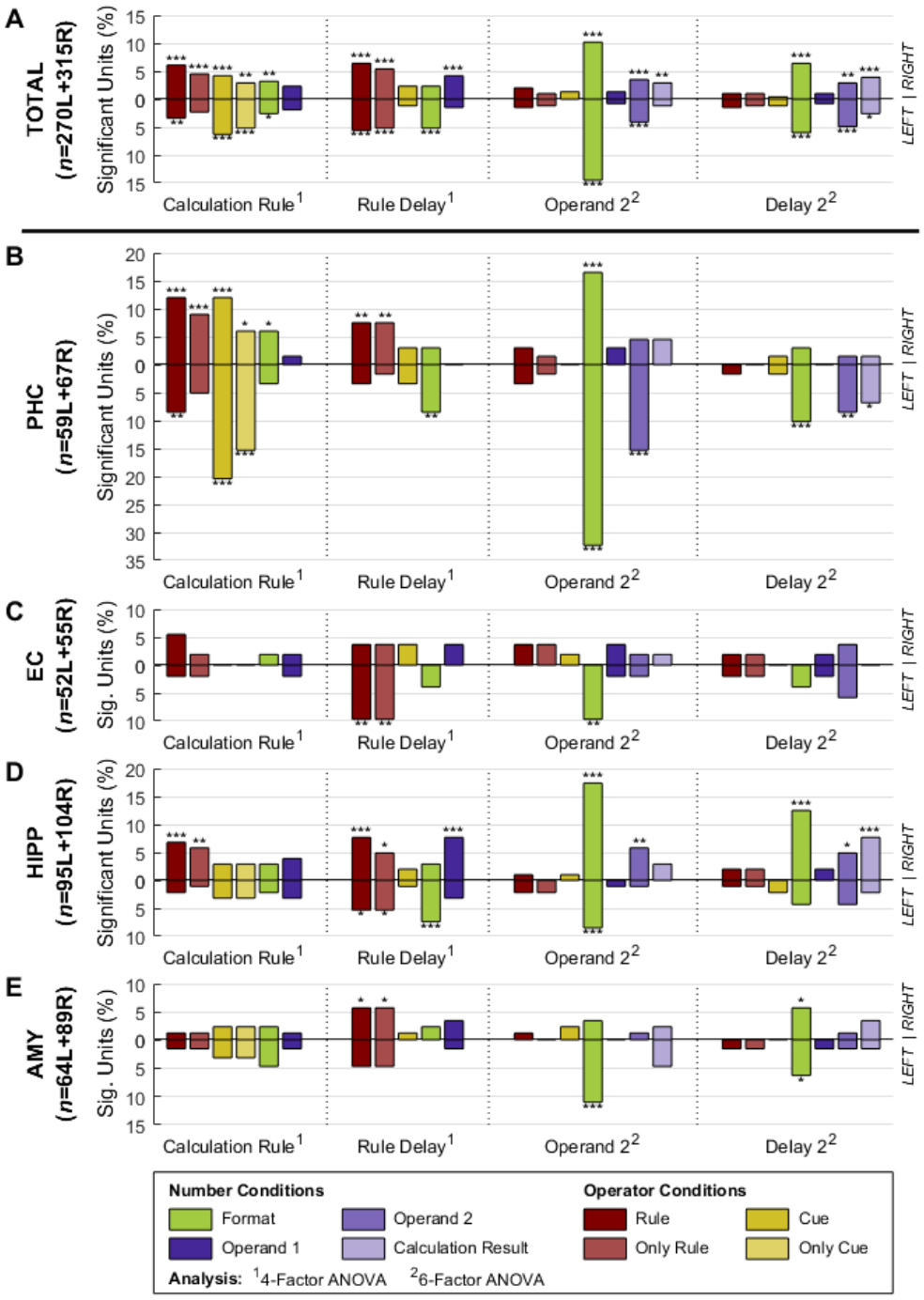


Figure S1: Neuronal Selectivity of MTL Single Units per Hemisphere. Related to Figure 3. Proportions of single units significant to different task factors for different MTL regions and hemispheres: **(A)** total population, **(B)** parahippocampal cortex, **(C)** entorhinal cortex, **(D)** hippocampus, and **(E)** amygdala. Proportions of units from the right and left hemisphere are depicted in the upper and lower rows, respectively. ANOVAs for the different task phases were evaluated at $\alpha = 0.01$. Neurons with an effect for 'arithmetic rule', but no concurrent other main effects are termed 'exclusively rule-selective' ("Only Rule"); same for factor 'rule cue'. Numbers of significant neurons were subjected to a binomial test, Bonferroni-corrected for multiple comparisons across areas ($n_1 = 4$) and hemispheres ($n_2 = 2$); asterisks indicate significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

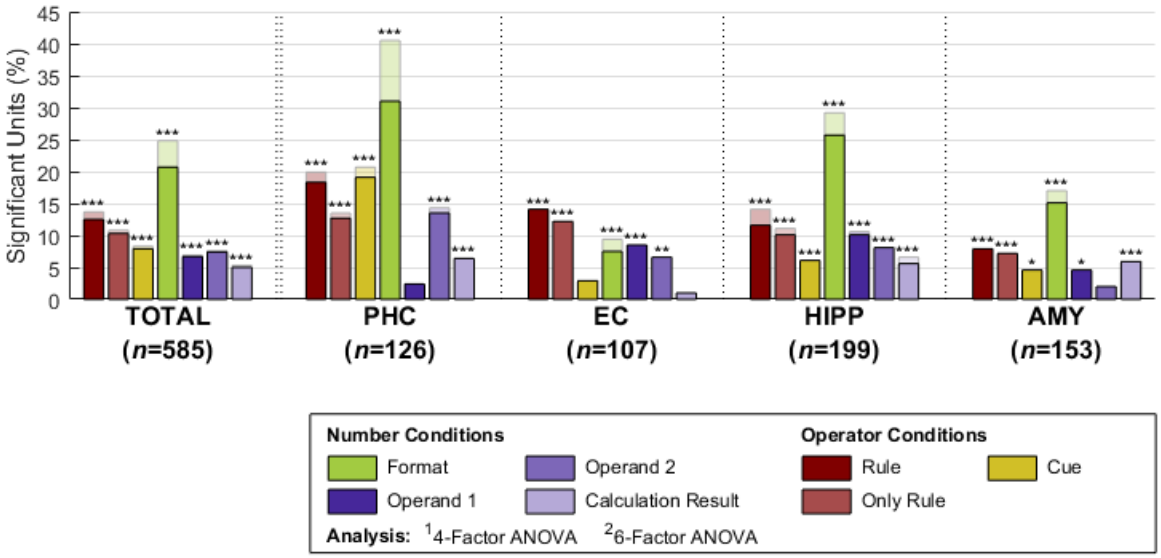


Figure S2: Neuronal Selectivity of MTL Single Units *a*cross Periods. Related to Figure 3. Proportions of single units significant in any (i.e., one or more) of the task periods (solid bars) along with percentages added up across *the four* trial periods (light bars). ANOVAs for the different task phases were evaluated at $\alpha = 0.01$. Neurons with an effect for 'arithmetic rule', but no concurrent other main effects are termed 'exclusively rule-selective' ("Only Rule"). Numbers of significant neurons were subjected to a binomial test, Bonferroni-corrected for multiple comparisons across areas ($n_1 = 4$) and task phases ($n_2 = 4$); asterisks indicate significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

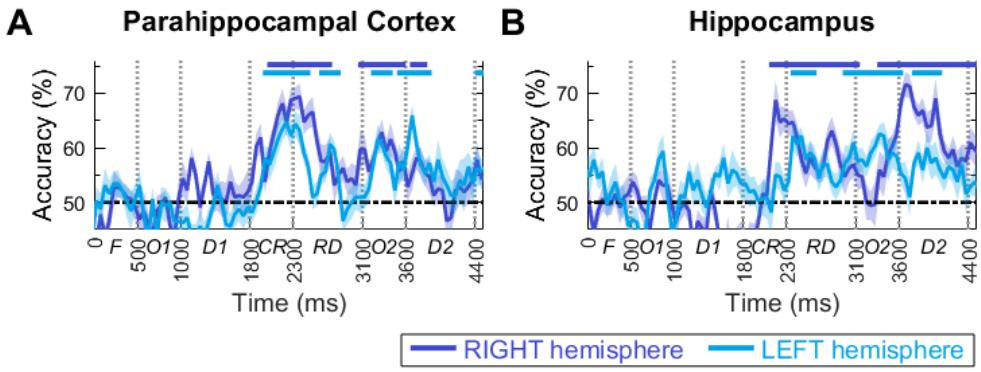


Figure S3: Rule Decoding in Different Hemispheres using an SVM Classifier. Related to Figure 4. Classification accuracy for decoding arithmetic rule information in **(A)** parahippocampal cortex: 67 units recorded from the right hemisphere, 59 units from the left hemisphere, and **(B)** hippocampus: 104 units recorded from the right hemisphere, 95 units from the left hemisphere. SVM classifiers were trained on the instantaneous firing rates across the trial period. The dashed line represents chance level (50 % for two classes). Light and dark blue bars above the data indicate significance ($p < 0.05$) in a permutation test for each hemisphere. Abbreviations at the axes indicate task phases: *F*, fixation; *O1*, operand 1; *D1*, delay 1; *CR*, calculation rule; *RD*, rule delay; *O2*, operand 2; *D2*, delay 2.

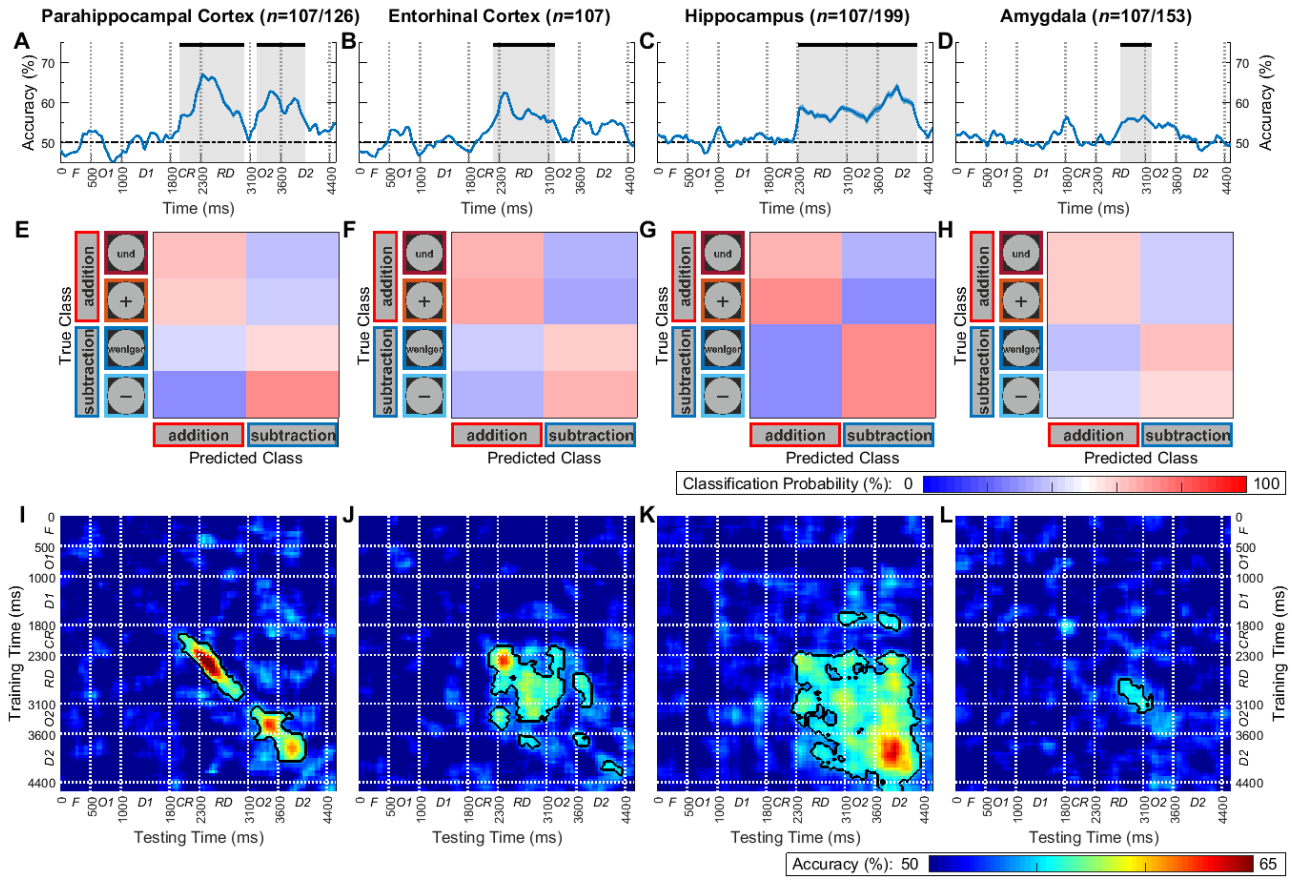


Figure S4: Rule Decoding in Sample-Equalized MTL Populations using an SVM Classifier. Related to Figure 4. Decoding performance when using random subsets of neurons per area, equalizing population size across all MTL regions (columns). **(A–D)** Average classification accuracy for decoding arithmetic rule information when training an SVM on the instantaneous firing rates across the trial period. The dashed line represents chance level (50 % for two classes). Black bars above the data and gray shaded areas indicate significance ($p < 0.05$) when testing against performance for SVMs trained on shuffled data in a permutation test. Abbreviations at the axes indicate task phases: *F,* fixation; *O1,* operand 1; *D1,* delay 1; *CR,* calculation rule; *RD,* rule delay; *O2,* operand 2; *D2,* delay 2. **(E–H)** Confusion matrix derived when training an SVM on firing rates averaged across the significant time window in (A–D), respectively. **E** shows the average of the confusion matrices obtained for each significant window (depicted in A). **(I–L)** Accuracy when training an SVM at a given time point of the trial and testing on another one (the main diagonals of the matrices correspond to the curves in A–D). Black contours indicate significance ($p < 0.05$) in a permutation test.

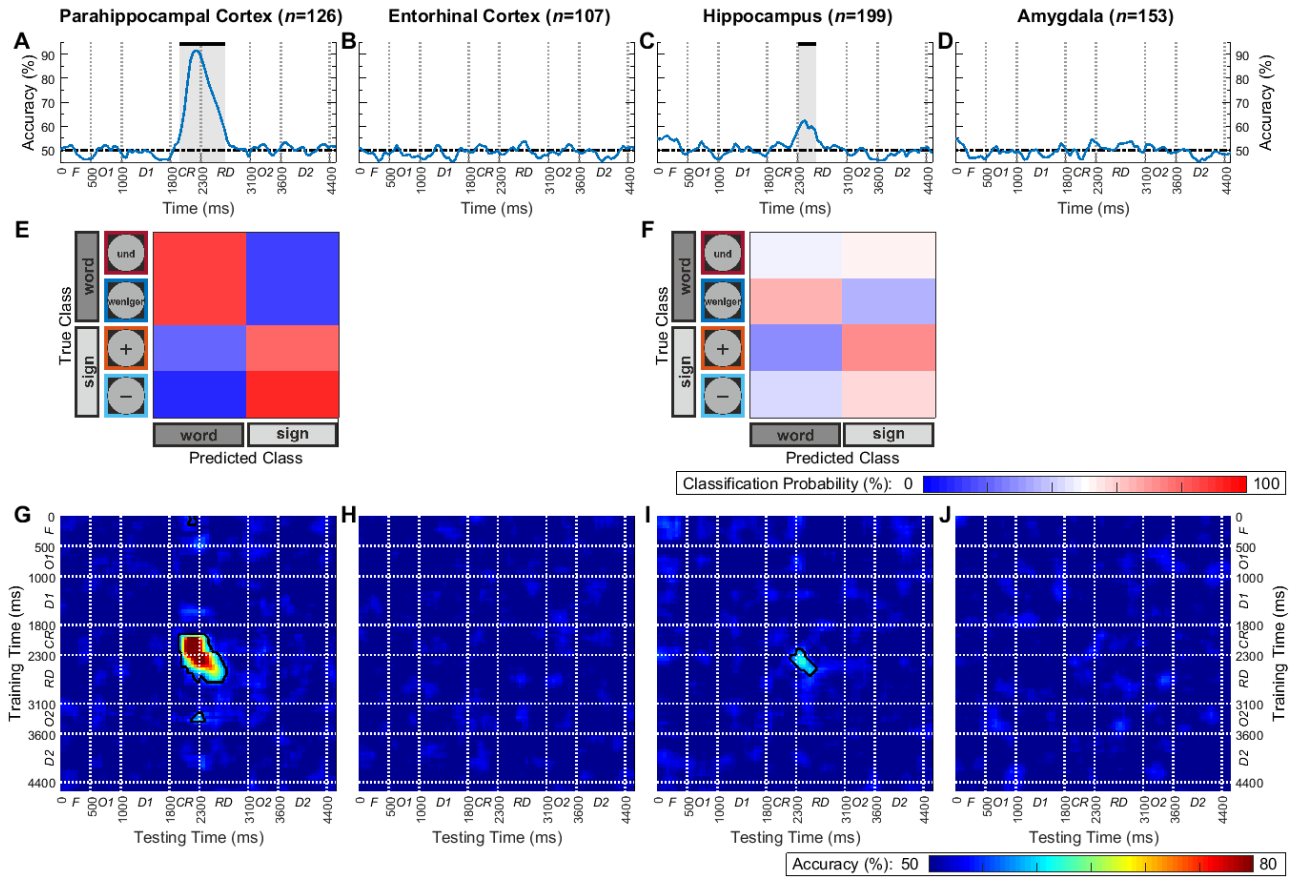


Figure S5: Rule Notation Decoding using an SVM Classifier. Related to Figure 4. Classifier performance for the four different MTL regions (columns). **(A–D)** Classification accuracy for decoding cue information when training an SVM classifier on the instantaneous firing rates across the trial period. The dashed line represents chance level (50 % for two classes). Black bars above the data indicate significance ($p < 0.05$) when testing against performance for SVM classifiers trained on shuffled data in a permutation test. Abbreviations at the axes indicate task phases: *F,* fixation; *O1,* operand 1; *D1,* delay 1; *CR,* calculation rule; *RD,* rule delay; *O2,* operand 2; *D2,* delay 2. **(E,F)** Confusion matrix derived when training an SVM on firing rates averaged across the significant time window in A and C, respectively. Significance was reached only in the PHC. **(G–J)** Accuracy when training an SVM classifier at a given time point of the trial and testing on another one (the main diagonals of the matrices correspond to the curves in A–D). Black contours indicate significance ($p < 0.05$) in a permutation test.