Highly task-specific and distributed neural connectivity in working memory revealed by single-trial decoding in mice and humans

5 Daniel Strahnen[1], Sampath K.T. Kapanaiah[1], Alexei M. Bygrave[2], Birgit Liss[1,3], David M.

6 Bannerman[4], Thomas Akam[4], Benjamin F. Grewe[5], Elizabeth L. Johnson[6,7], Dennis Kätzel[1,*]

7

8 [1]Institute of Applied Physiology, Ulm University, Ulm, Germany

9 [2] Department of Neuroscience, Johns Hopkins University, USA

10 [3]Linacre College and New College, University of Oxford, Oxford, UK

11 [4]Department of Experimental Psychology, University of Oxford, Oxford, UK

12 [5]Institute of Neuroinformatics, ETH Zürich, Zürich, Switzerland

13 [6]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA

14 [7]Life-Span Cognitive Neuroscience Program, Institute of Gerontology, Wayne State

15 University, Detroit, MI, USA

16 * Correspondence: dennis.kaetzel@uni-ulm.de; +49 731 500 33770; Fax +49 731 500

17 33779; Institute of Applied Physiology, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm,

18 Germany

19 Words: 5580 (excluding Abstract, Methods, References, Figure legends)

20 Figures: 8; Supplementary Figures: 14

21 Supplementary Tables: 6

Abstract

Working memory (WM), the capacity to briefly and intentionally maintain mental items, is key to successful goal-directed behaviour and impaired in a range of psychiatric disorders. To date, several brain regions, connections, and types of neural activity have been correlatively associated with WM performance. However, no unifying framework to integrate these findings exits, as the degree of their species- and task-specificity remains unclear. Here, we investigate WM correlates in three task paradigms each in mice and humans, with simultaneous multi-site electrophysiological recordings. We developed a machine learning-based approach to decode WM-mediated choices in individual trials across subjects from hundreds of electrophysiological measures of neural connectivity with up to 90% prediction accuracy. Relying on predictive power as indicator of correlates of psychological functions, we unveiled a large number of task phase-specific WM-related connectivity from analysis of predictor weights in an unbiased manner. Only a few common connectivity patterns emerged across tasks. In rodents, these were thalamus-prefrontal cortex delta- and beta-frequency connectivity during memory encoding and maintenance, respectively, and hippocampal-prefrontal delta- and theta-range coupling during retrieval, in rodents. In humans, task-independent WM correlates were exclusively in the gamma-band. Mostly, however, the predictive activity patterns were unexpectedly specific to each task and always widely distributed across brain regions. Our results suggest that individual tasks cannot be used to uncover generic physiological correlates of the psychological construct termed WM and call for a new conceptualization of this cognitive domain in translational psychiatry.

Introduction

Working memory (WM) refers to the capacity to store and manipulate contents of perception and thought at the forefront of attention over seconds to minutes[1,2]. Several psychiatric and neurological disorders are characterised by severe and pharmaco-resistant impairments of WM[3]. Studies in humans, non-human primates, and rodents[4] have centrally implicated the prefrontal cortex (PFC)[5–12], parietal cortex[11,13], medio-dorsal thalamus (MD)[14,15], and hippocampus[8,16–21] in WM processes. In particular, connectivity between these brain structures correlates with the performance in specific tasks[11,22–24] (see Supplementary Table 1 for an overview of rodent studies).

However, it remains elusive which kinds of neural connectivity actually mediate WM. Several - partially conflicting – reports suggest that specific brain regions, frequency bands[25–27] and types of inter-regional coupling (or metrics to analyse) are essential for WM. Studies in rodents, for example, have implicated network oscillations in either the delta (δ)[28], theta (θ)[27,29,30], beta (β)[14,26], or gamma (γ)[25] frequency, in coupling between such frequency bands (esp. $\theta-\gamma$)[31] or between oscillations and local spiking of neurons[14,27,31–33], or in interactions between either the ventral (vHC)[31,33] or dorsal hippocampus (dHC)[26,29,34] and PFC. On a T-maze rewarded-alternation task, both vHC→PFC connectivity[33] and dHC-PFC coupling via the thalamic nucleus reuniens[34] have been claimed to mediate the encoding of WM-contents using optogenetic inhibition experiments. In addition to this diversity of findings in rodents - which mainly originate from recordings in spatial alternation tasks (Supplementary Table 1) - it remains largely unknown to what extent the discovered WM substrates generalize across different tasks and species.

Moreover, it remains unclear if this diversity of reported associations between specific neural activities and WM performance implicates that, indeed, several types of connectivity and connections are engaged in WM simultaneously. Alternatively, it could be due to differences between task paradigms, species, specific rodent models, or analysis procedures. Several factors may explain the current uncertainty regarding the physiological correlates of WM.

3

71   This includes the difficulty of separating correlates of more basic functions like motivation,

72   spatial processing, and attention from the actual WM component of the task - especially in

73   rewarded-alternation assays that are mostly used in rodents (Supplementary Table 1)[35].

74   Further uncertainty arises from the typical analytical approach of correlating within-subject

75   averages of performance (WM accuracy score)[26,29,31], trial type (correct vs. incorrect)[33], or

76   trial phase (sample vs. choice phase)[30,33] with within-subject averages of a select

77   connectivity measure. Such correlations do not necessitate mechanistic causation but could

78   be indirect or even epiphenomenal. Finally, the *streetlight effect* has recently been

79   highlighted as a rather principal limitation of studies of physiological correlates of

80   psychological function[36]. In this context, this term denotes the missing of neurophysiological

81   correlates due to a biased selection of only few investigated brain regions, connections, or

82   activity parameters. We recently showed that the rich plethora of widely used functional

83   connectivity measures indeed contains a considerable number of non-redundant metrics[37],

84   which entails the possibility for seemingly contradictory findings and inconclusive null results.

85   To identify inter-regional neural connectivity that is associated with WM in mice and humans,

86   we addressed these issues in several ways: First, we comparatively assessed three distinct

87   visual WM tasks in each species. In mice, we applied an operant 5-choice delayed-

88   *matching*-to-sample (DMTS) task as well as a T-maze-based and an operant 2-choice

89   delayed-*non*-matching-to-sample (DNMTS) task. In humans, we analysed an existing

90   dataset where three distinct task types which featured the same temporal structure and were

91   inter-mixed within one session[20]. Therefore, in each species, the non-WM related differences

92   between the applied tasks were relatively small, which benefits an optimal comparison of the

93   neurophysiological basis of different types of WM. Second, during these tasks, local field

94   potentials (LFPs) were simultaneously recorded from four sites (PFC, dHC, vHC, and MD

95   thalamus) in rodents, and from three sites (PFC, medial temporal lobe, MTL, and

96   orbitofrontal cortex, OFC) in humans. Third, to maximize connectivity-related information

97   obtained from these multi-site recordings, we extracted a large set of measures of inter-

98   regional neural coupling. The composition of this set was optimized based on our prior

99   analysis of redundancies between the most commonly used LFP-based coupling metrics[37].

100   Each metric was determined in three or four task phases, four or five frequency bands, and

101   as both absolute and relative measures. This yielded a total of 960 and 1344 connectivity

102   measures in mice and humans, respectively, in addition to ~240 measures of local activity.

103   Finally and most importantly, we applied machine learning (ML) to predict WM choices on a

104   single-trial basis from these high-dimensional patterns of functional connectivity. Data-driven

105   approaches have recently been deployed to reduce the *streetlight effect*[36,38–40]. These

106   studies have applied *unsupervised* classification procedures to extract electrophysiological

107   correlates of stress responses[39], depression vulnerability[38], and anxiety[36] from connectivity

108   patterns obtained from multi-site LFP-recordings in mice, collectively termed the *electome*.

109   To establish a largely unbiased search for WM correlates within high-dimensional electome

110   patterns, in contrast, we translated the supervised, trial-based nature of WM tasks (pre-

111   defining choices as correct or incorrect) into our analysis[10,21,24,33,41,42]. Implementing a

112   *supervised* decoding analysis allowed us to harness the power of prediction to prove the

113   presence of behaviourally relevant information in complex patterns of neural signals[43]. The

114   possibility of trial-by-trial prediction of WM-based choices from the spiking activity of large

115   ensembles of neurons in rodent PFC[10,33], monkey inferior temporal cortex[41], or human

116   hippocampus[21] has recently been demonstrated. In these attempts, however, prediction was

117   done with classifiers that were trained specifically for a single animal and task session, since

118   the predictor variables were individual neurons. In contrast, we applied ML across datasets

119   from multiple subjects and sessions and used connectivity and activity metrics from multiple

120   sites as predictor variables. In this way, subsequent quantification of the predictive power of

121   each electrophysiological variable allowed a largely unbiased identification of connectivity

122   patterns associated with WM in mice and humans as verified by trial-by-trial predictive

123   power.

124

Results

Correct choices in DMTS working memory are associated with distinct signatures of connectivity

To evaluate associations between WM and electrophysiological measures of inter-regional connectivity, we first implanted mice with chronic field electrodes in PFC, MD, vHC, and dHC and tested them in three spatial working memory (SWM) tasks (Fig. 1a-f): first, an operant DMTS 5-choice SWM (5-CSWM) task, subsequently the T-maze rewarded alternation task, and finally an operant DNMTS 2-choice SWM (2-CSWM) task. To allow for simultaneous LFP recordings during operant tasks, we developed a custom-designed operant box optimized for implanted and tethered animals (Fig. 1c) that is tightly integrated with electrophysiological recordings via pyControl software and microcontroller modules[44]. The set of three tasks (Fig. 1d-f) was chosen to retain comparability due to their shared visuo-spatial nature and distinct individual differences (i.e., operant DMTS vs. DNMTS; maze-based DNMTS vs. operant DNMTS). In addition, the T-maze task was included due to its wide usage (see e.g., Supplementary Table 1). The 5-CSWM was specifically designed to limit the usage non-WM mediation strategies by the animal - due the large number of choice configurations, requirement to shuttle between opposite walls of the box, and delay-periods in total darkness, as we described in detail recently[45]. Also, both operant tasks provide tight control over the timing of behavioural events and deliver intrinsic control variables for WM-enabling psychological functions like attention, measured by sample phase (SP) accuracy, and motivational drive, measured by reward latency. In each task, we applied extensions of the delay between the SP and the choice phase (CP) across which the memorized information needed to be maintained in order to strongly engage WM capacity. As expected for WM assays[2], such delay challenges significantly decreased WM choice accuracy of the mice (Fig. 1g-i).
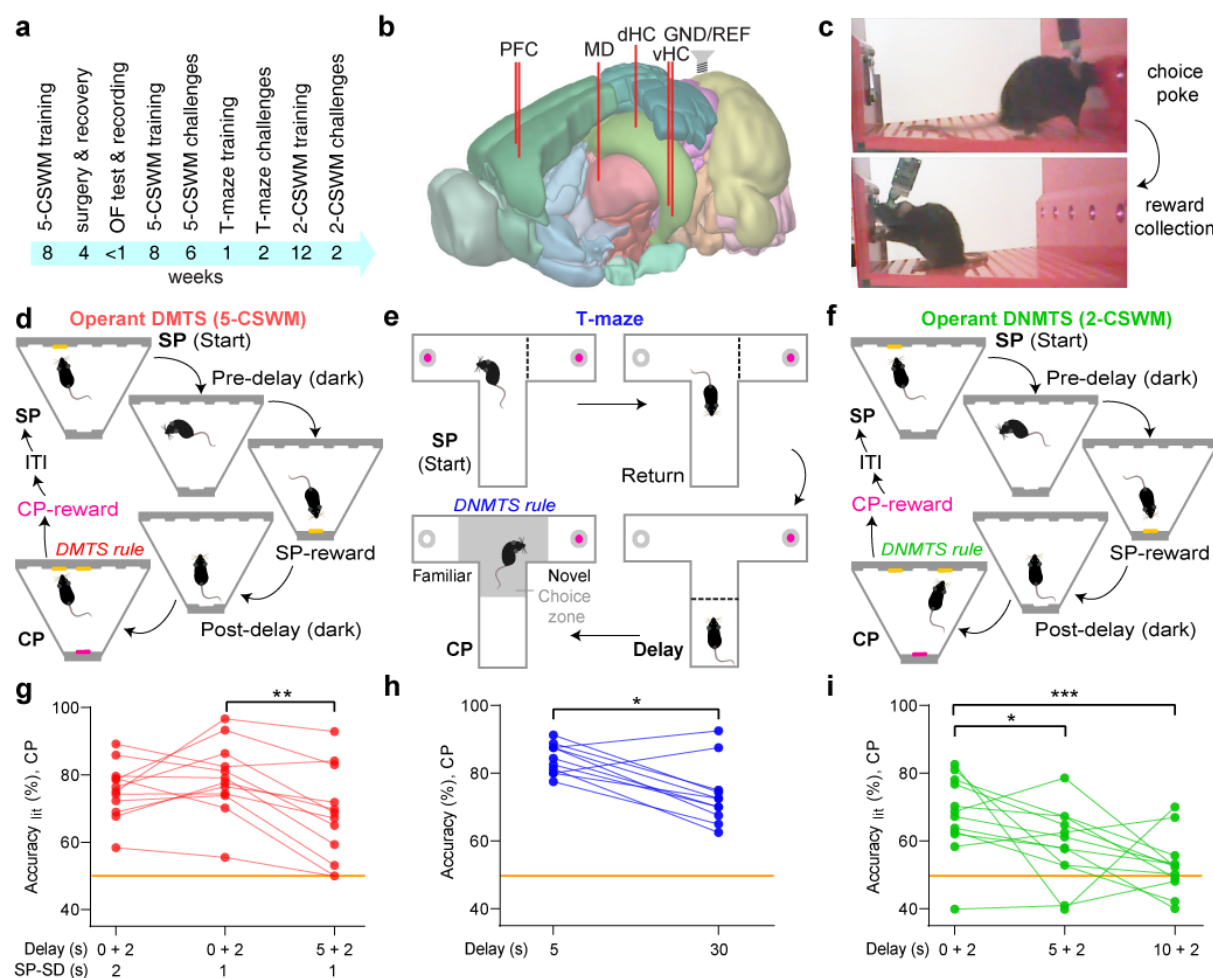
152

Fig. 1 | Rodent SWM tasks. (**a**) Timeline of experiments in the analysed cohort (see Methods). (**b**) Illustration of electrode placements in the mouse brain (image taken from Allen Brain Atlas); pairs of electrodes were inserted into PFC and vHC. (**c**) Lack of obstruction of tethered mice with mounted headstage during poking of choice hole (top) and reward collection (bottom) in custom-made pyControl operant boxes. (**d-f**) Illustration of DMTS 5-CSWM[45] (d), T-maze rewarded alternation (e), and DNMTS 2-CSWM (f) tasks; in (d,f), choices in SP and CP need to be made at the 5-choice wall (top), while rewards for correct responses in each phase are collected on the opposite wall (magenta, bottom). (**e-i**) WM performance measured as response accuracy in the CP (% correct choices relative to available indicated options) in 12 mice in each set of challenge conditions including their respective baseline with simultaneous electrophysiological recordings. Delay length determining the WM challenge stated on x-axes; for operant tasks (e,j) pre- + post-delay

165   (referring to set delays before and after SP-reward collection) are indicated. SP-SD, stimulus

166   duration in sample phase. Asterisks indicate differences between challenge, Sidak-post-hoc

167   tests conducted after significant main effect of challenge, RM-ANOVA. Orange line, chance

168   level performance. * $P < 0.05$; ** $P < 0.01$; *** $P \le 0.001$.

169

170   To initially explore possible WM correlates among measures of neural connectivity, we

171   computed time-resolved spectrograms of four largely non-redundant[37] connectivity metrics

172   aligned to the time of correct and incorrect SP and CP responses for the distinct phases of

173   the 5-CSWM task (*non-directed*: coherence, Coh; weighted Phase-Lag-index, wPLI;

174   *directed*: Granger causality, GC; partial directed coherence, PDC; Supplementary Fig. 1-2).

175   We subtracted spectra of correct SP responses or incorrect CP responses from those of

176   correct CP responses to eliminate neural representations of poking action, execution of a

177   reward-related response, and attention (Fig. 2a-d, Supplementary Fig. 3). This qualitative

178   analysis suggested a complex pattern of connectivity associated with correct WM decisions,

179   including elevated hippocampal-prefrontal connectivity in the low γ-range (30-48 Hz)

180   immediately after the CP response (Fig. 2a-b) and sustained connectivity in the PFC→MD

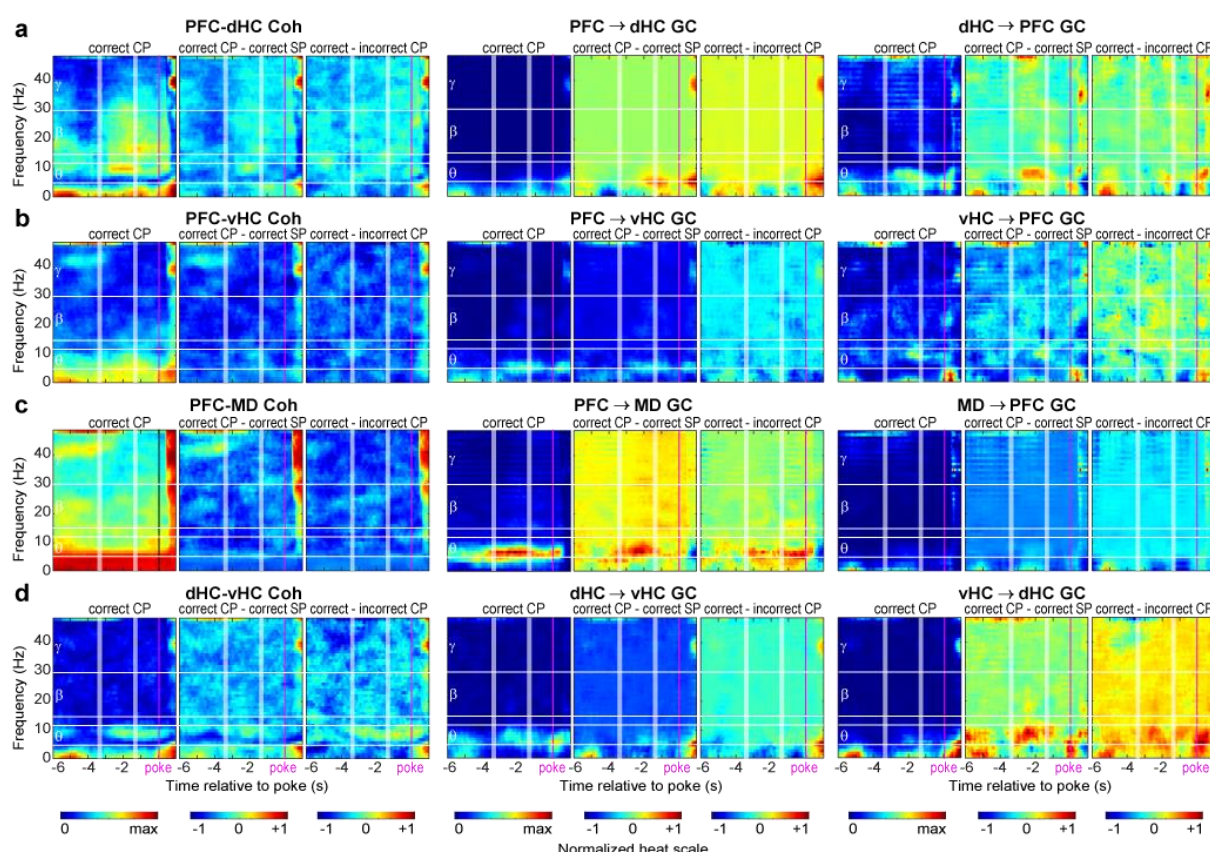181   and vHC→dHC θ-range (5-12 Hz) during the delay (Fig. 2c-d).

Fig. 2 | Connectivity in the DMTS 5-CSWM task. (**a-d**) Spectrograms depicting min-max normalized coherence (Coh) and Granger causality (GC) for the connections stated above each triplet panel for the delay and CP of the 5-CSWM task, temporally aligned to the choice poke entry (p, vertical magenta lines) showing 6 s before until 1 s after the choice poke (x-axes in (d)); the start and end of the post-delay shown by white stripes corresponding to mean±SD as determined by CP response latency. Each triplet shows the absolute value (left), the difference between the former and either the prior correct SP (middle), or incorrect CPs (right). Horizontal white lines show borders between analysed frequency bands, stated on the left. See Supplementary Fig. 1-2 for spectrograms of absolute values in correct and incorrect trials of the CP and SP, respectively, and Supplementary Fig. 3 for the same display as (a-d) for wPLI and PDC.

Trial-by-trial prediction of DMTS WM-mediated choices from local and long-range neural activity

The ability to predict behavioural choices from neural activity may be regarded as evidence that such activity encodes aspects of these choices[33,43]. We assessed individual connectivity variables as five separate connectivity metrics (Coh, wPLI, GC, PDC, and cross-regional $\theta-\gamma$ phase-amplitude coupling, PAC) in four task phases (SP, pre-reward delay, post-reward delay, CP; see Fig. 1d), and four frequency bands (δ, 1-4 Hz, θ, 5-12 Hz, β, 15-30 Hz, low-γ, 30-48 Hz) along 4 connections (as shown in Fig. 2a-d). In the same way, indicators of local activity (power and local $\theta-\gamma$ PAC) were calculated within the four involved regions (dHC, vHC, PFC, MD). Additionally, all metrics were calculated in *relative* terms by dividing their obtained value by the value of that specific metric during the inter-trial-interval (ITI) before the start of the respective trial. This resulted in 240 variables for each inter-regional connection and 56 intra-regional activity variables for each area characterizing every 5-CSWM trial. We trained subspace discriminant classifiers – which proved superior among 25 different types of linear and non-linear classifiers (Supplementary Fig. 4) - to predict WM-choice trial-by-trial using the parameters contributed by each connection or region separately for trials of the final 5-CSWM challenge (1 s SP stimulus duration, SP-SD, 5 s delay; Fig. 3a). A single decoding model was generated across all subjects and its decoding accuracy was determined by predicting trials that were not part of the training dataset, i.e. by cross-validation.

We found that individual 5-CSWM choices could be decoded with 79.4% and 79.8% maximum average accuracy when using measures of neural connectivity along the PFC-MD or the PFC-dHC connection as predictors, respectively (Fig. 3b). Using one-way ANOVA and pairwise Tukey post-hoc tests, we established a hierarchy between decoding accuracies obtained from each of the four connections and four regions. Decoding accuracies obtained from PFC-MD, PFC-dHC, local dHC and MD activity did not differ from each other and were superior to the remainder (Fig. 3b). Even though decoding accuracies varied by connection

10

222   and region, they were always significantly better than those of control classifiers trained with

223   shuffled labels ($P < 10^{-17}$, t-tests) which, in turn, decoded indistinguishable from the 50%

224   chance level on average (Fig. 3b). To evaluate the generality of the obtained classifiers, we

225   assessed if they could also decode WM-based choices in data from other 5-CSWM

226   challenge protocols. Even though the resulting decoding accuracies were generally lower

227   compared to those achieved with data from the same protocol, they were still significantly

228   higher than those of classifiers trained with shuffled labels (Supplementary Fig. 5). These

229   analyses reveal that WM-based choice is encoded in LFP-based connectivity and activity

230   measures in individual trials and that such information is widely distributed across multiple

231   brain regions.

232

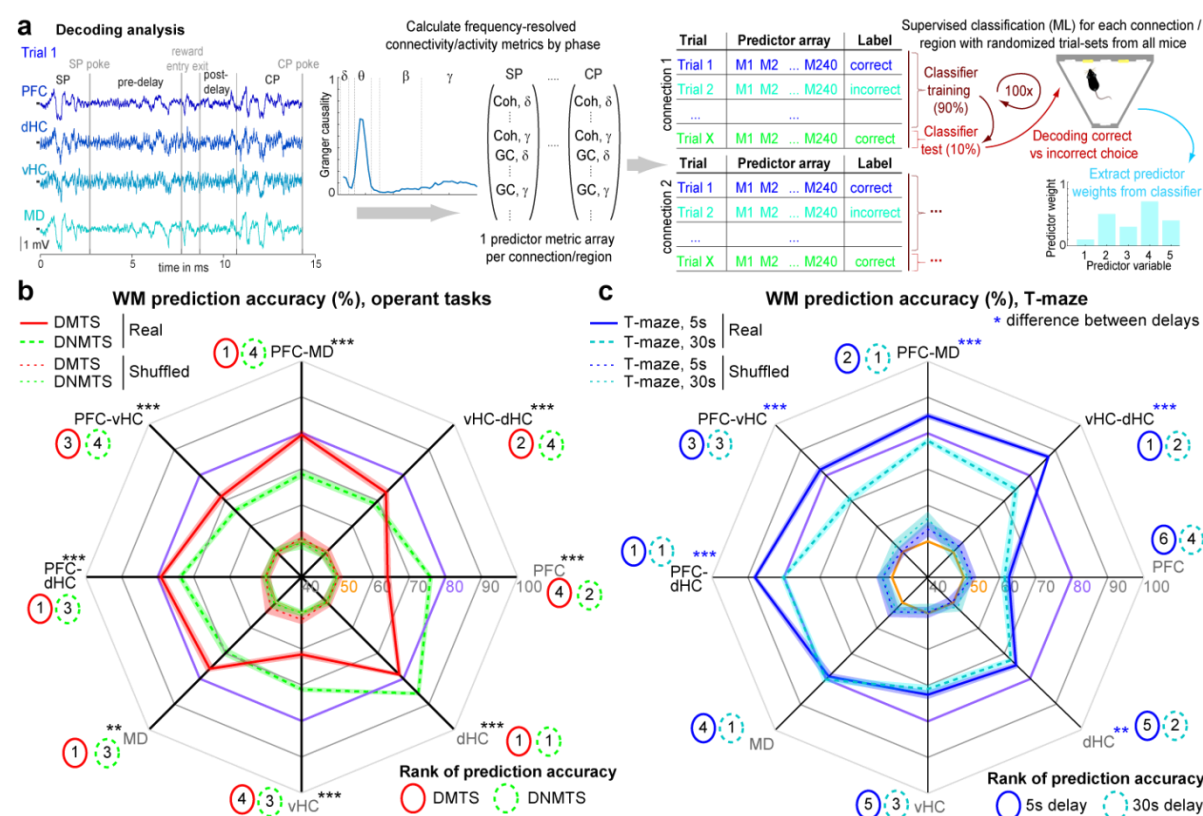233   **Fig. 3 | Trial-by-trial decoding of WM-based choice.** (**a**) Illustration of ML-based decoding

234   analysis (see Methods). (**b-c**) Cross-subject decoding accuracies achieved on average when

235   using connectivity or local activity parameters of the indicated individual connections (black)

236   or areas (grey), respectively to predict WM-based *correct* vs. *incorrect* choices in the DMTS

237   5-CSWM task (combined 1 s SP-SD, 5+2 s delay challenge, 2 sessions; red, b), the DNMTS

238   2-CSWM task (baseline, 2 sessions, green, b), the T-maze WM task with either 5s (solid

239   blue, 4 sessions, c) or 30s (dashed blue, 4 sessions, c). Thinner dotted lines show decoding

240   accuracies of corresponding classifiers trained with shuffled labels, remaining at chance

241   level (50%, orange). Classifiers trained with real labels perform better than those trained with

242   shuffled labels in all cases ($P < 10^{-17}$, t-tests, not indicated). The accuracy of 80% is coloured

243   in purple to aid comparison. Numbers in coloured ovals indicate the rank of the prediction

244   accuracies achieved on average by using data from the respective connection or region.

245   Ranks have been generated from pairwise comparisons with Tukey post-doc tests

246   conducted after significant effects of connection/region in one-way ANOVAs ($P < 0.0001$ in

247   all cases); connections/regions that were not significantly different from each other were

248   assigned the same rank. Black stars in (b) indicate differences of accuracy values achieved

249   in the two operant tasks (Tukey post-hoc tests after significant effect of task-type in ANOVA

250   across all three tasks). Blue stars in (c) indicate pairwise differences between the two delays

251   (uncorrected t-tests). ** $P < 0.01$; *** $P \leq 0.001$. Shaded regions around mean show s.e.m.

252   across 100 classifiers generated for each task and connection or region.

253   **Specific connections and regions are engaged differently in distinct rodent**
254   **WM tasks**

255   To investigate if this conclusion applies generally to rodent WM, we repeated the same

256   analysis for the operant DNMTS data (final baseline sessions, 2 s delay). In this case,

257   however, we obtained the maximum average prediction accuracy (86.1 %) from local dHC

258   activity, rather than PFC-MD (66.1%, lowest rank of all classifiers) or PFC-dHC (73.5%)

259   connections. Generally, in this task, local activities allowed relatively high decoding

260   accuracies (72-77% for PFC, MD, and vHC), while coupling metrics were significantly less

261   predictive (66-68%, $P < 0.001$, Tukey; except for dHC-PFC, Fig. 3b).
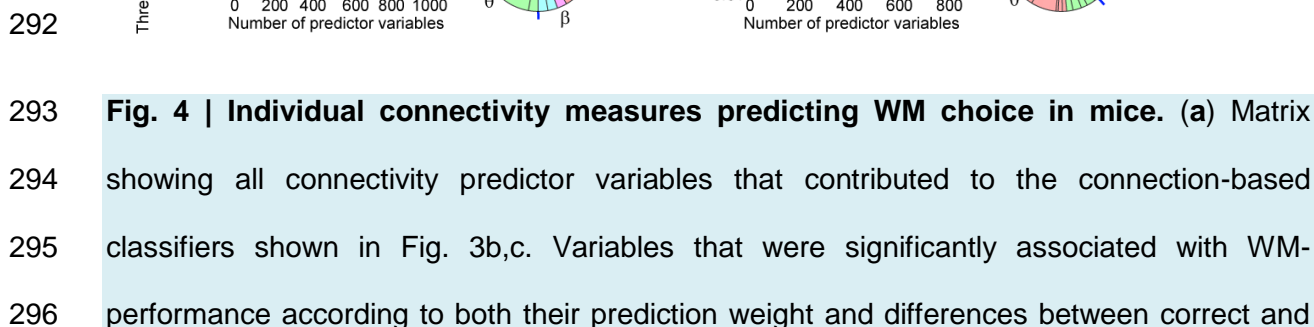
262   In reverse, trial-by-trial decoding of T-maze data achieved the highest average accuracies

263   (82-88%) when using connectivity data from either one of the four connections (with dHC-

264 connections being most predictive), whereas local activities were significantly less predictive

265 (62-79%; $P < 0.001$, Tukey, Fig. 3c). However, decoding accuracies for information from all

266 4 connections decreased when analysing data from the 30 s delay challenge, in which these

267 animals also showed lower behavioural performance (Fig. 1h, 3c), suggesting that not only

268 task type but also task difficulty affect the information encoded in each connection.

269 **Specific connections and regions are engaged differently in distinct phases of**
270 **the rodent DMTS 5-CSWM task**

271 The prior analyses entail at least two conclusions: Firstly, WM-related information in a single

272 trial is not encoded in any single region or connection, although some of them bear higher

273 predictive power regarding WM-choice than others. Secondly, the predictive power of a

274 given region or connection is not uniform but strongly depends on the type of WM task,

275 indicating that different mechanisms and regions are engaged to solve distinct behavioural

276 demands. These conclusions re-emphasize the question as to what extent oscillatory

277 processes in distinct frequency bands, of a distinct biological type, or in a specific task phase

278 (encoding, delay, choice) can be regarded as correlates of WM (Supplementary Table 1).

279 To answer this question, we took advantage of the fact that a linear classifier reveals the

280 predictive power of each involved predictor variable according to its assigned weight. We

281 performed Bonferroni-adjusted *t*-tests comparing the weights for each connectivity variable

282 with the weights assigned by the classifiers trained on label-shuffled control data, and,

283 additionally, conducted *t*-tests comparing the amplitudes of each variable between correct

284 and incorrect trials that contributed to the classifiers. Variables for which both *t*-tests were

285 significant were considered as bearing WM-related information (indicated by colour in Fig.

286 4a). This analysis revealed a relatively small set of consistent WM-related feature-classes as

287 correlates of DMTS 5-CSWM, a majority of them in the γ-range (Fig. 4b): (*1*) PFC→MD δ-

288 and γ-range connectivity in the *SP*, (*2*) MD→PFC β- and γ-range as well as dHC-vHC

289 θ−range connectivity in the *delay*, (*3*) dHC→PFC and MD-PFC γ-range as well as vHC-PFC

290   δ−θ-range coupling in the *CP,* and (*4*) intra-hippocampal δ-connectivity in *all phases* (Fig. 4c-

291   d; see also 5c).

Fig. 4 | Individual connectivity measures predicting WM choice in mice. (a) Matrix showing all connectivity predictor variables that contributed to the connection-based classifiers shown in Fig. 3b,c. Variables that were significantly associated with WM-performance according to both their prediction weight and differences between correct and

14

297   incorrect CP (*t*-tests, Bonferroni-adjusted for total number of variables, see (b)) in any of the

298   three WM tasks are indicated by the corresponding colour, remainder in grey (white squares

299   have no corresponding variable). Variables from the pre- and post-delay in the 5-CSWM are

300   combined in single lines. For θ, mean amplitude (m), peak amplitude (p) and frequency of

301   peak (f) are shown, while for all other frequency bands only the mean amplitude is used due

302   to the absence of a clear singular peak. At the bottom, all task phases are combined and

303   only connectivity metrics that are predictive in all three tasks (in at least one phase) are

304   indicated in black. (**b**) Share of each connection (coded by colour) and frequency band

305   (stated around pie chart with separations in blue) among all significant predictors (*N* stated in

306   centre) for the indicated task. (**c-d**) Values of absolute (c) or relative (d) predictor variables

307   for DMTS WM (extracted from (a)) in correct (blue) and incorrect (red) trials (left axes). Bars

308   in the background, referenced to by right axis, show absolute values of average predictor

309   weights normalized within each classifier (i.e., connection) coded by their colour.

310   Significance is not indicated as it applies to all shown variables. Blue font, directional

311   connectivity in the *opposite* direction compared to connection-name. Error bars, s.e.m. (**e, f**)

312   Left: Number of variables identified as significant based on ML predictor weight (light green),

313   difference between correct and incorrect trials (medium green), or both (dark green) in

314   dependence on the *P*-value adjustment (y-axes) in the DNMTS (e) and T-maze (f) tasks.

315   Right: Share of predictor variables as depicted in (b) for the *P*-levels indicated in green.

316

317   Given the prominence and high predictor weights of *CP* parameters (Fig. 4c-d) - which align

318   with the arising γ-band connectivity immediately after the CP-poke (Fig. 2a-d) - we wondered

319   if the predictability of WM-choice (Fig. 3b) actually relied mainly on identifying a

320   representation of anticipated reward. Therefore, we replicated the decoding analysis for *SP*

321   choices for which animals also expect reward. For the SP, however, average decoding

322   accuracies were – although still above the 50% chance level - considerably smaller, namely

323   64-68 % and 54-59 % for predictions based on connectivity and local activity, respectively

15

324   (Supplementary Fig. 6a). While this result shows that the attentional element of the task is

325   more difficult to predict from the available parameters than WM choice, it also demonstrates

326   that the obtained *CP* prediction accuracy was not simply based on representations of motor-

327   action (hole-poking), attention, or reward anticipation. We also repeated the decoding

328   analysis for CP choice with complete omission of all CP parameters. Even though average

329   decoding accuracies decreased significantly for some connections, including the most

330   predictive ones (PFC-dHC, 72%, PFC-MD, 73.9%) - but not for vHC-dHC (72%) - overall

331   accuracies remained far above those obtained from classifiers trained on shuffled labels,

332   and hence above chance level ($P < 10^{-30}$ and $P < 0.002$ for classifiers trained on connectivity

333   or local data, respectively; Supplementary Fig. 6b). Overall, these analyses demonstrate that

334   activities along distinct connections and in distinct frequency bands represent encoding (SP),

335   maintenance (delay), and recall (CP) of WM contents in the 5-CSWM task.

WM-related functional connectivity is highly task-specific in mice

337   To investigate if such phase-specific connectivity generalizes across tasks, we performed

338   the same analysis for the classifiers predicting performance in the operant DNMTS and the

339   T-maze assays. In both cases, considerably more parameters carried WM-related

340   information than in the 5-CSWM task (Fig. 4a-b). Compared to DMTS WM, T-maze

341   rewarded alternation choice was predicted by a much larger number of predictors, with a

342   prominence of θ- and β-range (as opposed to γ-range) variables, and a considerable

343   proportion of SP-parameters (Fig. 4a-b). This pattern is reflective of the diversity of

344   connectivity parameters that have been associated with T-maze performance in prior studies

345   (Supplementary Table 1), and the relatively detrimental effects of optogenetic manipulations

346   in the SP and delay, as compared to the CP[33,34,46].

347   Most astonishingly, the combined analysis of all three tasks revealed that *none* of the

348   specific connectivity parameters identified in *one* task bore significant predictive power in the

349   other two tasks, revealing a remarkable task-specificity of such parameters (Fig. 4a). It is,

350   however, possible that this finding is simply caused by a very conservative Bonferroni-

16

351 adjustment of the *P*-value used as significance threshold (0.05/number of *all* connectivity

352 and activity variables combined; 0.05/1184 for the 5-CSWM, 0.05/888 for the T-maze and 2-

353 CSWM). To test this possibility, we repeated the above analysis while relaxing this

354 adjustment incrementally over four orders of magnitude (Fig. 4e-f). However, the number of

355 identified significant parameters, the relative contribution of different frequency bands, and

356 especially the extreme sparseness of overlap between task-specific predictors changed

357 relatively little (Fig. 4e-f, Supplementary Fig. 7-8). This analysis also revealed that far more

358 connectivity parameters are identified according to their prediction weight than according to

359 their amplitude difference between correct and incorrect trials (Fig. 4e-f). This suggests that

360 the classical approach of correlating behavioural performance with the amplitude of a given

361 metric (Supplementary Table 1) likely misses a sizable proportion of WM-related functional

362 connectivity.

363 To investigate potential differences or similarities between the time-course of individual

364 parameters during the task, we extracted those spectral connectivity parameters that were

365 predictive across all three assays albeit in different phases: directed dHC→PFC γ-

366 connectivity and intra-hippocampal δ-range coupling (Fig. 4a). Inspection of the time-course

367 of these parameters over the delay and CP revealed that they behaved rather differently in

368 the individual tasks: dHC→PFC γ-connectivity showed a transient increase during the delay

369 of all three tasks, but only in the operant tasks a second increase occurred immediately after

370 correct choices (but not after incorrect choices; Fig. 5a-b). Intra-hippocampal δ-coupling

371 even showed a different time course in every task, including a correct choice-specific

372 decrease in the 2-CSWM delay which contrasted sharply with a steady rise during the T-

373 maze delay (Fig. 5a-b). Thus, even within the few predictor variables that are relevant across

374 all tasks, the actual physiological activity relating to the behaviour differed markedly.

375 Given these results, we directly tested the hypothesis that distinct activity patterns underlie

376 the different rodent WM tasks by rendering *task-type* a dependent variable: we trained

377 classifiers to decode which one of the three tasks a subject is currently conducting using

378   connectivity or local activity parameters from correct trials as input. Based on connectivity

379   data, task-type could be decoded with average accuracies of 97-99% when discriminating

380   between the two operant tasks (50% chance level) and with an accuracy >90-95% when

381   discriminating between all three tasks simultaneously (33.3% chance level; Supplementary
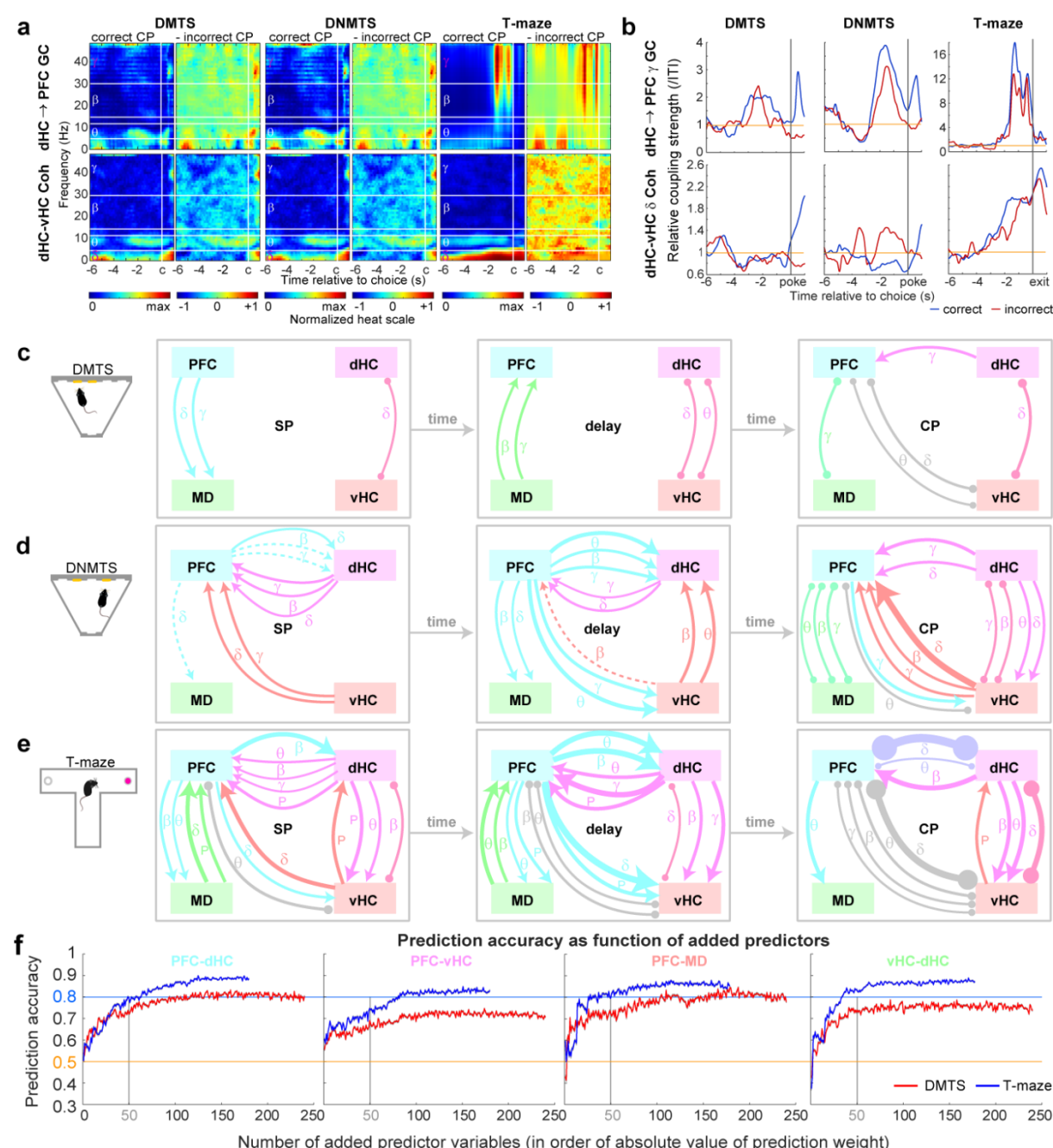
382   Fig. 9).

383

384   **Fig. 5 | Individual connectivity measures predicting WM choice in all tasks in ice.** (a)

385   Spectrograms for metrics (stated on the left) that were predictive in all three tasks (albeit in

386   different phases; named at the top) are shown as absolute values and as difference between

387   correct and incorrect CP responses, -6s until +1s around the choice; relevant frequency

388   bands are γ (top) and δ (bottom). (**b**) Average temporal evolution of those metrics in units of

389   their value during the preceding ITI aligned to the choice point in each task (vertical grey

390   line, poke or exit from decision zone) during correct (blue) and incorrect (red) trials. (**c-e**)

391   Depictions of directed connectivity (arrows, derived from GC or PDC) or non-directed

392   coupling (round-ended arcs, coherence or wPLI) during the three phases of each task

393   identified on the left, derived from significant predictor variables shown in Fig. 4a. For

394   connections where there is a directed connectivity metric of the same frequency, a non-

395   directed metric in the same frequency band is not depicted. Line weights indicate the change

396   in connectivity amplitude in the stated phase relative to the preceding ITI (irrespective if the

397   indicated metric was significant as absolute or relative measure). Measures that are

398   significant but where the amplitude in the indicated phase is identical or smaller than in the

399   ITI are shown as dotted lines. P, θ−γ-PAC. (**f**) Average decoding accuracies obtained with

400   classifiers calculated with a reduced number of predictor variables are shown as a function

401   of the number of added predictors, whereby the addition was done in order of normalized

402   prediction weight obtained with all variables (Fig. 3b,c) for DMTS (red) and T-maze (blue)

403   and the named connections. Chance level (50%) and 80% accuracy are indicated by

404   coloured lines. Shaded area, s.e.m.

Common WM-related connectivity patterns shared across rodent tasks

406   To extract commonalities of connectivity between the three tasks, we aggregated predictive

407   non-directed (coherence, wPLI) and directed (GC, PDC) metrics (extracted from Fig. 4a) and

408   depicted their amplitude increases relative to the preceding ITI for each task phase (Fig. 5c-

409   e). For the T-maze (the only task for which prior reference data exists), this revealed several

410   connectivity patterns associated before with rewarded alternation performance, including

411   vHC-PFC[33] and dHC-PFC[34] coupling during encoding and MD→PFC β-range activity during

412   maintenance across the delay[14,46]. Importantly, MD-PFC β-range delay activity was also

413   seen in the other two WM tasks, although their directionality differed (MD→PFC in the DMTS

414   task; PFC→MD in the DNMTS task; Fig. 5c-e). Likewise, further task-independent

415   connectivity patterns emerged in this analysis: prominent vHC-PFC $\delta/\theta-$coupling and vHC-

416   dHC $\delta-$coupling in the CP, and MD-PFC δ coupling in the SP (but with task-dependent

417   directionality; Fig. 5c-e). Some further patterns were shared only by the two DNMTS tasks,

418   e.g., directed PFC-dHC β- and γ-connectivities in the SP and delay (Fig. 5d-e). At the same

419   time, this analysis also confirmed that the vast majority of WM-related connectivity was task-

420   specific, especially when comparing the 5-CSWM DMTS to the other two tasks (Fig. 5c-e).

421   An important aspect of this analytical approach is that none of these individually highlighted

422   connectivity measures (Fig. 4a, 5c-e) is particularly predictive on its own: When performing

423   decoding analysis with reduced sets of predictor variables – starting with the parameter with

424   the single highest weight and adding variables incrementally – the inclusion of several dozen

425   predictor variables was necessary to achieve maximum decoding accuracy (Fig. 5f).

426   **Predictive power of local activity in a single area varies by task phase and type**

427   *Local* oscillatory activity in the four analysed regions also allowed considerable prediction

428   accuracy in all three tasks - sometimes even exceeding that obtained from connectivity

429   metrics (Fig. 3b-c). Therefore, to reveal WM-related local activity metrics, we repeated the

430   prior weight-based analysis for the respective variables (power, local PAC). In the 5-CSWM

431   DMTS task only CP parameters, mostly in the β/γ-range, were significantly associated with

432   WM (Fig. 6a). For the two other tasks, in contrast, significant predictors came from all three

433   phases and were somewhat less frequency-specific; the power of dHC-oscillations across all

434   frequency bands and phases constituted the most prominent cluster of choice-predictors in

435   both assays (Fig. 6a). In agreement with the high decoding accuracy obtained with local

436   activity (as opposed to connectivity) in the DNMTS 2-CSWM (Fig. 3b), many more significant

437   local predictor variables were found for this task compared to the other two, irrespective of

438   *P*-value threshold (Fig. 6b). Importantly, however, there was again hardly any overlap

439   between significant predictors from the three tasks. Also, while in DMTS WM all predictive

440 activity parameters had *higher* amplitudes in correct trials compared to incorrect trials, this

441 was not the case for the T-maze, where virtually all predictive hippocampal activity was

442 *lower* in correct trials compared to incorrect trials – only PFC and MD power were higher in

443 correct trials (Fig. 6c-d). Hence, as observed in inter-regional connectivity, local activities

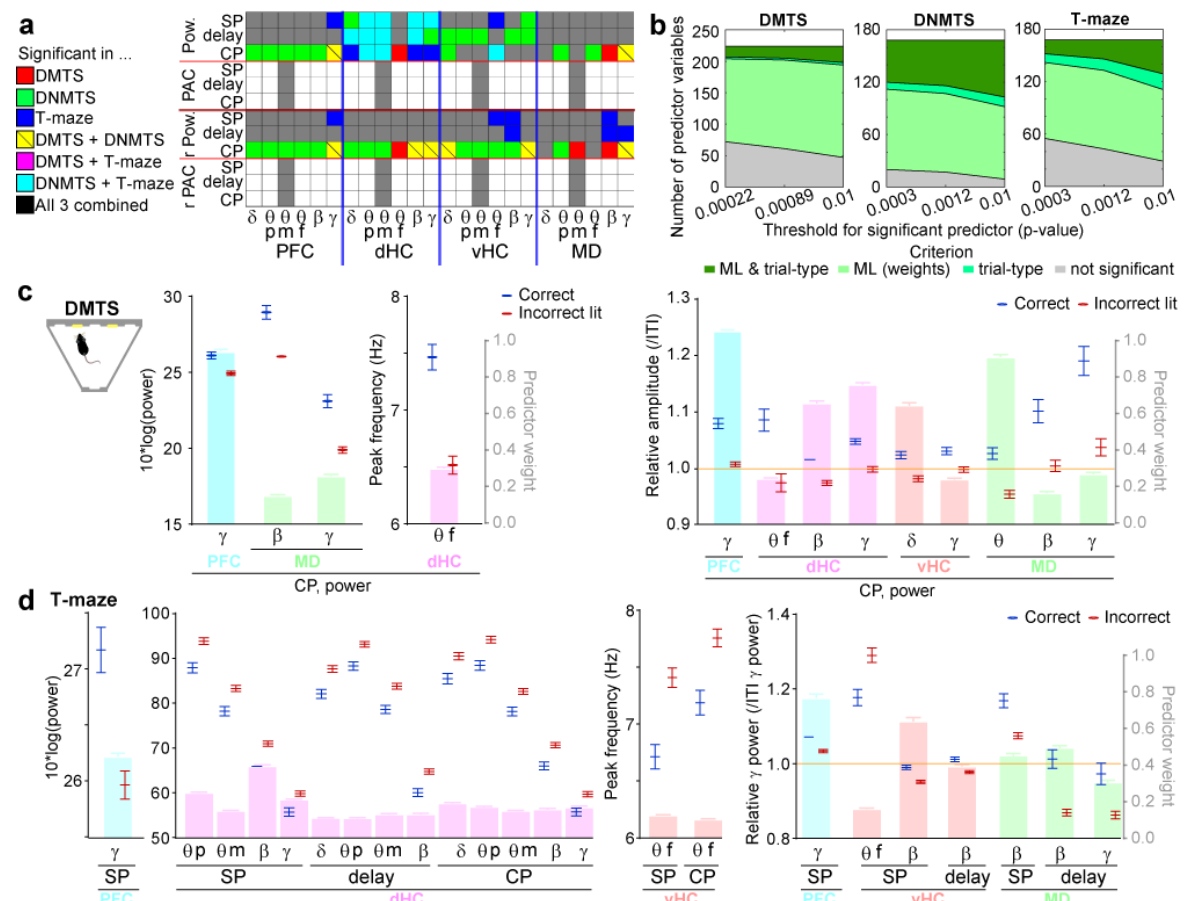444 related to WM-choice were highly task-specific in multiple respects.

445

Fig. 6 | Local activity measures predicting WM choice in all tasks. (**a**) Display and
analysis as in Fig. 4a, but for all local activity parameters that contributed to the region-
based classifiers shown in Fig. 3b,c. Variables that were significantly associated with WM-
performance according to both their prediction weight and differences between correct and
incorrect CP (*t*-tests, Bonferroni-adjusted for total number of variables) in any of the three
WM tasks are indicated by the corresponding colour, remainder in grey (white squares have
no corresponding variable). (**b**) Number of variables identified as significant based on ML
predictor weight (light green), difference between correct and incorrect trials (medium

21

454 green), or both (dark green) in dependence on the *P*-value adjustment (x-axes) in the named

455 tasks. (**d-e**) Values of absolute (left) or relative (right) predictor variables for DMTS (e) or T-

456 maze (f) WM (extracted from (a)) in correct (blue) and incorrect (red) trials (left axes). Bars in

457 the background, referenced to by right axis, show absolute values of average predictor

458 weights normalized within each classifier (i.e., connection) coded by their colour.

459 Significance is not indicated as it applies to all shown variables. Error bars, s.e.m.

Trial-by-trial decoding of WM-mediated choices from local and long-range neural activity in humans

462 It remains unclear if highly task-specific and widely distributed WM correlates are only found

463 in rodents or also in human WM. To clarify this question, we used a dataset of intracranial

464 LFP (iEEG) recordings made in 8 human subjects from three sites - PFC, OFC, and MTL –

465 during three types of WM assays whose trials were intermixed within a single test session:

466 identity-related WM (differentiating between identical and novel shapes), spatial WM, and

467 temporal WM (remembering the temporal order of two stimuli; Fig. 7a-b)[20]. For each of the

468 three tasks, we applied the same ML-approach as in mice, generating classifiers that use

469 activity data from four phases (SP, pre-cue- and post-cue delay phases, CP; see task

470 schedule in Fig. 7a) from only a single connection or region at a time.

471 Average decoding accuracies for trial-by-trial prediction of WM-choices were higher than

472 those achieved in mice, ranging consistently between 87-90% for predictions based on

473 connectivity and between 72-82% for predictions based on local activity, whereas

474 "predictions" based on shuffled control data remained significantly lower ($P < 10^{-40}$, *t*-tests)

475 and were not different from chance level (50%; Fig. 7c). We also trained classifiers on the

476 combined data from all three inter-regional connections and three regions – either separately

477 for each task-type or combining all types of trials indiscriminately. For task-specific

478 classifiers, average decoding accuracies reached 87.6%, 90.8%, and 89.8% for identity-

479 related, spatial, and temporal WM, respectively, i.e., no higher than what could be achieved

480 by connectivity data from the single best connection in each task (Fig. 7d). However,

481   decoding accuracy dropped to 79.4% if task-paradigms were intermixed (Fig. 7e) suggesting

482   that functional connectivity is, at least partially, task-specific. Task-specific prediction

483   accuracies of up to 91% could also be obtained without including CP connectivity measures

484   (Supplementary Fig. 10). Furthermore, in two cases, an average prediction accuracy of up to

485   81% could even be achieved if using connectivity data from only a single task phase – either

486   the SP in temporal WM or the post-cue delay in spatial WM (Fig. 7f; Supplementary Fig. 10).

487   Strikingly, in both cases, prediction accuracies – and hence information contents - of all

488   three connections were always similar to each other, suggesting a broad presence of WM-

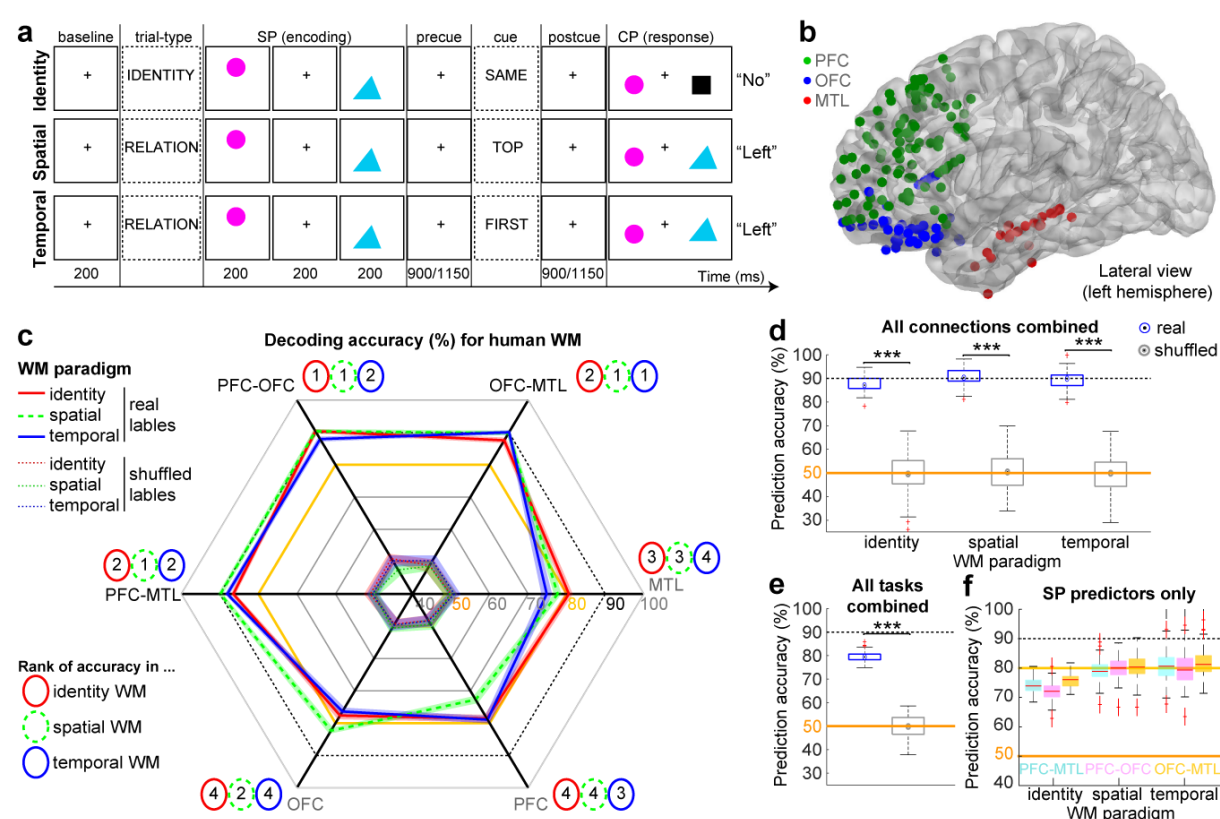489   related neural substrates across the brain.

490

491   **Fig. 7 | Single-trial based prediction of WM choice in humans.** (**a**) Display of the

492   structure of the three human WM tasks, according to ref. [20] (see Methods). (**b**) Placements

493   of electrodes as projected onto the left hemisphere, irrespective of actual hemisphere of

494   each electrode. (**c**) Cross-subject decoding accuracies achieved on average when using

495   connectivity or local activity parameters of the indicated connections (black) or areas (grey),

496 respectively, to predict WM-based *correct* vs. *incorrect* choices in the tasks coded by colour

497 on the left. Accuracies achieved by classifiers trained with randomly shuffled labels are

498 shown as coloured dotted lines; they are consistently lower than accuracies achieved by

499 classifiers trained with real labels in all tasks and connections/regions ($P < 10^{-40}$, *t*-tests, not

500 indicated) and assume chance level (50%, orange). Accuracies of 80% (yellow line) and

501 90% (black dotted line) are indicated to aid comparison. Shaded area represents s.e.m.

502 Numbers in circles colour-coded for the respective paradigm indicate the rank of the average

503 decoding accuracies achieved using data from the respective connection or region. Ranks

504 have been generated from pairwise comparisons with Tukey post-doc tests conducted after

505 significant effects of connection/region in one-way ANOVAs ($P < 0.0001$ in all cases);

506 connections/regions that were not significantly different from each other were assigned the

507 same rank. (**d**) Decoding accuracies achieved when using the classifiers trained on

508 predictors from all connections/regions combined (**e**) Similar analysis as (d) but with trials

509 from all three paradigms inter-mixed. Blue and grey in (d-e) depicts performance of

510 classifiers trained on trials with correct or incorrect (shuffled) labels (stars indicate *t*-test

511 comparisons between them). Error bars, data range without identified outliers which are

512 highlighted in red; boxes, range between 25th-75th percentile; dot, median. (**f**) Similar

513 analysis as in (d) but using only SP variables from the connection indicated by colour as

514 predictors. Red lines indicate mean, boxes the 25th and 75th percentile, whiskers indicate

515 data range without outliers, and red crosses indicate outliers. See also Supplementary Fig.

516 10 for analysis but using only predictors from single task-phases.

WM-related functional connectivity is highly task-specific and broadly distributed in humans

519 In order to identify possible correlates of WM in humans, we analysed the prediction weights

520 of the individual connectivity metrics similarly as for the mouse dataset, again extracting

521 WM-related metrics based on the two criteria of prediction weight and a different amplitude

522 of the metric in correct trials compared to incorrect trials (Bonferroni-adjusted *t*-tests). As in

523 mice, WM-related measures (185 out of 1344 connectivity predictor variables) were widely

524   distributed across connections, frequency bands, and metric types. When inspecting the

525   matrix of significant predictors more closely, some regularities emerged (Fig. 8a): First, WM-

526   related activity was highly task-specific with 88% of significantly WM-related connectivity

527   metrics being relevant in only a single paradigm. Only a single metric was predictive in all

528   three paradigms – OFC→PFC post-cue delay γ-PDC. The principle task-specificity was

529   maintained also with relaxed *P*-value thresholds (Supplementary Fig. 11-12). Second, by far

530   the most – and the most common - predictors emerged in the γ-band, irrespective of

531   significance threshold (Fig. 8a-b, Supplementary Fig. 11, 13-14). The δ-band - in contrast to

532   mice - contributed almost no WM-related variables (only one each in spatial and temporal

533   WM, confined to the OFC-MTL connection). Also, the θ-band bore relatively few WM-related

534   connectivity parameters, and these were mostly relevant for spatial WM and to a lesser

535   extent for identity WM, but hardly for temporal WM. Furthermore, θ-γ-PAC appeared rather

536   relevant (as found in the same data before[20]) in all three types of tasks, especially identity-

537   related WM. Third, *changes* of a metric relative to the ITI before each trial were rarely

538   predictive. Finally, despite the relatively high decoding accuracy achieved for temporal WM

539   (Fig. 7c-d), the number of connectivity metrics related to this WM-type was considerably

540   smaller (18 out of 1344 measures) than for the other two (72 and 95) and there was hardly

541   any overlap between these metrics and those relevant for the other two WM-paradigms (only

542   3 each, mostly in the γ-band; Fig. 8a-b). In summary, the analysis in humans confirms the

543   high task-specificity, and broad anatomical and frequency-range distribution of WM-related

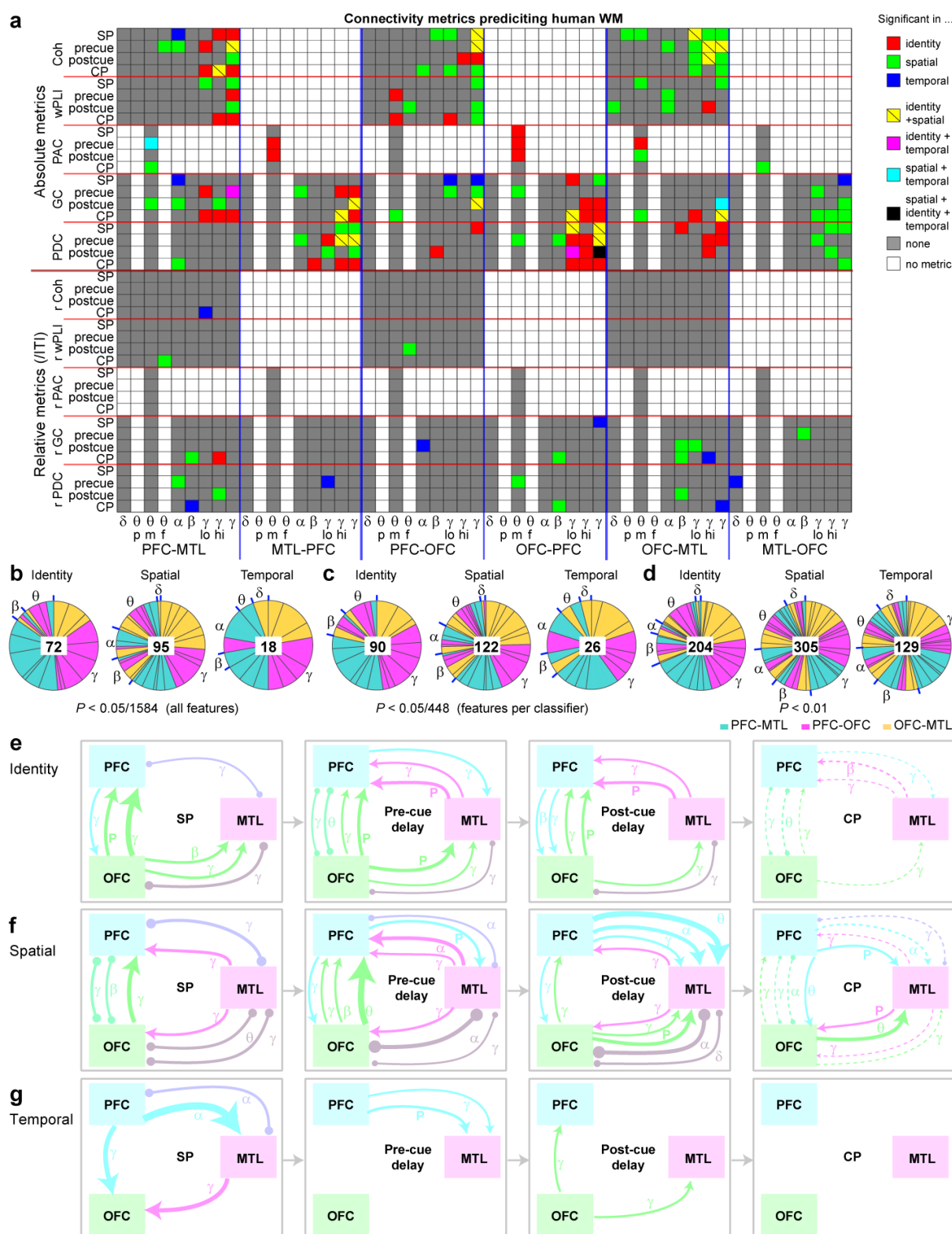544   neural activity already seen in mice.

Fig. 8 | Highly task-specific and broadly distributed correlates of human WM. (**a**) Matrix showing all connectivity predictor variables that contributed to the classifiers shown in (8d) and were significantly associated with WM-performance according to their weight and

549 differences between correct and incorrect CP (see Results) in the paradigms coded by

550 colour (see legend on the right). For θ, mean amplitude (m), peak amplitude (p), and

551 frequency of peak (f) are shown, while for all other variables only the mean amplitude is

552 used. The γ-band contributed three predictors each as this frequency was split into a high-

553 and low-γ-range in addition to using the whole range (30-100 Hz). See Supplementary Fig.

554 11 for the same analysis with relaxed *P*-value correction and Supplementary Fig. 13-14 for

555 prediction weights of the same variables. (**b-d**) Share of each connection (coded by colour)

556 and frequency band (stated around pie chart with separations in blue) among all significant

557 predictors (*N* stated in centre) for the indicated task. The significance threshold has been

558 Bonferroni-adjusted either by the total number of predictor variables from all connections and

559 regions (1548, corresponding to analysis in panel (a); b), or by the number of connectivity

560 variables for a single connection (448; c), or a standard threshold of 0.01 was chosen (d). (**e-**

561 **g**) Depictions of directed connectivity (arrows, derived from GC or PDC) or non-directed

562 coupling (round-ended arcs, coherence or wPLI) during the four phases of each task

563 identified on the left, derived from significant predictor variables shown in (a). Line weights

564 indicate the increase of connectivity in the stated phase relative to the preceding ITI.

565 Measures that are significant but *decrease* in the respective phase relative to the ITI are

566 shown as dashed lines. Significant coupling metrics are not depicted if directed measures

567 are represented in both directions. P, θ−γ-PAC.

Common γ-band connectivity across human WM tasks

569 To scrutinize this conclusion, we searched for commonalities between tasks by aggregating

570 predictive non-directed (coherence, wPLI) and directed (GC, PDC) metrics and the multiple

571 measures within the γ- and θ-bands (extracted from Fig. 8a), and depicted their amplitude

572 change relative to the preceding ITI for each task phase (Fig. 8e-g), as previously done for

573 the mouse dataset (Fig. 5c-e). In this analysis, OFC-PFC γ-coupling during encoding and

574 directed OFC→PFC/MTL γ-connectivity during the post-cue delay emerged as common

575 patterns present in every task. There were also more commonalities between spatial and

576   identity WM, namely γ-coupling between all three regions that was elevated throughout

577   encoding and delay phases and then decreased below ITI-levels in the CP (Fig. 8e-f).

578   Strikingly in fact, *all* significant predictors from the $\alpha-\beta-\gamma$-range in these tasks showed

579   *elevated* amplitudes during encoding and delay, but *decreased* amplitudes during the CP,

580   compared to their amplitude in the preceding ITI (Fig. 8e-g). Only θ and $\theta-\gamma$-PAC predictor

581   variables *increased* in amplitude during the CP in spatial WM (Fig. 8f). Finally, even with this

582   aggregated analysis, not a single connectivity pattern that was shared between any two

583   tasks emerged outside the γ-band in any task phase (Fig. 8e-g). This analysis suggests that

584   human WM generally relies on anatomically broad, task phase-specific modulation of γ-

585   connectivity between several brain regions irrespective of task, while the engagement of

586   oscillatory coupling in other frequency bands is task-specific.

587

Discussion

589   Here we demonstrate that WM-related choices can be predicted trial-by-trial in mice and

590   humans using linear decoding of high-dimensional arrays of LFP-based measures of inter-

591   regional connectivity or local activity (the *electome*). The high decoding accuracies of around

592   90% (compared to a chance level of 50%) achieved in both species are remarkable

593   considering the spatially coarse nature of the extracted neural signal, the short – often sub-

594   second – data traces used to calculate predictors, the intrinsic variability caused by merging

595   data from all analysed individuals with varying electrode placements, and the lack of precise

596   neuronal information as encoded in spike trains of individual neurons[10,21,33,41,47]. Using the

597   trial-by-trial predictive power of physiological activity as the indicator for its association with

598   WM[42], this approach enabled a largely unbiased top-down analysis to reveal an

599   unexpectedly rich pattern of frequency-specific connectivity changes during individual

600   phases of distinct WM assays in mice and humans.

601   The comparative analysis of multiple WM assays - including those that allow control over

602   basic motivational and attentional parameters - provides a unique advantage in that

603   neurophysiological activity patterns which might be truly relevant to WM may be isolated. In

604   this way, we could reveal MD-PFC $\delta-$ and β-range coupling during memory encoding and

605   maintenance, respectively, as well as vHC-PFC and vHC-dHC δ/θ-coupling during retrieval

606   as common connectivity patterns across all three rodent tasks - although the vast majority of

607   connectivity proved to be task-specific (Fig. 5c-e). In humans, γ-band connectivity across all

608   analysed connections was commonly linked to WM-choice across tasks, while WM-related

609   connectivity in other frequency bands was mostly task-specific (Fig. 8).

610   Against a backdrop of widely varying assumptions about which kind of neural connectivity

611   underlies WM (see Introduction), our analysis was initially motivated by the possibility to

612   extract "true" anatomical and frequency-related WM-correlates using the predictor weights

613   generated by the linear classifiers that decode WM-based choices with high accuracy. Our

614   results, however, refute some implicit key assumptions of this endeavour – and, by

615   extension, of many prior investigations of WM-correlates: First, there is no single – or small

616   set of – anatomical regions or connections, types of directional information transmission, or

617   frequency bands that can be regarded as a unique *WM correlate*. Indeed, previously

618   suggested "correlates", especially in the rodent literature (Supplementary Table 1), could

619   appear as such only because the sum total of connections and measures investigated in

620   each study was small (streetlight effect), as opposed to the 1184 and 1584 metrics analysed

621   here in mice and humans, respectively. In our study, virtually every analysed frequency

622   band, metric, connection and region bore some predictive power regarding WM-mediated

623   choice.

624   Second, there is no single behavioural WM-task that could be regarded as *representative* of

625   the generic psychological construct termed "working memory" in order to allow the

626   identification of *the* neurophysiological correlate of that construct. The latter is illustrated by

627   the enormous variability in the patterns of predictive connections and metrics across task-

628   paradigms in both species. In other words, a physiological variable that correlates with

629   choice accuracy in the T-maze represents a neurophysiological correlate of T-maze

630   performance, but not necessarily of WM. The same principle applies to our cross-species

631   comparison, as the uncovered candidates for a generic (task-independent) WM correlate

632   originated from different frequency bands in humans (γ) than in rodents ($\delta-\theta-\beta$). A

633   translational implication of these findings is that it is likely impossible to define

634   neurophysiological underpinnings of "working memory" as a uniform psychological construct.

635   However, the existence of certain tasks that represent such a psychological function (i.e.,

636   that engages a physiological mechanism that is central to all WM tasks across species) is an

637   implicit key assumption of the *Research Domain Criteria (RDoC)* approach, which envisions

638   to use those representative paradigms in search of WM-enhancing cellular and molecular

639   targets[48]. Our data suggest that the key target variable in the preclinical discovery of WM-

640   enhancing compounds might be the appropriate regulation of γ-range connectivity (given its

641   importance for human WM) rather than behavioural performance in any particular rodent

642   task.

643   The analysis of the human dataset, in particular, paints a rather different picture of what a

644   correlate of WM could be – at least when searched for in LFP-data. In all three tasks,

645   prediction accuracy calculated from connectivity (as opposed to local activity) metrics was

646   not only very high, but it was also roughly *equal* between the three analysed connections,

647   even though these are anatomically quite distinct. This was the case even with the limited

648   analysis incorporating only SP or post-cue delay connectivity as predictors of spatial WM.

649   These findings may be taken as an indication that WM-related information is extremely

650   broadly distributed, and - rather than specific activity located in a certain connection or

651   region - it is the ability to manipulate information flow across brain regions *as such*, that

652   determines task performance[49,50]. This model is in line with the ever growing list of brain

653   areas that are implicated in WM, including the superior frontal[51], anterior cingulate[52,53] and

654   sensory cortex[54,55], ventral tegmental area[28], and the nuclei of the midline and anterior

655   thalamus[4,9,34], and the concept that different areas may be involved depending on the

656   strategy used to solve the task[54]. An unexpectedly broad anatomical representation of

657   sensory and behavioural information across the brain has recently been uncovered by

658   decoding activity of individual neurons in multiple cortical areas[56,57], and our decoding

659   analysis of LFP-based connectivity in cognition underscores this phenomenon.

660   In conclusion, our multi-area decoding approach and cross-task cross-species comparative

661   analysis revealed not only a rich functional connectivity supporting WM, exceeding the WM-

662   associated connectivity described before. It also demonstrated an unexpected task- and

663   species-specificity of WM-related neural coupling that raises substantial caution regarding

664   the predictive translational value of each assay and demands to re-think our search for

665   physiological WM-correlates.

666

Methods

Animals and surgery

All experiments were performed in accordance to the German Animal Rights Law (Tierschutzgesetz) 2013 and were approved by the Federal Ethical Review Committee (Regierungsprädsidium Tübingen) of Baden-Württemberg, Germany (licence number TV1399). The rodent cohort, surgery and histology details have been described in our prior publication of the open-field data from the same cohort[37]. Briefly, 12 C57BL/6N wildtype mice, including 8 males, were selected from offsprings of an in-house colony (heterozygous *Gria1*[tm1Rsp] mice; MGI:2178057)[58], group-housed in Type II-Long individually ventilated cages (Greenline, Tecniplast, G), enriched with sawdust, sizzle-nest[TM], and cardboard houses (Datesand, UK), and subjected to a 13 h light / 11 h dark cycle. The mice were implanted with electrodes at ca. 9 months of age, after 2 months of training in the DMTS-WM task and prior habituation training (see below). Single polyimide-insulated tungsten wires of 50 µm diameter (WireTronic Inc., CA, US) were implanted, with reference to Bregma (in mm), into the PFC (AP +1.8-1.9, ML 0.3-0.35; 1.8-1.9 below pia), MD (AP -1.2, ML 0.3, 2.7 below pia), dHC (AP -1.9-2.0, ML 1.5, 1.4 below pia), and vHC (AP -3.1-3.2, ML 2.9-3.0, 3.4 mm for single and 3.8-3.9 mm for dual electrodes below pia). In most mice, dual electrodes were used for PFC and vHC, whereby the second electrode was placed about 0.5 mm higher than the stated distance from pia. Both hemispheres were implanted at roughly equal proportion. Stainless steel screws (1.2 mm diameter, Precision Technologies, UK) were implanted in the contralateral hemisphere ca. 1 mm from the midline above the cerebellum (AP -5.5) for ground and above the anterior frontal cortex (AP +4.0) for additional reference (used for the analysis in the open-field test, but not for the present analysis), and where connected with a 120 µm PTFE-insulated stainless steel wire (Advent Research Materials Ltd., UK; Fig. 1b). All electrode wires were connected to pins in a dual-row 6-pin or 8-pin connector (Mill-Max, UK). Electrode placements were determined *post-mortem* from electrolytic lesions made under terminal ketamine/medetomidine anaesthesia followed by

694   perfusion-fixation. Misplaced electrodes were excluded, leaving the following number of

695   used electrodes per region; PFC: 12; dHC: 7; vHC: 9; MD: 4. The resulting number of

696   contributing connections were PFC-dHC: 7; PFC-vHC: 9; vHC-dHC: 6; MD-PFC: 4.

697   ## Operant DMTS 5-CSWM task

698   The principal schedule of the task was as previously described[45] and emulates the *combined*

699   *attention and memory* (CAM) task previously developed in rats[59]. In the present study, the

700   task was conducted in custom-designed pyControl-based operant boxes that were optimized

701   for both simultaneous electrophysiological recordings and the acquisition of the 5-CSWM

702   task (operant box design-files and task-scripts implementing all task paradigms used in this

703   study available from https://github.com/KaetzelLab)[44]. The latter was done by using a 5-

704   choice poke-wall with a larger distance between the holes (3 cm edge-to-edge; 4.5 cm

705   centre-to-centre), which we chose because of the long learning time in our standard

706   commercial boxes used before[45]. Briefly, each trial of the task is divided into a sample phase

707   (SP), a delay-phase (delay) - which is further sub-divided into a pre-delay and a post-delay

708   by the time point of reward collection - the choice phase (CP), and the inter-trial interval (ITI).

709   In the default state, the task is conducted in the dark (house-light off) with the only

710   illumination deriving from the poke- and receptacle holes in certain task phases. The SP is

711   identical to that of the 5-choice-serial-reaction-time task (5-CSRTT), except that premature

712   responses are not punished: one of the 5 holes in the 5-choice wall is illuminated for a

713   certain stimulus duration (SP-SD) and mice need to poke into that hole within the limited

714   hole time (SP-SD plus 1 s) in order to obtain a small reward (20 µl or, in stages 2 and

715   onwards, 10 µl strawberry milk) at the receptacle at the opposite end of the wall during the

716   pre-delay time. The reward-collection (receptacle exit) starts the post-delay (2 s in all cases)

717   after which the originally presented hole and one randomly assigned other hole is illuminated

718   (thereby starting the CP) for a certain stimulus duration (CP-SD, 5 s for all protocols shown

719   in Fig. 1). Mice have to poke into the same hole as in the SP, realizing a DMTS-rule, in order

720   to obtain a large reward (60 µl). After a 5 s ITI a new trial starts with the SP. Incorrect pokes

721   or omissions in the SP or CP lead to a 5 s time-out period (house-light illuminated; reward

722   omitted) and the start of a new trial after an ITI of 5 s. Before training, mice were habituated

723   to the operant box, to consuming the strawberry milk (Müller®, Germany) reward from the

724   receptacle, and to poking into the 5-choice wall to obtain a reward (acquisition of the basic

725   operant cycle).

726   Subsequently, training was conducted through multiple stages across which the task

727   became incrementally more difficult due to a shortening of SDs and an increase in the

728   number of CP-stimulus configurations (see Supplementary Table 5). During pre-surgery

729   training – in order to compare performance between groups – no performance-based staging

730   was applied, but instead all mice were trained on the stage 1 for the same number of 21

731   days, and then transitioned through the remaining three training stages (2-4) with 2-3 d of

732   training per stage. Parameters defining the stages are found in Supplementary Table 5.

733   Sessions lasted 30 min throughout, and were conducted on 5 d/week. Training was

734   continued ~4-5 weeks after surgery on the baseline stage (4) for 5 weeks to allow the mice

735   to approach asymptotic performance. Subsequently, mice were trained further for 3 weeks in

736   a tethered mode, with the headstage (see below) mounted on their heads, in a baseline

737   protocol with shorter SD in SP (10 s) and CP (5 s; stage 5) to increase the number of

738   obtained trials and better standardize the encoding time. Subsequently, mice were taken

739   through three series of challenge protocols with intermittent training on the baseline stage 5.

740   The same challenge was conducted on 2-3 consecutive days in order to obtain sufficient

741   trials for later analysis. The three series were *(a)* pure delay challenges, where the pre-delay

742   was extended from 0 to 5 and 10 s, *(b)* a distraction challenge with an illumination of the

743   house-light for 0.5 s starting randomly timed between 0.8-1.3 s of the 2 s post-delay phase,

744   and *(c)* a combined attention (SP) and working memory challenge with an SP-SD of 1 s

745   (instead of the 2 s of the specific baseline stage 6 of this challenge) and a pre-delay of 5 s,

746   which was preceded by a sole attention challenge (1 s SP-SD, 0 s pre-delay). Most of the

747   analysis shown uses the data from the final (combined) challenge, although decoding

34

748   analysis was also conducted for the other challenges in order to replicate the analysis and,

749   additionally, to assess the capacity of cross-prediction between classifiers from entirely

750   different challenge-conditions (Supplementary Fig. 5). The post-delay remained 2 s, starting

751   with the exit from the reward receptacle and spent in darkness, throughout all challenge and

752   baseline protocols.

Operant DNMTS 2-CSWM task

754   The operant DNMTS task followed the same principal trial-schedule as the 5-CSWM task

755   except for two modifications: the implementation of a *non*-match-to-position rule (i.e. animals

756   are rewarded for choosing the illuminated CP poke-hole that is not the one, that they poked

757   in the SP) and a simplified set of choice options using always only holes 2 and 4 of the 5-

758   choice wall (making it similar to a task developed by Goto & Ito[60]). Mice were trained in this

759   task only after the T-maze (see below) in order to ease the switching from the prior, opposite

760   task-rule. All mice were trained for 30 training sessions on the 2-choice DNMTS task, then

761   were tested in two delay challenges, in which the pre-delay was extended to 5 and then 10 s

762   (2 d each). A subset was then trained in a 5-choice version of the same task (i.e. the

763   DNMTS-version of the 5-CSWM task) for 12 d *without* and for 3 d *with* headstage mounted,

764   but performance in many mice was not sufficient, and hence mice were returned to the 2-

765   choice version and trained for a further 5 d with mounted headstage. The other subset

766   moved directly to 5 d of training with mounted headstage. Then, recordings were conducted

767   on two days with baseline training and two days each for two delay challenges, in which the

768   pre-delay was extended to 5 and then 10 s. Given the relatively low performance in the delay

769   challenges, the data from the two baseline days was used for further analysis of

770   electrophysiology data. In all protocols, the SP-SD was 8 s, the CP-SD was 5 s, the limited

771   hold time exceeded the SD by 1 s, ITI and time-out were each 5 s, the post-reward delay

772   was 2 s (spent in darkness), the SP reward was 10 μl, and the CP reward 60 μl.

Non-matching-to-position T-maze rewarded alternation SWM task

Rewarded alternation-based SWM was tested in a T-Maze with transparent walls (10 cm high), intransparent floor (red PVC), and food wells (made from white Teflon®) placed in the end of all three arms. Sliding doors could be used to block off either of the two choice arms in the SP and the initial 10cm partition (containing the food-well) of the start arm during the delay and ITI. Before testing mice were habituated to the surroundings and to the condensed milk reward (10% Ja!-Kondensmilch®, Germany, diluted 1:1 with drinking water) at first cage-wise and later independently. The task schedule was identical to what we previously described[26], involving ten trials per day each consisting of an SP and a CP. During the SP, mice were placed at the beginning of the start arm facing the experimenter, and were left to run into the pseudo-randomly assigned goal arm that was not blocked to obtain a reward. Subsequently, mice returned to the end of the start arm to obtain another, small reward while being enclosed for a 5 s delay period, during which the door from the previously blocked goal arm was removed. Once the delay had passed, the CP started in which the mouse was allowed to choose between one of the goal arms of which the previously unvisited one was rewarded (correct choice), while the other one was not (incorrect choice). Once the CP was completed, mice were motivated to return to the start arm to obtain another small reward there while being enclosed for an ITI of 20 s before the next trial commenced. 10 consecutive trials were conducted per day; testing was conducted over 8 d with a delay of 5 s and for another 4 d with a delay of 30 s – all 12 sessions were conducted with simultaneous recording of LFP-activity, i.e. a headstage was mounted and tethered through an SPI-cable. In each session, a new sequence of open arms in the SP was used whereby half of the 10 trials were always assigned to the left arm, and the sample arm was not the same for more than three consecutive trials.

Acquisition and pre-processing of mouse data

Prior to testing, a 32-channel RHD2132 headstage (Intan Technologies, CA, US) was plugged into the implanted connector via a custom-built adaptor that interfaced a 36-pin Omnetics connector (A79022-001, MSA components, G) with another 6-pin or 8-pin Mill-Max

801   connector. The headstage was connected to an *Open-Ephys* acquisition board (https://open-

802   ephys.org, US; obtained through the Open-EPhys store at Champalimaud, Portugal) via two

803   light-weight flexible SPI-cables (Intan Technologies), sometimes daisy-chained through a

804   custom-connected miniature slip-ring (Adafruit, NY, US). The adaptor was wired so that all

805   signals were referenced to the ground-signal obtained from above the contralateral

806   cerebellum. Data were amplified and digitized, sampled at 20 kHz and band-pass filtered at

807   0.1 – 250 Hz for all subsequent analysis of LFP signals. In operant tasks, all individual task-

808   and behavioural events were recorded by pyControl. Additionally, all events relevant to time-

809   locked electrophysiological analysis of WM task-phases and choices (e.g., correct SP and

810   CP responses) were encoded as patterns of transistor-transistor logic (TTL) signals by

811   pyControl and recorded as time-stamps with the electrophysiological data by the *Open-*

812   *EPhys* acquisition software using the 8 analogue inputs of the acquisition board and a

813   dedicated BNC-HDMI interface board (Open-EPhys). For the T-maze task, ANY-maze

814   (Stoelting) was used to track the position of the animal in the different subdivisions of the

815   maze, and this positional information was encoded in patterns of TTL-signals recorded via

816   an AMi-interface board and a BNC-HDMI interface board as time-stamps with the

817   electrophysiological signals in the *Open-EPhys* acquisition software.

818

Human intracranial electrophysiology data during WM

820   A publicly available dataset of multi-site intracranial recordings during three WM tasks [20] was

821   downloaded from http://dx.doi.org/10.6080/K0VX0DQD. It includes data from 10 adult

822   human subjects (mean ± SD [range]: 37 ± 13 [22–69] y; 7 males) who were implanted with

823   intracranial electrodes to identify epileptic foci for surgical resection. Electrode placements

824   were in the medial temporal lobe (MTL, i.e., CA1; CA3/dentate gyrus; subiculum; or

825   parahippocampal, perirhinal, or entorhinal area), lateral PFC (inferior, middle, or superior

826   frontal area), and OFC (orbitofrontal, frontal polar, or medial prefrontal area). Only subjects

827   with electrodes localized in all three regions were included in the present analyses ($N = 8$).

828   Details of behavioural testing, data acquisition, and pre-processing have been described

829   previously[20]. In brief, three different WM tasks were conducted, whereby subjects had to

830   either identify a previously indicated object identity, location, or temporal order of two visual

831   stimuli. Trials from all these three tasks were pseudo-randomly mixed from trial to trial in a

832   single test session. Each trial started with a 1 s pre-trial fixation interval, after which a screen

833   indicated whether in the respective trial would be tested for object identity or spatiotemporal

834   position (800 ms). Subsequently, the SP started in which two shapes were presented

835   subsequently for 200 ms each, separated by 200 ms. After a subsequent pre-cue delay (900

836   or 1150 ms, varied pseudo-randomly), a cue appeared for 800 ms that specified which of the

837   two shapes would need to be identified in the later CP according to a rule of identity

838   (same/different), spatial location (top/bottom) or temporal order (first/second). After a post-

839   cue delay (900 or 1150 ms, varied pseudo-randomly), the CP started as two shapes were

840   presented, of which the participant had to choose the one that was correct according to the

841   prior cue. All trials from all patients were merged for subsequent analyses, just as was done

842   for the mouse WM data. The fully pre-processed data (as described[20]) was used for the

843   current analysis.

844   **Data analysis**

845   All signal analyses were done in MatLab (MathWorks). Mouse electrophysiology data were

846   exported to MatLab and, for all LFP analyses, down-sampled to 1 kHz and analysed with

847   custom-written scripts. To reduce low frequency drift, signals were first detrended using the

848   *locdetrend* function of the Chronux signal processing toolbox (http://chronux.org/) with 1 s of

849   data and a sliding window of 0.5 s. Trials were excluded from further analyses if the

850   amplitude exceeded the 5th standard deviation within each channel for more than 10% of the

851   trial duration. In mice, PFC and vHC were recorded with dual electrodes. In case both

852   electrodes were located in the intended target area (PrL/Cg1 for PFC and fissure for vHC),

853   as inferred from lesion sites, all metrics were calculated for both electrodes and the result

854   was averaged to obtain the final value. For the human dataset, the number of electrodes per

855   site varied between 1 and 28 per area. Therefore, for connectivity measures, each single

856   metric was calculated for every possible inter-regional pair of electrodes and the resulting

857   value was averaged across all combinations of a single connections. Analogously, for local

858   activity measures, each single metric was calculated for every electrode and the resulting

859   value was averaged across all electrodes of a given area for each subject.

860

Power and non-directional synchrony

861

862   Power and coherence spectra were calculated with Chronux routines implemented in the

863   Chronux toolbox using the multi-taper method[61]. Power values were expressed as $10*\log_{10}$

864   values for all analyses and the range of frequencies was set from 0.1 to 48 Hz. A time-

865   bandwidth product of 9 and 17 tapers were used to calculate power and coherence during

866   defined time-periods during ITI, sample phase, delay (if applicable) and choice phase. To

867   address the issue of volume conduction, we calculated the weighted phase lag index

868   (wPLI)[62] using routines implemented in the FieldTrip toolbox[63]. The original trials were further

869   divided into 99% overlapping "pseudo trials" with a length of 600 ms and padded to the next

870   power of two. The complex cross-spectrum was computed using a Hann taper with a

871   spectral smoothing of 2 Hz. Time-frequency spectral analyses were performed with routines

872   from the FieldTrip toolbox using Morlet wavelets with a width of 3 cycles steps of 10 ms[63].

873   Time periods before and after the time frame of interest were padded with real data to avoid

874   artifacts of too long wavelets at low frequencies.

Directional Synchrony

875

876   Non-parametric Granger Causality (npGC)[64] and partial directed coherence (PDC)[65] were

877   calculated using the FieldTrip toolbox[63]. The original trials were further divided into 50%

878   overlapping "pseudo trials" with a length of 1s and padded to the next power of two, differing

879   from power and non-directional synchrony measures because 99% overlap did not provide

880   substantially different results but came with a much higher computational effort. The complex

881   cross-spectrum was computed using a Hann taper with a spectral smoothing of 2 Hz. The

882   noise covariance matrix and transfer function were obtained by applying Wilson's spectral

883   matrix factorization to complex Fourier-spectra. This non-parametric approach was shown to

884   be better at capturing all spectral features, less error prone because no model order had to

885   be chosen, computationally faster than applying autoregressive modelling[66,67], and to deliver

886   virtually same results when used on our LFP data[37]. Time-frequency representations of

887   npGC and PDC were obtained by Morlet wavelets using the same configurations as

888   described above.

Phase-amplitude coupling (PAC)

890   Cross-frequency coupling (CFC)[68] was assessed using the measure of phase-amplitude

891   coupling (PAC), the statistical relationship between the phase of a low-frequency and the

892   amplitude of a high-frequency component, in a cross-regional analysis[69,70]. Time series data

893   were first band-pass filtered in the desired frequency ranges, followed by a Hilbert transform

894   using the MatLab function *hilbert* to calculate the real and imaginary parts of the signal to

895   obtain the instantaneous amplitude and phase. For each trial, intra- and inter-regional PAC

896   were determined by calculating the modulation of γ-amplitude by θ-phase using the phase-

897   locking technique proposed by Voytek et al. with routines described in[71].

Supervised Machine Learning

899   To validate the calculated measures of neural connectivity and to identify predictive variables

900   we employed supervised machine learning algorithms. Spectrally resolved parameters (e.g.,

901   γ-coherence, γ-power) from each inter-regional connection (e.g., dHC-PFC) and local brain

902   region (e.g., dHC) were analysed separately using different classifiers. For classification, we

903   used the absolute parameter values as well as the ratio of each parameter relative to the

904   preceding ITI. In mice, 240 connectivity metrics per connection and 56 local activity metrics

905   per region were used in the 5-CSWM DMTS task (1184 for all connections and regions

906   combined; 296 per task-phase), and 180 connectivity metrics and 42 local activity metrics

907   were used on the operant 2-CSWM DNMTS task and the T-maze (888 combined). The

908   difference originates from the usage of 4 task phases, including pre- and post-delay phases,

40

909　in the 5-CSWM task, but only 3 phases (one delay phase) in the two DNMTS tasks, which

910　was the post-delay in the 2-CSWM task. For all predictor matrices, the pre- and post-delay

911　phases were combined to allow uniform comparisons between all three tasks. In humans,

912　448 connectivity metrics per connection and 80 local activity metrics were generated for the

913　analysis shown in Fig. 7, using all 4 task phases including SP, pre-cue delay, post-cue

914　delay, and CP (classifiers omitting the CP were also calculated, see Supplementary Fig. 10).

915　Note that, in humans more predictors arise because frequency bands have been determined

916　additionally in the alpha-band (8-12 Hz), and in a higher γ-band (50-100 Hz) in addition to

917　the low γ-band (30-49 Hz) used in rodent analysis. We also included the combined band (30-

918　100 Hz) as separate predictor. To ensure a sufficient number of trials for classification and

919　the general validity of the identified predicting variables, we merged data from all subjects of

920　one group, i.e., from all mice provided correct electrode placement (see Supplementary

921　Table 6) or from the subset of human subjects with coverage of all three regions,

922　respectively.

923　Since rodent and human subjects performed proficiently above chance level resulting in

924　more correct than incorrect trials[20], we used a *synthetic minority over-sampling technique*

925　(SMOTE) to construct a balanced dataset with five nearest neighbours to consider[72]. All

926　electrophysiological predictor variables of a classifier were normalized between 0 and 1,

927　setting the maximum empirical value for each metric to 1. We used 90% of the data as a

928　training set and the remaining 10% for testing. Allocation to the training and test set was

929　done randomly and repeated 100 times to obtain a mean and its variance for the achieved

930　decoding accuracy the predictor weights. To identify the classification algorithm which fits

931　our data best we assessed the 25 most used classifiers implemented in MATLAB

932　(Supplementary Fig. 4). Focussing on easy interpretability and high predictive accuracy we

933　chose random subspace ensembles on a linear discriminant analysis (LDA) template

934　(subspace discriminant classifier) for our further analysis, which achieved the highest

935　prediction accuracies compared to all other tested linear classifiers. Briefly, LDA aims to

936   identify a hyperplane that maximizes the mean distance between the mean of the two

937   classes while minimizing variance between them. Since the sample size of our data was

938   relatively small compared to the number of features, we used the *random subspace method,*

939   which is a valid approach to resolve this issue and has been shown to be superior to single

940   classification algorithms[73,74]. It operates by creating a classifier ensemble where each

941   classifier is trained with a reduced, randomly sampled number of input features, e.g., they

942   are projected into a new subspace which leads to a relative increase of the number of

943   samples. The number of features to sample in each classifier and the number of learning

944   cycles were set to half of the total number of features and 30, respectively. The coefficient

945   magnitudes of each feature obtained by each subspace LDA classifier were averaged

946   across learning cycles to get a solid quantification of its predictive value. As measures of

947   classification performance, we used the prediction accuracy, the AUC of the receiver

948   operating characteristic (ROC) and the F1-score which is defined as follows: F1 =

949   (2*class1precision * class1recall / (class2precision + class1recall)). Precision is defined as

950   the True positives / True Positive + False Positive for class 1 and as the True negatives /

951   True Negative + False Negative for class 2. Recall is defined as True Positive / True Positive

952   + False Negative which is equivalent to the sensitivity for class1 and specificity for class2.

953   Since it has been shown that the theoretical chance level of 50% should not be expected

954   and it is favourable to obtain an empirical chance level, we randomly shuffled the data labels

955   (e.g. correct / incorrect) and repeated the analysis described above to create an empiric null

956   distribution[75].

Statistical analysis

958   Behavioural training and challenge data were analysed with repeated-measures ANOVA and

959   pairwise Sidak post-hoc tests for simple main effects. To determine the importance of

960   individual connectivity or activity measures (predictors), we used a two-step procedure: First,

961   pre-classification, we performed a paired *t*-test for each feature comparing its value in

962   correct vs. incorrect trials. Second, post-classification, we used the magnitude of the weight

963   of each predictor as delivered by the subspace discriminant classifiers to perform a paired *t*-

964   test between the obtained magnitudes and the magnitudes from the shuffled dataset.

965   Features were only recognized as significantly important if both criteria were met. *P*-values

966   were Bonferroni-corrected within-species for the total number of used features across all

967   classifiers calculated in mice (1184; $P < 0.05/1184$) and humans (1584; $P < 0.05/1584$) in

968   the most conservative analysis; additional analysis was conducted to evaluate the

969   dependency of the number of obtained significant predictors with less stringent adjustments.

970

Data and protocol availability

972   All source data for behavioural performance in mice can be obtained from the corresponding

973   author upon reasonable request. All electrophysiological data from mice will be made

974   publicly available at https://gin.g-node.org/KaetzelLab[76]. Human data are available at

975   http://dx.doi.org/10.6080/K0VX0DQD. All MATLAB analysis scripts for spectral analysis are

976   publicly available on GitHub (https://github.com/KaetzelLab/LFP_analysis)[77]. Design files of

977   custom-made operant boxes (https://github.com/KaetzelLab/Operant-Box-Design-Files) and

978   task-files for operant WM tasks (https://github.com/KaetzelLab/Operant-Box-Code) are

979   available on GitHub.

980

Acknowledgments

Author Contributions

991 SKTK and DS conducted all mouse experiments. DS analysed mouse and human data, with

992 initial support from AMB and BFG. DS and DK designed the study based on prior input from

993 AMB, BFG, and DMB. BL provided essential resources. ELJ provided the human data.

994 SKTK, TA, and DK developed the pyControl operant box enabling simultaneous recording

995 during operant behavioural testing. DS and DK wrote the manuscript with advice from all

996 authors.

Competing Interests Statement

998 The authors declare no competing interest.

999

References

1. Baddeley, A. D. & Hitch, G. Working Memory. in *Psychology of Learning and Motivation* (ed. Bower, G. H.) vol. 8 47–89 (Academic Press, 1974).

2. Baddeley, A. Working memory. *Science* **255**, 556–559 (1992).

3. Millan, M. J. *et al.* Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat. Rev. Drug Discov.* **11**, 141–168 (2012).

4. Lalonde, R. The neurobiological basis of spontaneous alternation. *Neurosci. Biobehav. Rev.* **26**, 91–104 (2002).

5. D'Esposito, M. *et al.* The neural basis of the central executive system of working memory. *Nature* **378**, 279–281 (1995).

6. Curtis, C. E., Rao, V. Y. & D'Esposito, M. Maintenance of spatial and motor codes during oculomotor delayed response tasks. *J. Neurosci.* **24**, 3944–3952 (2004).

7. Goldman-Rakic, P. S. The physiological approach: functional architecture of working memory and disordered cognition in schizophrenia. *Biol. Psychiatry* **46**, 650–661 (1999).

8. Yoon, T., Okada, J., Jung, M. W. & Kim, J. J. Prefrontal cortex and hippocampus subserve different components of working memory in rats. *Learn. Mem.* **15**, 97–105 (2008).

9. Hallock, H. L., Wang, A. & Griffin, A. L. Ventral Midline Thalamus Is Critical for Hippocampal–Prefrontal Synchrony and Spatial Working Memory. *J. Neurosci.* **36**, 8372–8389 (2016).

10. Stout, J. J. & Griffin, A. L. Representations of On-Going Behavior and Future Actions During a Spatial Working Memory Task by a High Firing-Rate Population of Medial Prefrontal Cortex Neurons. *Front. Behav. Neurosci.* **14**, (2020).

11. Johnson, E. L. *et al.* Bidirectional Frontoparietal Oscillatory Systems Support Working Memory. *Curr. Biol.* **27**, 1829-1835.e4 (2017).

12. Parto Dezfouli, M., Davoudi, S., Knight, R. T., Daliri, M. R. & Johnson, E. L. Prefrontal lesions disrupt oscillatory signatures of spatiotemporal integration in working memory. *Cortex* **138**, 113–126 (2021).

1028 13. Johnson, E. L. *et al.* Spectral Imprints of Working Memory for Everyday Associations in

1029 the Frontoparietal Network. *Front. Syst. Neurosci.* **12**, (2019).

1030 14. Parnaudeau, S. *et al.* Inhibition of Mediodorsal Thalamus Disrupts Thalamofrontal

1031 Connectivity and Cognition. *Neuron* **77**, 1151–1162 (2013).

1032 15. Fuster, J. M. & Alexander, G. E. Neuron Activity Related to Short-Term Memory.

1033 *Science* **173**, 652–654 (1971).

1034 16. Cahusac, P. M. B., Miyashita, Y. & Rolls, E. T. Responses of hippocampal formation

1035 neurons in the monkey related to delayed spatial response and object-place memory

1036 tasks. *Behav. Brain Res.* **33**, 229–240 (1989).

1037 17. O'Keefe, J. & Speakman, A. Single unit activity in the rat hippocampus during a spatial

1038 memory task. *Exp. Brain Res.* **68**, 1–27 (1987).

1039 18. Watanabe, T. & Niki, H. Hippocampal unit activity and delayed response in the monkey.

1040 *Brain Res.* **325**, 241–254 (1985).

1041 19. Hallock, H. L., Arreola, A. C., Shaw, C. L. & Griffin, A. L. Dissociable roles of the dorsal

1042 striatum and dorsal hippocampus in conditional discrimination and spatial alternation T-

1043 maze tasks. *Neurobiol. Learn. Mem.* **100**, 108–116 (2013).

1044 20. Johnson, E. L. *et al.* Dynamic frontotemporal systems process space and time in

1045 working memory. *PLOS Biol.* **16**, e2004274 (2018).

1046 21. Boran, E. *et al.* Persistent hippocampal neural firing and hippocampal-cortical coupling

1047 predict verbal working memory load. *Sci. Adv.* **5**, eaav3687 (2019).

1048 22. Miller, E. K., Lundqvist, M. & Bastos, A. M. Working Memory 2.0. *Neuron* **100**, 463–475

1049 (2018).

1050 23. Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S. & Glahn, D. C. Meta-analysis of

1051 41 functional neuroimaging studies of executive function in schizophrenia. *Arch. Gen.*

1052 *Psychiatry* **66**, 811–822 (2009).

1053 24. Jacob, S. N., Hähnke, D. & Nieder, A. Structuring of Abstract Working Memory Content

1054 by Fronto-parietal Synchrony in Primate Cortex. *Neuron* **99**, 588-597.e5 (2018).

46

1055   25. Yamamoto, J., Suh, J., Takeuchi, D. & Tonegawa, S. Successful Execution of Working

1056   Memory Linked to Synchronized High-Frequency Gamma Oscillations. *Cell* **157**, 845–

1057   857 (2014).

1058   26. Bygrave, A. M. *et al.* Hippocampal–prefrontal coherence mediates working memory and

1059   selective attention at distinct frequency bands and provides a causal link between

1060   schizophrenia and its risk gene GRIA1. *Transl. Psychiatry* **9**, 142 (2019).

1061   27. Jones, M. W. & Wilson, M. A. Theta Rhythms Coordinate Hippocampal–Prefrontal

1062   Interactions in a Spatial Memory Task. *PLOS Biol.* **3**, e402 (2005).

1063   28. Duvarci, S. *et al.* Impaired recruitment of dopamine neurons during working memory in

1064   mice with striatal D2 receptor overexpression. *Nat. Commun.* **9**, 2822 (2018).

1065   29. Sigurdsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A. & Gordon, J. A. Impaired

1066   hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature*

1067   **464**, 763–7 (2010).

1068   30. O'Neill, P.-K., Gordon, J. A. & Sigurdsson, T. Theta Oscillations in the Medial Prefrontal

1069   Cortex Are Modulated by Spatial Working Memory and Synchronize with the

1070   Hippocampus through Its Ventral Subregion. *J. Neurosci.* **33**, 14211–14224 (2013).

1071   31. Tamura, M., Spellman, T. J., Rosen, A. M., Gogos, J. A. & Gordon, J. A. Hippocampal-

1072   prefrontal theta-gamma coupling during performance of a spatial working memory task.

1073   *Nat. Commun.* **8**, 2182 (2017).

1074   32. Tamura, M., Mukai, J., Gordon, J. A. & Gogos, J. A. Developmental Inhibition of Gsk3

1075   Rescues Behavioral and Neurophysiological Deficits in a Mouse Model of Schizophrenia

1076   Predisposition. *Neuron* **89**, 1100–1109 (2016).

1077   33. Spellman, T. *et al.* Hippocampal-prefrontal input supports spatial encoding in working

1078   memory. *Nature* **522**, 309–314 (2015).

1079   34. Maisson, D. J.-N., Gemzik, Z. M. & Griffin, A. L. Optogenetic suppression of the nucleus

1080   reuniens selectively impairs encoding during spatial working memory. *Neurobiol. Learn.*

1081   *Mem.* **155**, 78–85 (2018).

1082   35. Sanderson, D. J. & Bannerman, D. M. The role of habituation in hippocampus-
1083   dependent spatial working memory tasks: Evidence from GluA1 AMPA receptor subunit
1084   knockout mice. *Hippocampus* **22**, 981–994 (2012).

1085   36. Cunniff, M. M., Markenscoff-Papadimitriou, E., Ostrowski, J., Rubenstein, J. L. & Sohal,
1086   V. S. Altered hippocampal-prefrontal communication during anxiety-related avoidance in
1087   mice deficient for the autism-associated gene Pogz. *eLife* **9**, e54835 (2020).

1088   37. Strahnen, D., Kapanaiah, S. K. T., Bygrave, A. M. & Kätzel, D. Lack of redundancy
1089   between electrophysiological measures of long-range neuronal communication. *BMC*
1090   *Biol.* **19**, 24 (2021).

1091   38. Hultman, R. *et al.* Brain-wide Electrical Spatiotemporal Dynamics Encode Depression
1092   Vulnerability. *Cell* **173**, 166-180.e14 (2018).

1093   39. Hultman, R. *et al.* Dysregulation of Prefrontal Cortex-Mediated Slow-Evolving Limbic
1094   Dynamics Drives Stress-Induced Emotional Pathology. *Neuron* **91**, 439–452 (2016).

1095   40. Kirkby, L. A. *et al.* An Amygdala-Hippocampus Subnetwork that Encodes Variation in
1096   Human Mood. *Cell* **175**, 1688-1700.e14 (2018).

1097   41. Woloszyn, L. & Sheinberg, D. L. Neural Dynamics in Inferior Temporal Cortex during a
1098   Visual Working Memory Task. *J. Neurosci.* **29**, 5494–5507 (2009).

1099   42. Johannesen, J. K., Bi, J., Jiang, R., Kenney, J. G. & Chen, C.-M. A. Machine learning
1100   identification of EEG features predicting working memory performance in schizophrenia
1101   and healthy adults. *Neuropsychiatr. Electrophysiol.* **2**, 3 (2016).

1102   43. Combrisson, E. *et al.* From intentions to actions: Neural oscillations encode motor
1103   processes through phase, amplitude and phase-amplitude coupling. *NeuroImage* **147**,
1104   473–487 (2017).

1105   44. Akam, T. *et al.* pyControl: Open source, Python based, hardware and software for
1106   controlling behavioural neuroscience experiments. *bioRxiv* 2021.02.22.432227 (2021)
1107   doi:10.1101/2021.02.22.432227.

1108   45. Teutsch, J. & Kätzel, D. Operant Assessment of DMTP Spatial Working Memory in Mice.
1109   *Front. Behav. Neurosci.* **13**, (2019).

1110   46. Bolkan, S. S. *et al.* Thalamic projections sustain prefrontal activity during working

1111   memory maintenance. *Nat. Neurosci.* **20**, 987–996 (2017).

1112   47. Brincat, Scott *et al.* Which Neural Signals are Optimal for Brain-Computer Interface

1113   Control? doi:10.3217/978-3-85125-260-6-104.

1114   48. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars

1115   of RDoC. *BMC Med.* **11**, 126 (2013).

1116   49. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nat.*

1117   *Neurosci.* **17**, 347–356 (2014).

1118   50. Johnson, E. L., Kam, J. W. Y., Tzovara, A. & Knight, R. T. Insights into human cognition

1119   from intracranial EEG: A review of audition, memory, internal cognition, and causality. *J.*

1120   *Neural Eng.* **17**, 051001 (2020).

1121   51. Majerus, S. *et al.* Cross-Modal Decoding of Neural Patterns Associated with Working

1122   Memory: Evidence for Attention-Based Accounts of Working Memory. *Cereb. Cortex N.*

1123   *Y. NY* **26**, 166–179 (2016).

1124   52. Niki, H. & Watanabe, M. Prefrontal and cingulate unit activity during timing behavior in

1125   the monkey. *Brain Res.* **171**, 213–224 (1979).

1126   53. Niki, H. & Watanabe, M. Cingulate unit activity and delayed response. *Brain Res.* **110**,

1127   381–386 (1976).

1128   54. Gilad, A., Gallero-Salas, Y., Groos, D. & Helmchen, F. Behavioral Strategy Determines

1129   Frontal or Posterior Location of Short-Term Memory in Neocortex. *Neuron* **99**, 814-

1130   828.e7 (2018).

1131   55. Esmaeili, V. & Diamond, M. E. Neuronal Correlates of Tactile Working Memory in

1132   Prefrontal and Vibrissal Somatosensory Cortex. *Cell Rep.* **27**, 3167-3181.e5 (2019).

1133   56. Insanally, M. N. *et al.* Spike-timing-dependent ensemble encoding by non-classically

1134   responsive cortical neurons. *eLife* **8**, e42409 (2019).

1135   57. Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N. & Komiyama, T. Area-Specificity and

1136   Plasticity of History-Dependent Value Coding During Learning. *Cell* **177**, 1858-1872.e15

1137   (2019).

1138   58. Zamanillo, D. *et al.* Importance of AMPA receptors for hippocampal synaptic plasticity

1139   but not for spatial learning. *Science* **284**, 1805–11 (1999).

1140   59. Chudasama, Y. & Robbins, T. W. Dopaminergic Modulation of Visual Attention and

1141   Working Memory in the Rodent Prefrontal Cortex. *Neuropsychopharmacology* **29**, 1628–

1142   1636 (2004).

1143   60. Goto, K. & Ito, I. The asymmetry defect of hippocampal circuitry impairs working memory

1144   in β2-microglobulin deficient mice. *Neurobiol. Learn. Mem.* **139**, 50–55 (2017).

1145   61. Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S. & Mitra, P. P. Chronux: A platform for

1146   analyzing neural signals. *J. Neurosci. Methods* **192**, 146–151 (2010).

1147   62. Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F. & Pennartz, C. M. A. An

1148   improved index of phase-synchronization for electrophysiological data in the presence of

1149   volume-conduction, noise and sample-size bias. *NeuroImage* **55**, 1548–1565 (2011).

1150   63. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software

1151   for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput.*

1152   *Intell. Neurosci.* **2011**, (2010).

1153   64. Dhamala, M., Rangarajan, G. & Ding, M. Analyzing information flow in brain networks

1154   with nonparametric Granger causality. *NeuroImage* **41**, 354–362 (2008).

1155   65. Baccalá, L. A. & Sameshima, K. Partial directed coherence: a new concept in neural

1156   structure determination. *Biol. Cybern.* **84**, 463–474 (2001).

1157   66. Dhamala, M., Rangarajan, G. & Ding, M. Estimating Granger Causality from Fourier and

1158   Wavelet Transforms of Time Series Data. *Phys. Rev. Lett.* **100**, 018701 (2008).

1159   67. Dhamala, M., Liang, H., Bressler, S. L. & Ding, M. Granger-Geweke causality:

1160   Estimation and interpretation. *NeuroImage* **175**, 460–463 (2018).

1161   68. Canolty, R. T. & Knight, R. T. The functional role of cross-frequency coupling. *Trends*

1162   *Cogn. Sci.* **14**, 506–515 (2010).

1163   69. Sirota, A. *et al.* Entrainment of Neocortical Neurons and Gamma Oscillations by the

1164   Hippocampal Theta Rhythm. **60**, 683–697 (2008).

1165 70. Tort, A. B. L. *et al.* Dynamic cross-frequency couplings of local field potential oscillations

1166   in rat striatum and hippocampus during performance of a T-maze task. *Proc. Natl. Acad.*

1167   *Sci.* **105**, 20517–20522 (2008).

1168 71. Voytek, B. *et al.* Shifts in Gamma Phase–Amplitude Coupling Frequency from Theta to

1169   Alpha Over Posterior Cortex During Visual Tasks. *Front. Hum. Neurosci.* **4**, (2010).

1170 72. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic

1171   minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

1172 73. Skurichina, M. & Duin, R. P. W. Bagging, Boosting and the Random Subspace Method

1173   for Linear Classifiers. *Pattern Anal. Appl.* **5**, 121–135 (2002).

1174 74. Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE*

1175   *Trans. Pattern Anal. Mach. Intell.* **20**, 832–844 (1998).

1176 75. Combrisson, E. & Jerbi, K. Exceeding chance level by chance: The caveat of theoretical

1177   chance levels in brain signal classification and statistical assessment of decoding

1178   accuracy. *J. Neurosci. Methods* **250**, 126–136 (2015).

1179 76. Strahnen, D., Kapanaiah, S., Bygrave, A. M. & Kätzel, D. KaetzelLab/GluA1-

1180   KO_LFP_data. *Ger. Neuroinformatics Node* (2020) doi:10.12751/g-node.3lvs3q.

1181 77. Strahnen, D. & Kätzel, D. KaetzelLab/LFP_analysis. *Github* (2020)

1182   doi:10.5281/zenodo.4384980.

1183

1184