

1 **Measuring instability in chronic human intracortical neural recordings towards stable, long-term brain-**
2 **computer interfaces**

4 Tsam Kiu Pun^{1,2,3*}, Mona Khoshnevis⁴, Tommy Hosman^{2,5}, Guy H. Wilson⁹, Anastasia Kapitonava⁶, Foram
5 Kamdar⁹, Jaimie M. Henderson^{9,10}, John D. Simeral^{5,2,3}, Carlos E. Vargas-Irwin^{3,5,7}, Matthew T. Harrison^{3,4**}, Leigh
6 R. Hochberg^{2,3,5,6,8**}

7 ¹ Biomedical Engineering Graduate Program, School of Engineering, Brown University, Providence, RI, USA.

8 ² School of Engineering, Brown University, Providence, RI, USA.

9 ³ Carney Institute for Brain Science, Brown University, Providence, RI, USA.

10 ⁴ Division of Applied Mathematics, Brown University, Providence, RI, USA.

11 ⁵ VA RR&D Center for Neurorestoration and Neurotechnology, Rehabilitation R&D Service, Providence VA
12 Medical Center, Providence, RI, USA.

13 ⁶ Center for Neurotechnology and Neurorecovery, Dept. of Neurology, Massachusetts General Hospital, Boston,
14 MA, USA.

15 ⁷ Department of Neuroscience, Brown University, Providence, RI, USA

16 ⁸ Department of Neurology, Harvard Medical School, Boston, MA, USA.

17 ⁹ Department of Neurosurgery, Stanford University, Stanford, CA, USA.

18 ¹⁰ Wu Tsai Neurosciences Institute and Bio-X Institute, Stanford University, Stanford, CA, USA.

19 ** These authors jointly supervised this work

20 * Corresponding author, Email: tsam_kiu_pun@brown.edu

21 **Abstract**

22 Intracortical brain-computer interfaces (iBCIs) enable people with tetraplegia to gain intuitive cursor control from
23 movement intentions. To translate to practical use, iBCIs should provide reliable performance for extended
24 periods of time. However, performance begins to degrade as the relationship between kinematic intention and
25 recorded neural activity shifts compared to when the decoder was initially trained. In addition to developing
26 decoders to better handle long-term instability, identifying when to recalibrate will also optimize performance.
27 We propose a method to measure instability in neural data without needing to label user intentions. Longitudinal
28 data were analyzed from two BrainGate2 participants with tetraplegia as they used fixed decoders to control a
29 computer cursor spanning 142 days and 28 days, respectively. We demonstrate a measure of instability that
30 correlates with changes in closed-loop cursor performance solely based on the recorded neural activity (Pearson ρ
31 = 0.93 and 0.72, respectively). This result suggests a strategy to infer online iBCI performance from neural data
32 alone and to determine when recalibration should take place for practical long-term use.

36 **Introduction**

37

38 Intracortical brain-computer interfaces (iBCIs) have enabled people with tetraplegia to control external devices by
39 decoding movement intentions from neural recordings^{1–7}. iBCIs can also restore communication by providing
40 rapid point-and-click cursor control for applications such as typing, web browsing and navigating apps on a tablet
8–13, and can enable speech-to-text decoding for people with severe dysarthria¹⁴. Decoders are typically trained
41 during explicit calibration epochs that allow for simultaneous collection of recorded neural signals during
42 instructed motor intentions^{6,15–17}. After training, decoding performance varies over time because of complex
43 biological and device-related instabilities that are not fully understood^{18,19}. Persistent periods of decreased
44 performance are commonly observed with existing decoding paradigms and remain one of the challenges that
45 hinders wider adoption of iBCIs for people with paralysis^{12,18–22}. Restoration of good control after performance
46 has degraded often requires the user to repeat a calibration task to retrain the decoder^{5,18}. Reducing the
47 frequency and duration of explicit recalibration tasks is important for improving the utility of iBCIs. A step in this
48 direction would be a method that could monitor performance and automatically determine when recalibration or
49 other measures were necessary.

50

51 Here, we show that decoding performance on timescales of tens of seconds can be estimated from the same
52 recorded neural signal used for motor decoding. The main idea is that persistent changes in performance likely
53 result from statistical changes of some kind in the recorded neural signals. Measures of statistical changes might
54 then be a good surrogate for measures of performance changes. We call this approach **MINDFUL** (measuring
55 instabilities in neural data for useful long-term iBCI). More specifically, given a target period for which average
56 decoding performance is unknown, we calculate a statistical distance between the distribution of neural activity
57 patterns during the target period and a similar distribution collected when performance was known to be good
58 (such as when the decoder was first trained), as illustrated in **Fig. 1a**. The MINDFUL score obtained using
59 Kullback-Leibler divergence (KLD) to compare neural activity patterns was found to correlate with decoding
60 performance.

61

62 Changes in decoding performance can be attributed to many types of variability in the recorded neural signals.
63 BCI decoding algorithms typically model rapid fluctuations in neural features (on timescales of tens of
64 milliseconds) that have no apparent correlation with motor intention as stochastic *noise*. Noise is a useful
65 explanation for why decoding performance changes from instant to instant, but it generally does not account for
66 persistent changes in (average) decoding performance that last seconds or more. We ascribe such persistent
67 changes to *model drift*, which we define as changes in the relationship between recorded neural signals and motor
68 intention. Nonstationarity, feature shift, dataset shift are terms that have been used synonymously in the
69 literature for this type of phenomenon, but they sometimes refer to any changes in the recorded signals rather
70 than changes in the signal-decoder relationship^{23–30}. Model drift can be attributed to various factors such as
71 changes in action potential waveforms^{18,31,32}, neural tuning profiles^{33–35}, cognitive strategy or plasticity due to
72 learning^{36–38}, material degradation and tissue responses to the recording device^{39,40}, and array micro-movements
73⁴¹. The type and magnitude of model drift results in various forms of performance degradation¹⁹, sometimes
74 necessitating decoder recalibration to restore control. Existing solutions to reduce the need for recalibration tasks
75 include adaptive decoders that require shorter recalibration sessions to maintain or restore stable performance
76^{3,10,42}, self-supervised recalibration using retrospective labeling that avoids explicit recalibration sessions^{26,27,43,44},
77 and robust decoders that experience less model drift by extracting stable, time-invariant features from high-
78 dimensional recordings^{20–22,45–51} or by adaptively adjusting decoder parameters^{12,52}.

79

80 The model drift that influences performance is necessarily a property of the joint distribution of recorded neural
81 signals and motor intention. It is not *a priori* clear, however, that model drift related to performance can be
82 meaningfully identified from the recorded neural signals alone, which is what MINDFUL attempts to capture.
83 MINDFUL differs from previously studied statistical tests for model drift that additionally require knowledge of
84 movement intention^{28,53,54}. Since true movement intention is often unavailable in iBCI applications where people
85 with paralysis control an external device without being cued to acquire targets (e.g., a cursor on a tablet computer
86 being used to send an email¹²), an approach like MINDFUL based only on the (marginal) distribution of recorded
87 neural activity is much more widely applicable.

88

89
90 Here, we present and validate an approach to predict closed-loop decoding performance without the knowledge
91 of true movement intention. MINDFUL was applied on longitudinal datasets where performance changed over
92 long time periods as two people with tetraplegia, designated as T11 and T5, were using an iBCI. Each participant
93 used an iBCI to control a computer cursor to perform target acquisition tasks on a screen. Target acquisition tasks
94 permit observation of (presumed) motor intention and, hence, can be used to directly measure decoding
95 performance. The kinematic decoders were held fixed across all sessions so that persistent changes in
96 performance could be ascribed to model drift and not to changes associated with the decoder. Briefly, MINDFUL
97 represents changes in neural distributions relative to a reference distribution where the decoder was initially
98 applied. MINDFUL is solely based on the recorded neural activity, without requiring information about the target
99 locations. The resulting MINDFUL score was highly correlated with changes in closed-loop cursor performance
100 over time.

101
102 **Results**
103

104 **Fixed decoders result in initially stable and then unstable performance across month-long sessions**
105 To first establish a baseline for decoder performance, we deployed fixed decoders^{27,51} for the purpose of
106 identifying, over a comparatively long period, how neural instabilities may lead to deteriorating control. Data
107 were collected from 15 research sessions, each from a separate day that spanned across 142 days, of T11
108 performing a center-out-and-back task using a fixed RNN (recurrent neural network) decoder, as previously
109 described⁵¹ (see **Methods**). As part of another study²⁷, T5 performed a random target task for six sessions
110 spanning 28 days using a fixed linear decoder (see **Methods**). To quantify closed-loop cursor performance, *angle*
111 *error* (AE) between the inferred intended directional vector (cursor-to-target position) and the decoded velocity
112 vector was used (see **Methods**). AE is a valuable metric for capturing performance as it is sensitive to
113 instantaneous cursor direction change, and can be averaged across any range of time. T11 achieved stable, high
114 performance online cursor control for the first three months. The median AE per trial for T11 for sessions during
115 the first three months was lower than later sessions on average (trial day 658-751: $26.8^\circ \pm 22.6^\circ$; trial day 758-800:
116 $88.4^\circ \pm 46.1^\circ$; $p < 0.001$, Wilcoxon rank sum). For T5, the first three sessions demonstrated lower AE than the later
117 three sessions (trial day 2121-2128: $39.6^\circ \pm 23.9^\circ$; trial day 2133-2149: $58.8^\circ \pm 31.7^\circ$; $p < 0.001$, Wilcoxon rank sum).
118 Brief recovery from decrease in performance was observed in both participants (93 days after the initial session
119 for T11 and 28 days after the initial session for T5), indicating fixed decoders may not necessarily result in a steady
120 decline in cursor control over time (**Supplementary Fig. 1**).
121

122 **Comparing distributions of neural activity patterns**

123 MINDFUL is based on comparing the distribution of recorded neural activity patterns in a target dataset (usually
124 with unknown decoding performance) to a similar distribution in a reference dataset (usually with known and
125 good decoding performance). **Fig. 1a** provides an illustration. The choices of neural features and measure of
126 statistical dissimilarity are important for practical use. Here we used a measure of statistical dissimilarity based on
127 the well-known Kullback-Leibler divergence (KLD). In principle other measures of dissimilarity could be used (see
128 **Methods**). Neural features were all derived from the underlying inputs to the kinematic decoder: threshold-
129 crossing spike rate and spike power for T11 and spike rate only for T5, extracted in 20 ms non-overlapping bins
130 (see **Methods**).
131

132 **Statistical distance between neural activity patterns correlates with performance**

133 Having established datasets where fixed decoders result in periods of both stable and fluctuating closed-loop
134 performance, we first investigated the underlying premise of MINDFUL that the distribution of neural activity
135 patterns varies systematically with decoder performance. First, neural features pooled from all sessions were
136 categorized into groups based on performance in terms of instantaneous (20 ms) AE. KLD was computed to
137 assess the differences in neural feature distribution at instances with low AE ($< 4^\circ$) to other distributions at
138 instances with varying levels of AE (see **Methods**). For both participants, the relationship between the KLD and
139 performance was found to be remarkably linear and strongly correlated (T11: Pearson $\rho = 0.985$, $p < 10^{-33}$; T5:
140 Pearson $\rho = 0.983$, $p < 10^{-31}$; see **Fig. 1b**). Neural feature distributions of instances with low AE demonstrated high

141 similarity (lower KLD) to the reference distribution of neural features from low AE instances, and the KLD
142 increased linearly as the compared neural feature distributions were drawn from instances with larger AE.
143

144 This is a proof-of-concept that statistical distance between distributions of neural activity patterns can correlate
145 strongly with decoding performance. It does not, however, provide a measure of performance that would be
146 useful in a clinical setting for detecting persistent changes in decoding performance that might arise from model
147 drift. Instantaneous AE must be known a priori to define the collections of neural features that are compared in
148 each point in **Fig. 1b** and MINDFUL is designed to be used in situations where AE is not known, at least not for the
149 target distribution. Moreover, AE can be a result of noise (transient variability) or model drift (persistent changes)
150 or both. **Fig. 1b** does not distinguish among these even though model drift is the phenomenon of interest here.
151 The linear relationship observed in **Fig. 1b** can be recreated in simulation using noise or using model drift or using
152 both (see **Supplementary Fig. 2**).
153

154 **MINDFUL correlates with decoding performance**

155 Towards developing a predictor of decoder performance based on neural activity for online application, we define
156 a measure called the MINDFUL score to study the effect of drifts that persist over timescales relevant to chronic
157 continuous iBCI use. Using the same concept as illustrated in **Fig. 1a**, but instead of grouping by AE as in **Fig. 1b**,
158 neural feature distributions were estimated from collections of time bins aggregated using a 60-second sliding
159 window, regardless of the performance during that window. The reference distribution is also estimated from
160 instances of low AE as in **Fig. 1b**, but it is sub-selected from only the initial session(s) when the decoder is first
161 deployed. As time progresses, we update the MINDFUL score which is based on the KLD between the reference
162 distribution and the subsequent neural feature distributions from the sliding window (see **Methods**). To validate
163 this method, MINDFUL score is correlated against the median AE calculated in the same 60-second sliding
164 intervals as in estimating the neural distributions. Strong correlations were found between MINDFUL score and
165 median AE across sessions, especially for T11 (See **Fig. 1c**. T11: Pearson $\rho = 0.91$, $p < 0.001$; Spearman $\rho = 0.90$, $p <$
166 0.001 ; T5: Pearson $\rho = 0.59$, $p < 0.001$; Spearman $\rho = 0.60$, $p < 0.001$; see **Methods**). This suggests that MINDFUL
167 can be a viable measure for tracking performance in real-time, as the statistical properties of neural features
168 aggregated over a longer timescale window reflects information about how the decoder performance drifts over
169 time without needing to know performance (AE) in advance.
170

171 **Correlation to performance increases by combining neural features and decoder outputs.** The neural features
172 (NF) in **Fig. 1b-c** were derived from principal components analysis (PCA; see **Methods**) and need not reflect the
173 sources of variability most predictive of decoder performance. Using features more closely related to the decoder
174 output might strengthen the relationship between the MINDFUL score and AE. One such feature is the output
175 from the decoder itself – in this case, the predicted 2-dimensional velocity vector, \hat{v} . Note that \hat{v} alone cannot be
176 used to define AE. It is the relationship between \hat{v} and the true intended direction that defines AE. Nevertheless,
177 changes in the distribution of \hat{v} over a time interval may reflect changes in decoder performance. Likewise to **Fig.**
178 **1c**, a 60-second sliding window when estimating the mean and the variance of the distributions. Using only the 2-
179 dimensional feature \hat{v} showed reduced correlations between the MINDFUL score and AE than from just NF for
180 T11, but increased for T5 (**Fig. 2a**. T11: Pearson $\rho = 0.464$, $p < 0.001$; Spearman $\rho = 0.407$, $p < 0.001$; T5: Pearson $\rho =$
181 0.704 , $p < 0.001$; Spearman $\rho = 0.763$, $p < 0.001$). When using the 4-dimensional feature created by concatenating
182 \hat{v} and \hat{v}_{prev} , where \hat{v}_{prev} comes from the previous time bin of \hat{v} (20ms earlier), there was an increased correlation to
183 AE for both T11 and T5 (**Fig. 2b**. T11: Pearson $\rho = 0.819$, $p < 0.001$; Spearman $\rho = 0.840$, $p < 0.001$; T5: Pearson $\rho =$
184 0.702 , $p < 0.001$; Spearman $\rho = 0.765$, $p < 0.001$). Lastly, the combination of inputs and outputs of the decoder, i.e.
185 the low-dimensional NF, \hat{v} , and \hat{v}_{prev} resulted in the highest correlation between KLD and AE for both participants
186 (**Fig. 2c**. T11: Pearson $\rho = 0.926$, $p < 0.001$, Spearman $\rho = 0.913$, $p < 0.001$, T5: Pearson $\rho = 0.719$, $p < 0.001$,
187 Spearman $\rho = 0.759$, $p < 0.001$).
188

189 **The MINDFUL score reflects changes in feature tuning.** We next investigated the types of changes in neural
190 data captured by the MINDFUL score. Changes in directional tuning have been shown to reduce performance in
191 both online and offline BCI studies^{10,12,18–22}. Directional tuning was quantified by fitting a cosine function to
192 normalized neural features to obtain estimates of preferred direction (PD) and modulation depth (MD)^{33,55–57}. 154

193 out of 384 features and 85 out of 192 features, for T₁₁ and T₅ respectively, had significant directional tuning for at
194 least half of all recording sessions (F-test, $p < 0.05$, see **Methods**). Changes in tuning in these features were
195 tracked over time (see **Methods**). (Although our analysis here is focused on the tuning of *individual* features, in the
196 supplement we use a different approach to show that changes in the conditional distribution of *population* activity
197 given motor intention are statistically significant; see **supplementary Fig. 3**.)
198

199 Tuning properties shifted gradually for the majority of features in T₁₁ (**Fig. 3a-b**). 125 out of 151 tuned features
200 exhibited significant change in both MD and PD³⁵ in at least one session (see **Methods**). Fitted tuning curves
201 across sessions for two example features illustrated changes in modulation depth and modulation lost,
202 respectively for T₁₁ (**Fig 3c**). The average change in PD across these features was larger on days where
203 performance was worse (day 7-93: $46.8^\circ \pm 31.2^\circ$; day 100-142: $62.4^\circ \pm 34.3^\circ$; $p < 10^{-7}$ Wilcoxon rank sum). Average
204 absolute change in MD in later sessions was also found to be significantly larger (day 7-93: 0.107 ± 0.067 ; day 100-
205 142: 0.159 ± 0.129 ; $p < 10^{-8}$, Wilcoxon rank sum).

206 Similar to T₁₁, gradual changes in T₅'s tuning properties were also observed (**Fig. 3d-e**). Some features illustrated
207 changes in either MD or PD, or both (**Fig 3f**). 71 out of 85 tuned features exhibited significant change in both MD
208 and PD in at least one session. The average change in PD across these features was larger on days where
209 performance was worse (day 12 and 14: $69.9^\circ \pm 40.7^\circ$; day 5, 7, and 28: $58.1^\circ \pm 36.6^\circ$; $p = 0.0346$, Wilcoxon rank
210 sum). Average absolute change in MD was not found to be significant (day 12 and 14: 0.147 ± 0.099 ; day 5, 7, and
211 28: 0.139 ± 0.093 ; $p = 0.664$, Wilcoxon rank sum).
212

213 To quantify the changes in encoding on a population level, we use tuning maps^{56,58}, defined by matrices of fitted
214 tuning parameters of significantly tuned features on each session. Tuning similarity between days was assessed
215 by calculating the correlation between the corresponding tuning maps (see **Methods**). In general, nearby sessions
216 in time with similar performance were more correlated. For T₁₁, tuning maps among early sessions (up to day 93)
217 with high performance were highly correlated, as well as among later sessions with low performance, but not
218 across these two epochs (**Fig. 3g**). For T₅, sessions that were closer together in time (along the diagonal) had
219 higher correlation than those further apart (**Fig. 3h**). These results suggest that model drift (tuning changes)
220 occurred across sessions. We were thus interested in determining how the MINDFUL score reflects changes in
221 tuning similarity. To compare with the tuning map correlation between sessions, we obtain a mean KLD between
222 each pair of sessions. Instead of fixing a reference distribution, pairwise KLDs of neural features between sessions
223 were calculated using a sliding window of 60 seconds updating every 10 seconds (see **Methods**). The KLDs from
224 the same session were averaged to get a mean of the neural distribution difference between pairs of sessions.
225

226 For T₁₁, pairs of sessions closer in time had smaller distribution shifts, while pairs of sessions further from each
227 other in time had larger distribution shifts in terms of KLD, which consequently strongly correlated with tuning
228 map correlation (Pearson $\rho = -0.812$, $p < 10^{-30}$, **Fig. 3i**). For T₅, the same trend was observed, except for the pairs
229 of sessions which compared the first three sessions to the last session where cursor control had recovered (dots in
230 blue shades). The correlation between mean KLD and tuning map correlation was also strong and significant ($\rho = -$
231 0.776 , $p < 10^{-4}$, **Fig. 3j**). Together, these findings suggested that the MINDFUL score using KLD captures day-to-
232 day changes in directional tuning, even though the metric can be calculated without information reflecting target
233 position or movement intentions.
234

235 **The MINDFUL score captures low-dimensional neural latent space drifts.** We further investigated how this
236 method relates to the changes in the low-dimensional neural latent space using demixed principal component
237 analysis⁵⁹. The top two direction-dependent principal components (PCs) on neural population from decoder day 0
238 were calculated to compare changes across sessions by projecting the neural population from subsequent
239 sessions on this PC space (see **Methods**). The average neural trajectories per target directions became less
240 distinct as the session dates progressed, reflecting changes in the underlying population activity over time
241 consistent with the decline in task performance (see **Fig. 4**). The amount of direction-related neural activity in
242 each session was quantified by the variance accounted for (VAF) by the top two direction-dependent components
243 on the subspace of day 0. For T₁₁, the VAF was initially 50.0% on day 0 and remained above 20% on days in which
244 clear separation of target trajectories were observed. As the decoder performance declined, the VAF dropped to
245

246 4.4% on decoder day 142 (**Fig. 4a**). This change in neural representation in low-dimensional space is strongly and
247 significantly correlated to the mean KLD per session (Pearson $\rho = -0.892$, $p < 10^{-5}$). Mean KLD between day 0 and
248 other sessions is calculated the same way as the mean KLD in the previous section (averaged KLDs using a sliding
249 window of 60 seconds updating every 10 seconds, see **Methods**). For T5, VAF was initially 42.2% on day 0, and
250 subsequently dropped to 2.9% and recovered to 11.3% on the last session (**Fig. 4b**). There was also a strong and
251 significant correlation between the top 2 VAF and mean KLD (Pearson $\rho = -0.858$, $p = 0.0029$).
252

253 In addition to correlating with model drift in neural data, MINDFUL was found to detect large momentary
254 deviations in the signal, likely attributable to device-related reasons such as signal transmission errors⁵. The sharp
255 spikes in both KLD and median AE in T11 neural data (**Fig. 2**) correspond to time steps during outlier trials (see
256 **Supplementary Fig. 4**). Outlier trials were defined as having more than a 5% drop of wireless neural data packets
257 or large “neural” responses greater than 8 standard deviations from the mean. Furthermore, when excluding
258 these trials, the MINDFUL score was still highly correlated to the median AE (see **Supplementary Fig. 5**). This
259 suggests that our method is capable of tracking both model drift over time, as well as short timescale technical-
260 related variability. Although instantaneous events are less relevant for decoder recalibration, a method to capture
261 these events may prove useful in other iBCI troubleshooting with both the current and future fully implanted
262 systems.
263

264 **Selecting reference and window length further optimizes correlation** We explored the role of sub-selecting
265 time bins with different AE ranges as the reference in the **MINDFUL** pipeline. When limiting the reference to
266 collection of time bins with low AE only (0-4°) as shown in **Fig 1c, 2**, there are strong correlations between the
267 KLD of derived neural features ($NF + \hat{X} + \hat{X}_{lag}$) to AE. It was higher than when using all time steps of any cursor
268 control quality for both participants (see **Fig. 5a**). This also held true when taking other combinations of derived
269 neural features into calculating KLD (**Supplementary Table 1**). In addition, sub-selecting instances with high AE
270 (50°-100°, 100°-180°) as the reference distribution reduced the KLD-AE correlation for both participants,
271 especially in T5.
272

273 Another important consideration for improving the MINDFUL correlation to performance is the duration of neural
274 data required to obtain a reasonable representation of neural distribution. Since neural activity recorded in the
275 precentral gyrus modulates with the direction of intended movement, too small of a window length may reflect
276 task-dependent differences in reaching to different directions, rather than a persistent model drift that affects
277 cursor control regardless of trial direction. A longer window length can avoid this issue when directional
278 distribution differences are averaged out over a longer period. However, the KLD would be smoothed out if
279 window length is too long, resulting in a larger delay to detect the need to update the decoder. To balance
280 between accuracy and efficiency for online implementation, the optimal duration was empirically determined to
281 be at least 60 seconds (**Fig. 5b**), where KLD-AE correlation began to plateau, and longer windows did not offer a
282 higher correlation. Using a 60-second window to estimate neural distributions provides a sufficiently large
283 number of samples to average out directional-dependent differences due to variations in trial-to-trial movement
284 directions.
285

286 **MINDFUL is robust to the reference task.** We repeated the analysis of tracking the correlation of
287 MINDFUL to performance (see **Fig 2c**) for T11 except we used reference data collected during different
288 tasks. The comparison data collected during center-out-and-back tasks remains the same as in **Fig 2c**.
289 Since the relationship between neural activity and movement can be context and task dependent^{60,61}, it is
290 unclear to what degree the reference and comparison tasks must be matched for MINDFUL to correlate
291 well to performance. MINDFUL is likely to be most useful in practice if it is robust across tasks and contexts.
292 Collecting reference data from a different task was not part of our original experimental design, but for T11
293 the appropriate data was collected for other purposes. In addition to the center-out-and-back task, T11 used
294 the same fixed decoder for random target tasks (day 7), analogous to T5’s task described above, and during
295 personal iBCI use (i.e. browsing the web; day 0; See **Methods**). Despite different cursor tasks being used
296 as the reference from the target distributions, MINDFUL still correlates highly to performance, even without
297 the help of subsampling based on AE (see **Fig 6b-c**). This is also true when all three types of tasks were

298 combined together for estimating the reference distribution (mixed tasks, see **Fig 6d**). The range of KLD is
299 slightly higher for the random target task, and lower for personal use and mixed tasks. This suggests that
300 MINDFUL might be robust to cursor task changes for reference.

301

302 Discussion

303 Apparent model drift during chronic iBCI use – resulting from changes in the information encoded in neural
304 ensembles, changes in the recorded neural elements themselves, or changes in the recording devices – are one of
305 the major challenges for developing decoders that will provide stable, accurate cursor control for long-term use by
306 people with paralysis. Existing solutions to mitigate more substantial model drift have limitations. Explicit
307 recalibration comes at the expense of interrupting the user in the midst of iBCI use to collect additional data for
308 training. While self-supervised recalibration doesn't require the user to perform daily repetition of a task, it relies
309 on stable online performance for effective pseudo-labeling. Currently, one cannot predict accurately the moment
310 at which the decoder may fail to sustain performance and thus require a supervised recalibration. Similarly, robust
311 algorithms reduce the need for frequent retraining but may require a retraining of the model *de novo*. MINDFUL
312 fills in the gap in the development of a better decoder recalibration strategy for practical everyday use, by
313 identifying, quantifying, and monitoring the degree of neural recording instability that contributes to degradation
314 of real-time decoder performance.

315

316 The MINDFUL score, which is based on measuring the Kullback-Leibler divergence (KLD) between distributions of
317 neural features, reflects online performance without the need to incorporate knowledge of intended targets. In
318 two participants, the MINDFUL score was strongly correlated to online angle error during iBCI cursor control
319 across session days spanning up to four months (T11) or one month (T5). With the goal of translating this method
320 to an online setting for the purpose of personal iBCI use, the MINDFUL score can reflect performance accurately
321 for different cursor tasks examined in different participants (**Fig. 2**). Importantly, the MINDFUL score is consistent
322 with tuning and latent space changes which can't be directly measured without information about movement
323 intention. This suggests that the MINDFUL score provides an intrinsic measure to track model drift affecting
324 decoder performance during long-term iBCI use.

325

326 Our study confirmed the well-acknowledged observation that recording instabilities can impact online
327 performance when the decoder cannot accommodate neural changes over long-term iBCI use. Model drift was
328 quantified by tracking changes in tuning and latent space representations of neural population activity across
329 sessions. MINDFUL was found to be highly correlated with both of these measures. It should be noted that our
330 method did not track mean firing rate shifts which are known to correlate with declines in decoder performance
331¹⁹. In our datasets, adaptive mean correction such as z-scoring or bias correction were applied to the neural
332 features during online cursor control to combat this type of model drift (see **Methods**). Therefore, performance
333 drops observed in this dataset were largely due to other types of model drift. The MINDFUL score, which
334 measures changes in the distribution of z-scored neural data, aims to discover model drift such as changes in
335 tuning or latent representation which may necessitate a decoder recalibration to restore control. Furthermore,
336 since MINDFUL was designed to be applied during online iBCI control where threshold crossings were primarily
337 used as neural features^{62–65}, we chose to investigate the functional stability in the thresholded neural activity'
338 tuning properties and latent neural representation rather than neuronal stability of discriminated single units
339 processed via spike sorting techniques^{18,31}.

340

341 Another plausible contributor to the observed changes in neural distributions may be participants' compensatory
342 neuro-motor strategies in response to suboptimal cursor control. Non-human primates that encounter artificial
343 perturbations in a previously learned BCI motor control task elicit new neural patterns with learning^{54,66–68}. Neural
344 activity during closed-loop, online control contexts are also different from open-loop, offline control which has
345 added real-time visual feedback^{69,70}. In this closed-loop iBCI study, when participants experienced a directional
346 bias in cursor kinematics, it is plausible that the participants might compensate for decoding errors with different
347 strategies such as moving against the bias (eliciting a larger magnitude of velocity), temporarily pausing
348 attempting movement control, or moving towards the bias in hope to reset the bias (when automatic bias
349 correction is applied). These alternative strategies are valid responses in attempting to improve control, but they

350 may result in a larger change in distribution of neural features, amplifying the original model drift when intention
351 context remains consistent. This highlights one of the challenges when studying model drift during closed-loop
352 control when the ground-truth intended movement cannot be observed independently of the decoded outputs.
353

354 While the MINDFUL score based on KLD consistently correlates with performance and changes in neural
355 representation, there are a number of noticeable differences in the results between subjects. First, the relevance
356 of the chosen neural features appears to be different for different participants. For example, for participant T₁₁,
357 the MINDFUL score derived from the neural features is more strongly correlated to the AE than the MINDFUL
358 score derived from decoded velocity features, slightly improving when combining both features; for participant
359 T₅, the MINDFUL score derived from the decoded velocity features is more strongly correlated to the AE than
360 neural features alone and is not substantially improved by adding the neural features (Fig 1c and 2). The features
361 used, choice of decoder and cursor task, and duration of data collection, can all influence the value of KLD, hence
362 correlation to AE. Second, there are noticeable differences in the range of KLD between subjects ([0, ~2.75] for
363 T₁₁ and [0, ~0.9] for T₅ in Fig 1c). It is possible that variability between subjects and between the dataset may
364 affect the range of KLD in MINDFUL. Interestingly, the range of KLD calculated by binning by performance is
365 much smaller than calculated across time ([0, ~0.3] for both participants; Fig 1b). In Fig 1b, since data was
366 collected across a wider range of time, both for the reference and comparison distributions, any model drift would
367 likely cause these distributions to have larger variance and, hence, smaller KLD. In support of this conjecture, we
368 found that the determinant of the empirical covariance matrix of the reference distribution by binning by
369 performance (Fig 1b) is 3.4 times larger than the reference set to the first session (Fig 1c) for T₁₁, whilst for T₅,
370 these determinants are relatively equal. Despite a number of methodological differences between the datasets
371 and chosen features, it's encouraging that MINDFUL robustly measured model drift for both participants.
372

373 Choosing an appropriate reference when calculating KLD influences the reliability of decoder performance
374 predictions using MINDFUL. First, selecting low AE time steps for reference was found to provide a higher
375 correlation between the KLD and AE (**Fig. 5a**). Using low AE as reference helps to identify the model drift where
376 the neural-kinematics distribution during decoder training has changed from that during testing when the
377 decoder was applied online. The training data of the decoder typically represents periods of relatively high
378 performance: For T₁₁, the LSTM decoder was trained on selected historical data with angle error < 45°, while T₅'s
379 decoder was initially trained on open loop blocks then immediately updated with a closed-loop block ^{27,51}. As
380 neural representations shift from the training distribution, the decoder is more likely to produce a subpar
381 performance with a higher error rate. Therefore, when sub-selecting only high-performance data as the baseline,
382 future neural shifts can be more accurately reflected by the KLD. However, there exists subject variability and
383 ambiguity regarding precisely how much data are needed for reference. For example, for participant T₅, using
384 timesteps with high AE (50°-100° or 100°-180°) as reference resulted in a more drastic decrease in the correlation
385 between KLD and AE than T₁₁. Also, using the first two sessions as reference resulted in a slightly higher Pearson
386 correlation than just the first session alone for T₅ (**Supplementary Fig. 6**). Nevertheless, assuming that a newly
387 trained decoder returns decent initial performance, our findings suggest that data from high-performance time
388 steps from the initial session where the decoder was first applied would be an appropriate choice for the
389 reference.
390

391 In this study, the MINDFUL score based on neural activity alone reflects performance in cases where fixed
392 decoders were used online. But it remains unclear how it can be applied to other types of adaptive or robust
393 decoders that aim to stabilize decoding by periodically realigning neural data to the initial session. In such
394 circumstances, if MINDFUL is calculated before data alignment, it may not directly correlate to performance as
395 adaptive alignment may keep AE low even when neural representations are changing. However, MINDFUL could
396 still be useful in several ways. First, MINDFUL may be applied on transformed neural data after manifold
397 alignment methods. Even if alignment approaches will result in a reduction in KLD (KLD is a common choice of
398 loss function), MINDFUL can measure the remaining differences. Second, if the features used by MINDFUL are
399 the outputs from the adapted decoder, rather than non-adapted features such as PCA components, then
400 MINDFUL might continue to correlate to online performance despite decoder adaptation. Lastly, instead of
401 periodically aligning (or recalibrating), MINDFUL can be used to signal the need for recalibration when model drift

402 is detected. Future experiments with other closed-loop adaptive decoders will be required to test these
403 approaches.

404

405 We proposed a statistical method to detect model drift when fixed decoders were used during consecutive days-
406 to-months of iBCI cursor control by two people with tetraplegia. This is crucial towards the goal of clinical
407 translation of iBCI systems for practical everyday use, as it requires stable and reliable decoders to maintain high
408 performance despite drifts in neural representations over time. MINDFUL was shown to be able to track model
409 drift based on the intrinsic properties of neural features and decoder outputs, which correlates to long-term
410 changes in decoder performance, without needing to be aware of the movement intention. This approach is well-
411 suited for future online adaptive iBCI systems aiming to provide continuous long-term control in a practical,
412 personal setting outside of standardized research sessions, where it is not possible to directly track intended
413 movements. For instance, MINDFUL and related methods could be used to trigger either a user-engaged or
414 background update as the decoder begins to degrade.

415

416 There are additional several considerations of applying the MINDFUL score online. First, during personal iBCI use,
417 the movement directions could be less symmetric and more sparsely distributed than the cursor tasks used in this
418 study. When sub-selecting reference time bins based on movement direction (up/down/left/right), KLD became
419 higher in magnitude and generally less correlated to AE (see **Supplementary Fig. 10, 11**). A longer time window
420 or careful time bin selection for both reference and target distributions may be needed to reduce the directional-
421 related differences. In addition, we believe that the MINDFUL approach may be useful for detecting changes in
422 the relationship between the signal and the decoder(s), even when the multiple disparate tasks and contexts are
423 incorporated into the reference set. For T11, MINDFUL was found to be robust to using neural reference data
424 collected during different cursor tasks, including periods of personal iBCI use for which we had no control over the
425 balance of intended directions or angle error. We did not have the data to investigate this in T5. Future work will
426 investigate this robustness in additional participants and more varied changes in tasks and contexts, all of which
427 will be important for practical iBCI use.

428

429 Second, as previously described, the range of KLD varies between participants. It will be crucial to set an
430 appropriate threshold for triggering a recalibration for individual users. One possible strategy to set a user-specific
431 threshold would be to initialize a threshold based on a AE cut-off from previously collected datasets and
432 iteratively fine-tune the threshold sensitivity by incorporating user's feedback. Lastly, the abovementioned large
433 noise instances which can be easily detected by MINDFUL should not trigger a recalibration, as it does not imply a
434 change in the neural-kinematic relationship estimated by the decoder. The frequency or pattern of these events
435 could, however, be valuable in further iBCI development.

436

437 Methods

438

439 **Human participants.** The Institutional Review Boards of Mass General Brigham/Massachusetts General Hospital,
440 Brown University, Providence VA Medical Center, and Stanford University granted permission for this study.
441 Intracortical neural signals were recorded from participant T11, a 37-year-old right-handed male with a C4 AIS-B
442 spinal cord injury (SCI) that occurred approximately 11 years prior to study enrollment, and T5, a 65-year-old right-
443 handed male, with a C4 AIS-C SCI that occurred approximately 9 years prior to study enrollment. Both
444 participants are enrolled in the BrainGate2 pilot clinical trial (NCT00912041), permitted under an Investigational
445 Device Exemption (IDE) by the US Food and Drug Administration (Investigational Device Exemption #Gog0003;
446 CAUTION: Investigational device. Limited by Federal law to investigational use). Informed consent was obtained
447 from all participants. All research sessions were performed at the participant's place of residence. All ethical
448 regulations relevant to human research participants were followed.

449

450 **Intracortical neural recordings and neural features.** Each participant had two 96-channel microelectrode arrays
451 (Blackrock Neurotech, Salt Lake City, UT) placed in the dominant (left) hand/arm knob area of the precentral
452 gyrus². T11's intracortical neural signals were recorded via a wireless broadband iBCI system⁷¹ while T5's recorded
453 was acquired via the cabled iBCI system. All research sessions were performed at the participant's place of

454 residence. The average signal across the array per electrode was subtracted with a common average reference
455 filter to reduce common mode noise. Neural features were extracted from the neural recording in 20 ms non-
456 overlapping bins. For real-time decoding and offline analysis, multi-unit threshold-crossing spike rates (RMS < -
457 3.5) per electrode were used for T₅, and two types of features: spike rates (RMS < -3.5) and power in the spike-
458 band (250 – 5000 Hz) per electrode were used for T₁₁. Across the 15 sessions, 34 of the 1840 trials were labeled as
459 outlier trials, which was defined as having more than a 5% drop of wireless neural data packets or large “neural”
460 responses greater than 8 standard deviations from the mean. No outlier trials were identified in T₅’s sessions.
461

462 **BCI behavioral task.** To assess decoder performance for cursor trajectories, T₁₁ performed a closed-loop 2D
463 point-and-click center-out-and-back task for each of 15 sessions on separate days that spanned 4 months (trial
464 days 658-800). For each trial, T₁₁ was prompted to attempt hand or finger movements to continuously move the
465 neural cursor from the center target to one of the eight pseudo-randomly selected peripheral targets and to then
466 attempt a hand gesture (right index finger down) to click on the target. T₁₁ was encouraged to maintain the same
467 set of motor imagery for all sessions presented in this study. Upon target selection, in the next trial T₁₁ was asked
468 to move the cursor back to the center target. A trial is successful when the cursor is inside the target and a click
469 action is decoded. Otherwise, a trial is considered failed after a 10-second timeout. Each session consists of two 5-
470 mins task blocks, except for trial day 751 with only one block. The cumulative task time of all sessions is 145 mins,
471 with a total of 1840 trials. Neural features, cursor position, target position, and decoder velocity outputs were
472 logged.
473

474 T₅ performed a closed-loop 2D random target selection task with a fixed-size target appearing in random
475 locations on the screen. T₅ attempted to move the cursor over the target and dwell on it for a consecutive 500ms
476 period before the 10-second timeout to complete the trial. Audio feedback was provided right after the end of a
477 trial to indicate trial success. A new random target is immediately presented with no delay. This task was repeated
478 for 6 sessions on separate days that spanned 28 days (trial days 2121-2149). Each session consists of two to four 4-
479 mins closed-loop blocks, which provide 84 mins of 1200 total trials across all sessions. Training blocks for
480 calibrating the decoder on trial day 2121 were not included in the test data.
481

482 **Angle error.** The instantaneous angle error is defined as the absolute angle difference between the inferred
483 intended directional vector (cursor position to target position) and the decoded velocity vector, $\hat{\theta}$ (best = 0°; max =
484 180°). In each 60-second interval to estimate the neural distribution for calculating the KLD, the median of the
485 angle error time bins during the same interval is computed. Median is chosen over mean because AE in our
486 datasets for both participants is not uniformly distributed between 0° and 180° (skew towards lower AE).
487

488 **Closed-loop neural decoding.** Decoders in this study were previously described in ^{27,51}. Briefly, for T₁₁, an LSTM
489 decoder was used to infer movement intentions from neural recordings. An LSTM is a variant of recurrent neural
490 network (RNN) with improved capability for long-term temporal dependencies ⁷². Previous studies described the
491 advantage of using a RNN for neural decoding over linear methods such as the Kalman filter ^{20,73-75}. The LSTM
492 decoder was trained and validated on closed-loop point-and-click cursor tasks from the 18 most recent sessions of
493 T₁₁, spanning 70 days from trial day 576 to 646. Of these sessions, only task blocks with a median angle error less
494 than 45° were included, which yielded a total of 331 minutes or 8441 trials of training data. Input neural features
495 were passed directly to the RNN layer whose outputs went to three densely connected activation functions,
496 decoding the x- and y-velocity and the distance to target. During online control, each neural feature was
497 adaptively z-scored using the mean and variance from a 3-minute rolling average window.
498

499 Clicks were decoded with a linear discriminant analysis (LDA) followed by a hidden Markov model. The LDA
500 calculates a subspace that maximally discriminates between a click and a movement state. Coefficients were
501 estimated with a regularization term of 0.001. Emission means and covariances used the empirical mean and
502 covariance from the training data. The selected z-scored neural features were smoothed with a 100ms boxcar
503 window before projected onto the LDA space. The estimated class probabilities were normalized using the
504 SoftMax function, then smoothed with a 400ms boxcar window. A click is returned when the click probability was
505 above a threshold of 0.98.
506

507 For T₅, non-overlapping 20ms-binned extracted features were fed through a linear regression model trained to
 508 predict the cursor-to-target distance. An initial decoder was trained based on T₅'s neural activity while he
 509 engaged in an open loop block on day 0 (trial day 2121). This decoder was then used to drive closed-loop control in
 510 a subsequent block. The final decoder parameters were then updated based on the first closed-loop block, and
 511 they were fixed for later closed-loop blocks and future sessions. The raw decoded velocity \hat{v}_t was exponentially
 512 smoothed with the running velocity average $\hat{\bar{v}}_t$ via $\hat{X}_t = \alpha\hat{X}_{t-1} + (1 - \alpha)\beta v_t$, where α is the smoothing factor and
 513 β is the gain parameter. Smoothing and gain were manually adjusted during the first session and fixed on
 514 subsequent days.

515
 516 To accommodate for session-to-session variability in recordings, we applied per channel z-scoring at every time
 517 bin for T₁₁ and a bias correction for T₅. For T₁₁, mean and variance were initialized from the previous block and
 518 adaptively update them using a 3-min rolling window. Neural features were decoded into cursor velocities by a
 519 real-time LSTM decoder. For T₅, a bias correction was applied to mitigate mean shifts in the decoded output by
 520 subtracting a running estimate of the decoder bias from the velocity outputs (with an adaption rate of 0.3)¹¹. Bias
 521 correction was first initialized from the previous blocks. The intercept term in the decoder is then updated to the
 522 negative resulting bias vector (obtained by pushing the mean firing rate vector through the decoder weights).

523
 524 **Derived neural features.** MINDFUL can be based on any collection of neural features. In this paper we experiment
 525 with three different types of features. The first is extracted neural features as described above (384 dimensions for
 526 T₁₁ and 192 dimensions for T₅). Individual neural features were z-scored per channel using a 3-min rolling window
 527 as implemented during online iBCI control for T₁₁. The same procedure is applied to T₅ despite a bias correction
 528 approach was applied during online control to offset means drifts. The second is based on principal components
 529 analysis (PCA) of the extracted neural features (after z-scoring). The recorded neural features were projected onto
 530 the PCA subspace defined by the top M principal components (PCs) of a reference dataset that we call the *PCA-*
 531 *reference data*; see below. The third is based on the output of the decoder (which can be viewed as a type of neural
 532 feature), \hat{v}_t . We also consider \hat{v}_{t-20} , the previous time bin (20ms earlier) of \hat{v}_t .

533
 534 **Kullback-Leibler divergence (KLD).** The MINDFUL score is based on comparing two datasets of neural features,
 535 defined in our case by the neural feature vectors from two collections of time bins. We use the derived neural
 536 features as described above. The first collection of time bins defines the reference *data* μ_1 and the second defines
 537 the *comparison data* μ_2 . Our choices for the reference and comparison time bins are described below. We first
 538 compute the sample mean (column) vectors μ_1 and μ_2 of the neural feature vectors in the reference data Σ_1 and
 539 comparison data Σ_2 , respectively. These mean vectors are the same dimension k as the derived neural features.
 540 We similarly compute the $k \times k$ sample covariance matrices Σ_1 and Σ_2 in the two datasets. The MINDFUL score
 541 that we use is

$$542 d_{KL}(P_1 || P_2) = \frac{1}{2} \left(\text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) - k + \ln \left(\frac{\det \Sigma_2}{\det \Sigma_1} \right) \right)$$

543
 544 where tr and det denote the trace and determinant of a matrix, respectively, and ln is the natural logarithm. This
 545 formula is the Kullback-Leibler divergence (KLD) between two k-dimensional multivariate Gaussian distributions
 546 with respective mean vectors μ_1 and μ_2 and respective covariance matrices Σ_1 and Σ_2 . Although it is motivated by
 547 a multivariate Gaussian model, its utility as a score for comparing two datasets does not rely on a Gaussian
 548 assumption. In developing MINDFUL, we experimented with other measures of statistical difference based only
 549 on means and covariances, such as the Jeffrey's (symmetric KL), Bhattacharyya, and Wasserstein distances
 550 between multivariate Gaussians, and found qualitatively similar results. We selected KLD as the example for this
 551 paper, because it consistently gave the best results in many different scenarios and it is widely known.

552
 553 **KLD grouped by angle error.** For the results in Fig. 1b, the reference and comparison time bins are defined by the
 554 angle error (AE) of the time bin, regardless of movement intention and session date. The reference data consist of
 555 all time bins for which the AE < 4°. We used these same time bins for the PCA-reference data. The comparison
 556 data consist of time bins for which the AE is in a particular 4° interval. We used 45 different comparison datasets
 557 defined by the AE intervals [0, 4°], [4°, 8°], ..., [172°, 176°], [176°, 180°] giving 45 different KLD scores, calculated as

described above. These scores are plotted versus AE (using the middle of each AE interval for the AE value) in Fig. 1b, and these 45 pairs define the reported correlations for Fig. 1b. The derived neural features used are the top M=5 PCs.

MINDFUL score to track model drift during closed-loop iBCI control. MINDFUL quantifies the neural distribution shifts over time relative to a historical reference distribution. We use KLD as described above and experiment with different choices of derived neural features and reference and comparison time bins. The reference time bins are restricted to the first session when the decoder was first deployed for T11, and the first two sessions for T5 (the first session is shorter than the others; see **supplementary Fig. 6** for KLD using only the first session for reference). We additionally restricted the reference time bins to those with AE < 4° in **Fig. 1c, 2, 5b**. We varied this AE threshold for inclusion in the reference data in **Fig. 5a**. In all of these figures the PCA-reference data is the same as the reference data. In **Fig. 1c, 2, 5a**, the comparison data consists of all time bins in a 60-second interval. In **Fig. 5b**, we varied the length of the comparison data interval. In all of these figures the comparison data interval is shifted in increments of 1 second to investigate how the MINDFUL score varies over time. In **Fig. 1**, the derived neural features are the top M=5 PCs. In **Fig. 2a**, the derived features are the 2-dimensional decoder output \hat{x} . In **Fig. 2b**, the derived features are 4-dimensional and consist of \hat{x} and the 2-dimensional decoder output from the previous time step \hat{X}_{lag} . In **Fig. 2c, 5**, the derived features are \hat{x} , \hat{X}_{lag} , and the top M=5 PCs, for a total of 9 dimensions. **Supplementary table 1** shows results with additional choices of the derived features (no z-score, M=10, or no PCA).

MINDFUL robustness to reference data across tasks. In **Fig 6**, we repeated the process of calculating MINDFUL as in **Fig 2c** except using different task data for the reference distribution. Additionally, no subsampling based on AE was applied for the reference as the ground truth performance is not available during personal use (no target was cued and therefore performance metrics are not readily available). Approximately 16 minutes of personal use and 10 minutes of random target tasks were included for the reference. The target distributions were still estimated from a 60-second sliding window during center-out-and-back tasks on subsequent sessions. Thus, median AEs calculated over time bins of these target distributions remain the same as **Fig 1c** and **2c** in this analysis.

Cosine tuning. We fit cosine tuning curves to estimate the tuning properties per feature per session. Cosine tuning has been used to describe the relationship between the neuronal firing rate to movement directions, and it forms a basis of using a linear decoder for neural decoding⁵⁷. In a cosine model $y = b_0 + b_1 \cos \theta + b_2 \sin \theta$, the firing rate of a neuron, is regressed on θ , the movement direction. b_0, b_1, b_2 are regression coefficients that can be estimated with least squares unbiased estimators. The model can also be expressed equivalently as $y = b_0 + \alpha \cos(\theta - \theta_0)$, where $\alpha = \sqrt{b_1^2 + b_2^2}$ representing the modulation depth (MD) of the cosine curve. and θ_0 representing the preferred direction (PD) where the largest firing rates are recorded. It is noted that since features in this study were z-scored, the bias term b_0 is close to zero, and is therefore omitted in the calculation of modulation depth. Tuning parameters were estimated from 20ms-binned features from the first second after the go-cue of non-outlier trials (offset by 160 ms reaction time) to capture neural activity associated with reach initiations. Feature tunings are considered significant if the F-test on the regression model has a p-value < 0.05. See **Supplementary Fig. 7** for the estimated tuning and the empirical firing rate of example features.

Changes in MD and PD. For each feature on each day, change in MD was calculated by $\Delta MD = MD - MD_{ref}$, and change in PD was calculated by $|\Delta PD| = \min(360^\circ - |PD - PD_{ref}|, |PD - PD_{ref}|)$, where ΔMD and ΔPD refer to the tuning of the first day for which the feature was significantly tuned. The CircStat toolbox was used⁷⁶. Only features that have significant directional tuning for more than half of all sessions were considered in the tuning map described below. (For all features including the non-significantly tuned, see **Supplementary Fig. 8**). Significance of tuning change was assessed by bootstrapping samples of PD (or MD) to obtain a distribution of ΔMD and ΔPD ^{33,35}. If the 95% confidence interval for the difference distribution does not contain 0, then we reject the null hypothesis at the 5% significance level³⁵. To visualize the patterns of tuning changes, the features were ordered by their tuning parameters using hierarchical clustering on Matlab. Euclidean distance was used to estimate the similarity of standardized ΔMD and ΔPD from all sessions between two

610 features, and ward linkage was applied to arrange the order of the clusters. The same ordering was used in the
611 heatmap of Δ MD and Δ PD for the same participant.

612
613 **Changes in tuning map.** We quantified changes in directional tuning on a population level by comparing tuning
614 maps over recorded sessions. A tuning map on each session is a $3 \times N$ matrix comprises the fitted tuning curve
615 parameters, b_0, b_1, b_2 , of N number of features with significant tuning on that day. Pairwise Pearson correlations
616 of maps was performed to assess the similarity of tuning across sessions. In a pair of daily tuning maps, only
617 features that were significant on both maps were considered to calculate the correlation. This pairwise correlation
618 was plotted in a heatmap which was interpolated to account for the irregular number of days apart between
619 sessions.

620
621 **Mean KLD between sessions.** To estimate the average neural distribution difference between pairs of sessions, a
622 mean KLD between distributions on two given sessions was calculated (see Fig. 3e, 3j). The 9-dimensional
623 derived neural features are the same as Fig. 2c, 5, namely, the top $M=5$ PCs, $\hat{\mathbf{X}}$, and $\hat{\mathbf{X}}_{lag}$. The PCA-reference data
624 is the same as Fig. 1c, 2, and 5, namely, time bins from the first session (T_{11}) or first two sessions (T_5) that have
625 $AE < 4^\circ$. The reference and comparison data are each 60-second intervals updated every 10 seconds. If there are
626 M such intervals in session \square and N in session \blacksquare , then the mean KLD between the sessions is the average KLD of
627 all $(\square \times \blacksquare)$ pairs of intervals from sessions \square and \blacksquare . Outlier trials were excluded from the intervals. For visualizing
628 the complete pairwise comparison matrix containing the KLD of all pairs of intervals from all sessions, see
629 Supplementary Fig. 9.

630
631 **Latent space dPCA projection.** We applied demixed principal component analysis (dPCA)⁵⁹ to population neural
632 activity from the initial decoder day for T_{11} and T_5 respectively. For subsequent trial days, we projected neural
633 data onto the top two task-relevant neural dimensions dPCA space of day α . PCs were computed from all features
634 of the first second after the go-cue of non-outlier trials (offset by 160 ms reaction time). Neural features were
635 smoothed using a gaussian kernel with standard deviation of 50 ms. T_5 random-target task trials were discretized
636 into eight movement directions in order to show comparable results to T_{11} 's center-out-and-back task with eight
637 peripheral targets. We further quantified the amount of task-related neural activity in each session by comparing
638 the variance accounted for (VAF) by the top two task-related neural components from the first session. VAF was
639 computed by

$$640 R^2 = \frac{||\bar{Y}||^2 - ||\bar{Y} - FD\bar{Y}||^2}{||\bar{Y}||^2}$$

641
642 Where \bar{Y} is trial-average neural features across conditions on a subsequent day, and encoder components \square and
643 decoder components \blacksquare estimated from the first session.

644

645 **Data Availability**

646 All data required to reproduce the findings in this study are publicly available on Dryad
647 (<https://doi.org/10.5061/dryad.n2z34tn5s>). The dataset contains intracortical neural signals recorded from both
648 participants along with detailed information about the BCI behavior tasks and performance metrics.

649

650 **Code Availability**

651 The code for reproducing the figures is made available on <https://github.com/ewinapun/MINDFUL>.

652

653 **Statistics and Reproducibility**

654 All statistical tests were reported using either Pearson's or Spearman's correlation coefficients for correlation, or
655 Wilcoxon rank sum for across days comparison; p-values are reported along with the name of the statistical test. A
656 number that follows the ' \pm ' sign is a standard deviation. Blinding and randomization were not relevant for this
657 two-participant study.

658
659

660 **Author contributions**
661
662 T.K.P., M.T.H., and L.R.H. initiated the study. T.K.P., M.K., and M.T.H. designed and investigated the
663 methodology. T.K.P. conducted the analysis. M.K. performed additional statistical analysis. T.K.P., T.H., G.H.W.,
664 A.K., F.K. were responsible for data collection. A.K., F.K. were responsible for session scheduling, logistics,
665 equipment setup/disconnection, and T.K.P., T.H., G.H.W., implemented code, designed the tasks and sessions,
666 and curated the datasets. T.K.P. drafted the manuscript, and C.E.V., M.T.H., and L.R.H. provided feedback and
667 revisions. All authors reviewed and edited the final manuscript. J.M.H., J.D.S., and L.R.H. were responsible for
668 funding acquisition and clinical research oversight. The study was co-supervised and guided by M.T.H., and L.R.H.
669

670 **Acknowledgement**
671 The authors would like to thank participants T11, T5, their families and caregivers, Beth Travers, Dave Rosler,
672 Maryam Masood, Ronnie Gross, Kristi Emerson, Sandrin Kosasih, Beverly Davis, and Kathy Tsou for their
673 contributions to this research. We acknowledge Krishna Shenoy for his inspiration and effort toward creating an
674 environment that spawned this work. Research was supported by the Office of Research and Development,
675 Rehabilitation R&D Service, Dept of Veterans Affairs (N2864C, A2295R, A2827R, A3803R), NIH NIMH
676 (T32MH115895), NIH-NINDS (UH2NS095548, U01NS098968), NIH-NIDCD (U01DC017844, R01DC014034), the
677 Croucher Foundation, Larry and Pamela Garlick, Wu Tsai Neurosciences Institute.
678

679 **Declaration of Interests**
680 The authors declare the following competing interests: The content is solely the responsibility of the authors and
681 does not necessarily represent the official views of NIH or the Department of Veterans Affairs or the United States
682 Government. The MGH Translational Research Center has a clinical research support agreement (CRSA) with
683 Axoft, Neuralink, Neurobionics, Precision Neuro, Synchron, and Reach Neuro, for which LRH provides
684 consultative input. LRH is a co-investigator on an NIH SBIR grant with Paradromics, and is a non-compensated
685 member of the Board of Directors of a nonprofit assistive communication device technology foundation (Speak
686 Your Mind Foundation). Mass General Brigham (MGB) is convening the Implantable Brain-Computer Interface
687 Collaborative Community (iBCI-CC); charitable gift agreements to MGB, including those received to date from
688 Paradromics, Synchron, Precision Neuro, Neuralink, and Blackrock Neurotech, support the iBCI-CC, for which
689 LRH provides effort. G.H.W. is a consultant for Artis Ventures. J.M.H. is a consultant for Neuralink and
690 Paradromics and is a shareholder in Maplight Therapeutics and Enspire DBS. He is also an inventor on intellectual
691 property licensed by Stanford University to Blackrock Neurotech and Neuralink.
692

- 693 **References**
- 694 1. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. Instant neural control of a
695 movement signal. *Nature* **416**, 141–142 (2002).
- 696 2. Hochberg, L. R. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*
697 **442**, 164–171 (2006).
- 698 3. Hochberg, L. R. *et al.* Reach and grasp by people with tetraplegia using a neurally controlled robotic arm.
699 *Nature* **485**, 372–375 (2012).
- 700 4. Schwartz, A. B., Cui, X. T., Weber, D. J. & Moran, D. W. Brain-controlled interfaces: movement restoration
701 with neural prosthetics. *Neuron* **52**, 205–220 (2006).
- 702 5. Simeral, J. D., Kim, S.-P., Black, M. J., Donoghue, J. P. & Hochberg, L. R. Neural control of cursor trajectory
703 and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. *J.*
704 *Neural Eng.* **8**, 025027 (2011).
- 705 6. Collinger, J. L. *et al.* High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* **381**,
706 557–564 (2013).
- 707 7. Gilja, V. *et al.* Clinical translation of a high-performance neural prosthesis. *Nat. Med.* **21**, 1142–1145 (2015).
- 708 8. Kim, S.-P. *et al.* Point-and-click cursor control with an intracortical neural interface system by humans with
709 tetraplegia. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**, 193–203 (2011).
- 710 9. Bacher, D. *et al.* Neural Point-and-Click Communication by a Person With Incomplete Locked-In Syndrome.
711 *Neurorehabil. Neural Repair* **29**, 462–471 (2015).
- 712 10. Brandman, D. M. *et al.* Rapid calibration of an intracortical brain-computer interface for people with
713 tetraplegia. *J. Neural Eng.* **15**, 026007 (2018).
- 714 11. Pandarinath, C. *et al.* High performance communication by people with paralysis using an intracortical brain-
715 computer interface. *Elife* **6**, (2017).
- 716 12. Jarosiewicz, B. *et al.* Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-
717 computer interface. *Sci. Transl. Med.* **7**, 313ra179 (2015).
- 718 13. Nuyujukian, P. *et al.* Cortical control of a tablet computer by people with paralysis. *PLoS One* **13**, e0204566
719 (2018).

- 720 14. Willett, F. R. *et al.* A high-performance speech neuroprosthesis. *Nature* 1–6 (2023).
- 721 15. Taylor, D. M., Tillery, S. I. H. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science*
722 **296**, 1829–1832 (2002).
- 723 16. Andersen, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. Multimodal representation of space in the posterior
724 parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* **20**, 303–330 (1997).
- 725 17. Homer, M. L., Nurmikko, A. V., Donoghue, J. P. & Hochberg, L. R. Sensors and decoding for intracortical
726 brain computer interfaces. *Annu. Rev. Biomed. Eng.* **15**, 383–405 (2013).
- 727 18. John E Downey, Nathaniel Schwed, Steven M Chase, Schwartz, A. B. & Collinger, J. L. Intracortical recording
728 stability in human brain–computer interface users. *Journal of neural Engineering* **15**, 046016 (2018).
- 729 19. Perge, J. A. *et al.* Intra-day signal instabilities affect decoding performance in an intracortical neural interface
730 system. *J. Neural Eng.* **10**, 036004 (2013).
- 731 20. Sussillo, D., Stavisky, S. D., Kao, J. C., Ryu, S. I. & Shenoy, K. V. Making brain-machine interfaces robust to
732 future neural variability. *Nat. Commun.* **7**, 13749 (2016).
- 733 21. Degenhart, A. D. *et al.* Stabilization of a brain–computer interface via the alignment of low-dimensional
734 spaces of neural activity. *Nature Biomedical Engineering* (2020).
- 735 22. Nuyujukian, P. *et al.* Performance sustaining intracortical neural prostheses. *J. Neural Eng.* **11**, 066003 (2014).
- 736 23. Davis, S. E. *et al.* A nonparametric updating method to correct clinical prediction model drift. *J. Am. Med.
737 Inform. Assoc.* **26**, 1448–1457 (2019).
- 738 24. Tsymbal, A. *The Problem of Concept Drift: Definitions and Related Work*. (Tech. Rep. Department of Computer
739 Science, Trinity College, Dublin., 2004).
- 740 25. Gama, J., Medas, P., Castillo, G. & Rodrigues, P. Learning with Drift Detection. in *Advances in Artificial
741 Intelligence – SBIA 2004* 286–295 (Springer Berlin Heidelberg, 2004).
- 742 26. Fan, C. *et al.* Plug-and-Play Stability for Intracortical Brain-Computer Interfaces: A One-Year Demonstration
743 of Seamless Brain-to-Text Communication. in *Conference on Neural Information Processing Systems* **37** (2023).
- 744 27. Wilson, G. H. *et al.* Long-term unsupervised recalibration of cursor BCIs. *bioRxiv* 2023.02.03.527022 (2023).
- 745 28. Kim, S.-P., Wood, F., Fellows, M., Donoghue, J. P. & Black, M. J. Statistical Analysis of the Non-stationarity of
746 Neural Population Codes. in *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and*

- 747 *Biomechatronics*, 2006. *BioRob 2006*. 811–816 (2006). doi:10.1109/BIOROB.2006.1639190.
- 748 29. Ahmadipour, P., Yang, Y., Chang, E. F. & Shanechi, M. M. Adaptive tracking of human ECoG network
749 dynamics. *J. Neural Eng.* **18**, 016011 (2021).
- 750 30. Driscoll, L. N., Duncker, L. & Harvey, C. D. Representational drift: Emerging theories for continual learning
751 and experimental future directions. *Curr. Opin. Neurobiol.* **76**, 102609 (2022).
- 752 31. Chestek, C. A. *et al.* Long-term stability of neural prosthetic control signals from silicon cortical arrays in
753 rhesus macaque motor cortex. *J. Neural Eng.* **8**, 045005 (2011).
- 754 32. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical
755 population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
- 756 33. Chestek, C. A. *et al.* Single-neuron stability during repeated reaching in macaque premotor cortex. *J.*
757 *Neurosci.* **27**, 10742–10750 (2007).
- 758 34. Carmena, J. M., Lebedev, M. A., Henriquez, C. S. & Nicolelis, M. A. L. Stable Ensemble Performance with
759 Single-Neuron Variability during Reaching Movements in Primates. *J. Neurosci.* **25**, 10712–10716 (2005).
- 760 35. Stevenson, I. H. *et al.* Statistical assessment of the stability of neural movement representations. *J.*
761 *Neurophysiol.* **106**, 764–774 (2011).
- 762 36. McGinley, M. J. *et al.* Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron* **87**,
763 1143–1161 (2015).
- 764 37. Vinck, M., Batista-Brito, R., Knoblich, U. & Cardin, J. A. Arousal and locomotion make distinct contributions
765 to cortical activity patterns and visual encoding. *Neuron* **86**, 740–754 (2015).
- 766 38. Hennig, J. A. *et al.* Learning is shaped by abrupt changes in neural engagement. *Nat. Neurosci.* **24**, 727–736
767 (2021).
- 768 39. Barrese, J. C. *et al.* Failure mode analysis of silicon-based intracortical microelectrode arrays in non-human
769 primates. *J. Neural Eng.* **10**, 066014 (2013).
- 770 40. Woepel, K. *et al.* Explant Analysis of Utah Electrode Arrays Implanted in Human Cortex for Brain-Computer-
771 Interfaces. *Front Bioeng Biotechnol* **9**, 759711 (2021).
- 772 41. Santhanam, G. *et al.* HermesB: a continuous neural recording system for freely behaving primates. *IEEE*
773 *Trans. Biomed. Eng.* **54**, 2037–2050 (2007).

- 774 42. Gilja, V. *et al.* A high-performance neural prosthesis enabled by control algorithm design. *Nat. Neurosci.* **15**,
775 1752–1757 (2012).
- 776 43. Bishop, W. *et al.* Self-recalibrating classifiers for intracortical brain-computer interfaces. *J. Neural Eng.* **11**,
777 026001 (2014).
- 778 44. Jarosiewicz, B. *et al.* Retrospectively supervised click decoder calibration for self-calibrating point-and-click
779 brain-computer interfaces. *J. Physiol. Paris* **110**, 382–391 (2016).
- 780 45. Brandman, D. M. *et al.* Robust Closed-Loop Control of a Cursor in a Person with Tetraplegia using Gaussian
781 Process Regression. *Neural Comput.* 1–23 (2018) doi:10.1162/neco_a_01129.
- 782 46. Farshchian, A. *et al.* Adversarial Domain Adaptation for Stable Brain-Machine Interfaces. in *International
783 Conference on Learning Representations* (2019).
- 784 47. Burkhardt, M. C., Brandman, D. M., Franco, B., Hochberg, L. R. & Harrison, M. T. The Discriminative Kalman
785 Filter for Bayesian Filtering with Nonlinear and Nongaussian Observation Models. *Neural Comput.* **32**, 969–
786 1017 (2020).
- 787 48. Ma, X., Rizzoglio, F., Perreault, E. J., Miller, L. E. & Kennedy, A. Using adversarial networks to extend brain
788 computer interface decoding accuracy over time. *bioRxiv* 2022.08.26.504777 (2022)
789 doi:10.1101/2022.08.26.504777.
- 790 49. Karpowicz, B. M. *et al.* Stabilizing brain-computer interfaces through alignment of latent dynamics. *bioRxiv*
791 2022.04.06.487388 (2022) doi:10.1101/2022.04.06.487388.
- 792 50. Jude, J., Perich, M. G., Miller, L. E. & Hennig, M. H. Capturing cross-session neural population variability
793 through self-supervised identification of consistent neuron ensembles. in *Proceedings of the 1st NeurIPS
794 Workshop on Symmetry and Geometry in Neural Representations* (eds. Sanborn, S., Shewmake, C., Azeglio, S.,
795 Di Bernardo, A. & Miolane, N.) vol. 197 234–257 (PMLR, 2023).
- 796 51. Hosman, T., Pun, T. K., Kapitonava, A., Simeral, J. D. & Hochberg, L. R. Months-long high-performance fixed
797 LSTM decoder for cursor control in human intracortical brain-computer interfaces. in *2023 11th International
798 IEEE/EMBS Conference on Neural Engineering (NER)* (2023).
- 799 52. Homer, M. L. *et al.* Adaptive Offset Correction for Intracortical Brain–Computer Interfaces. *IEEE Trans. Neural
800 Syst. Rehabil. Eng.* **22**, 239–248 (2014).

- 801 53. Kim, S.-P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P. & Black, M. J. Neural control of computer cursor
802 velocity by decoding motor cortical spiking activity in humans with tetraplegia. *J. Neural Eng.* **5**, 455–476
803 (2008).
- 804 54. Rokni, U., Richardson, A. G., Bizzi, E. & Seung, H. S. Motor learning with unstable neural representations.
805 *Neuron* **54**, 653–666 (2007).
- 806 55. Truccolo, W., Fries, G. M., Donoghue, J. P. & Hochberg, L. R. Primary motor cortex tuning to intended
807 movement kinematics in humans with tetraplegia. *J. Neurosci.* **28**, 1163–1178 (2008).
- 808 56. Ganguly, K. & Carmena, J. M. Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biol.* **7**,
809 e1000153 (2009).
- 810 57. Georgopoulos, A. P., Kalaska: Roberto, z. John & Massey, J. T. On the relations between the direction of two-
811 dimensional arm movements and cell discharge in primate motor cortex. *The Journal of Neuroscience* **2**, 1527–
812 1537 (1982).
- 813 58. Orsborn, A. L. *et al.* Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic
814 control. *Neuron* **82**, 1380–1393 (2014).
- 815 59. Kobak, D. *et al.* Demixed principal component analysis of neural population data. *Elife* **5**, (2016).
- 816 60. Downey, J. E. *et al.* Motor cortical activity changes during neuroprosthetic-controlled object interaction. *Sci.
817 Rep.* **7**, 16947 (2017).
- 818 61. Gallego, J. A. *et al.* Cortical population activity within a preserved neural manifold underlies multiple motor
819 behaviors. *Nat. Commun.* **9**, 4233 (2018).
- 820 62. Fraser, G. W., Chase, S. M., Whitford, A. & Schwartz, A. B. Control of a brain–computer interface without
821 spike sorting. *J. Neural Eng.* **6**, 055004 (2009).
- 822 63. Masse, N. Y. *et al.* Non-causal spike filtering improves decoding of movement intention for intracortical BCIs.
823 *J. Neurosci. Methods* **236**, 58–67 (2014).
- 824 64. Todorova, S., Sadtler, P., Batista, A., Chase, S. & Ventura, V. To sort or not to sort: the impact of spike-
825 sorting on neural decoding performance. *J. Neural Eng.* **11**, 056005 (2014).
- 826 65. Christie, B. P. *et al.* Comparison of spike sorting and thresholding of voltage waveforms for intracortical
827 brain-machine interface performance. *J. Neural Eng.* **12**, 016009 (2015).

- 828 66. Jarosiewicz, B. *et al.* Functional network reorganization during learning in a brain-computer interface
829 paradigm. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19486–19491 (2008).
- 830 67. Golub, M. D. *et al.* Learning by neural reassociation. *Nat. Neurosci.* **21**, 607–616 (2018).
- 831 68. Oby, E. R. *et al.* New neural activity patterns emerge with long-term learning. *Proc. Natl. Acad. Sci. U. S. A.*
832 **116**, 15210–15215 (2019).
- 833 69. Jarosiewicz, B. *et al.* Advantages of closed-loop calibration in intracortical brain-computer interfaces for
834 people with tetraplegia. *J. Neural Eng.* **10**, 046012 (2013).
- 835 70. Shanechi, M. M. *et al.* A real-time brain-machine interface combining motor target and trajectory intent
836 using an optimal feedback control design. *PLoS One* **8**, e59049 (2013).
- 837 71. Simeral, J. D. *et al.* Home Use of a Percutaneous Wireless Intracortical Brain-Computer Interface by
838 Individuals With Tetraplegia. *IEEE Trans. Biomed. Eng.* **68**, 2313–2325 (2021).
- 839 72. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
- 840 73. Tseng, P.-H., Urpi, N. A., Lebedev, M. & Nicolelis, M. Decoding Movements from Cortical Ensemble Activity
841 Using a Long Short-Term Memory Recurrent Network. *Neural Comput.* **31**, 1085–1113 (2019).
- 842 74. Premchand, B. *et al.* Decoding movement direction from cortical microelectrode recordings using an LSTM-
843 based neural network. in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine
844 and Biology Society (EMBC)* 3007–3010 (2020). doi:10.1109/EMBC44109.2020.9175593.
- 845 75. Hosman, T. *et al.* BCI decoder performance comparison of an LSTM recurrent neural network and a Kalman
846 filter in retrospective simulation. *9th International IEEE/EMBS Conference on Neural Engineering (NER)* 1066–
847 1071 (2019).
- 848 76. Berens, P. CircStat: A MATLAB Toolbox for Circular Statistics. *J. Stat. Softw.* **31**, 1–21 (2009).
- 849 77. Härdle, W. *Applied Nonparametric Regression*. (Cambridge University Press, 1990).
850 doi:10.1017/CCOL0521382483.

851

852 **Fig. 1 | MINDFUL score correlates with performance over time.**
 853 (a) Illustration of how the MINDFUL metric is calculated. Each dot symbolizes neural features at a given time bin in any given session, colored
 854 by whether the time bin is included for estimating the reference distribution (yellow), or for comparison (blue), or neither (gray). The
 855 difference in distributions is quantified by Kullback-Leibler divergence (KLD) between the reference distribution, p_0 and the comparison
 856 distribution, p_i .
 857 (b) Binned samples of neural features were grouped according to decoder performance (AE), including data from all sessions. A total of 45
 858 different distributions were generated, with AE increasing in 4° intervals from 0° to 180° . Bins with low AE ($< 4^\circ$) were chosen as the reference
 859 distribution, and compared against the other 44 for T11 (left panel) and T5 (right panel). The dotted line which represents the best linear
 860 regression fit, along with the Pearson correlation coefficients, r , is shown.
 861 (c) The reference distribution was estimated from neural features (NF) time bins where AE $< 4^\circ$, limited to day 0 where the decoder was first
 862 deployed for T11, and day 0 and 5 for T5. For subsequent sessions, neural distributions for comparison were constructed using an overlapping
 863 sliding window of 60 seconds at 1 second intervals. The KLD (right y-axis) is overlaid onto median AE (left y-axis in blue) across all recorded
 864 sessions for T11 (left panel) and T5 (right panel). Gray lines indicate the beginning of the session. Pearson and Spearman rank correlation, r
 865 and ρ respectively, quantify the relationship between the KLD and median AE. Insets present examples of cursor control of the task in the first
 866 and last session. For T11, cursor trajectories for all trials during a 5-min block are shown. Each color represents a peripheral target in a center-
 867 out-and-back task. For T5, cursor trajectories of the first 20 trials of a block are shown, along with the corresponding target presented at a
 868 random location on the screen in each trial.
 869

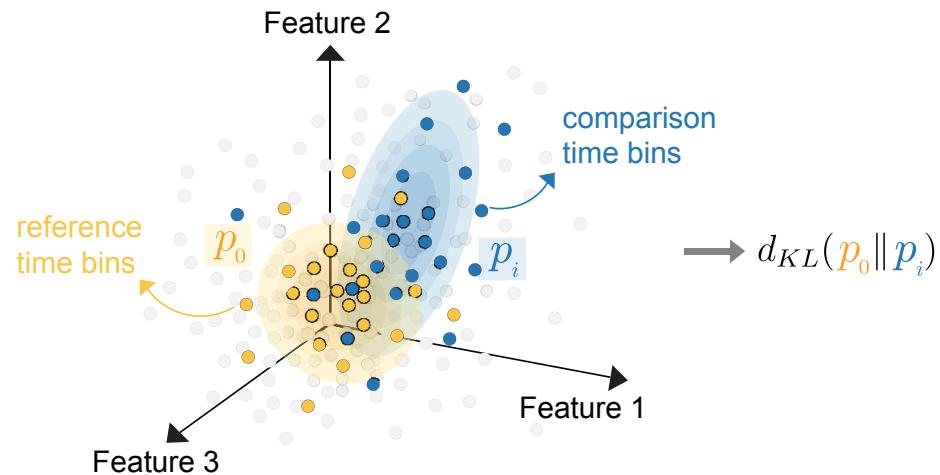
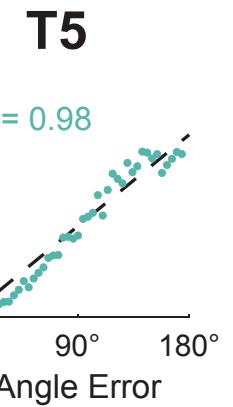
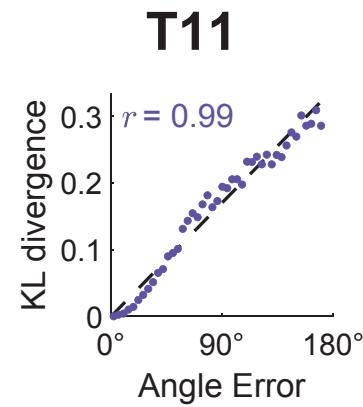
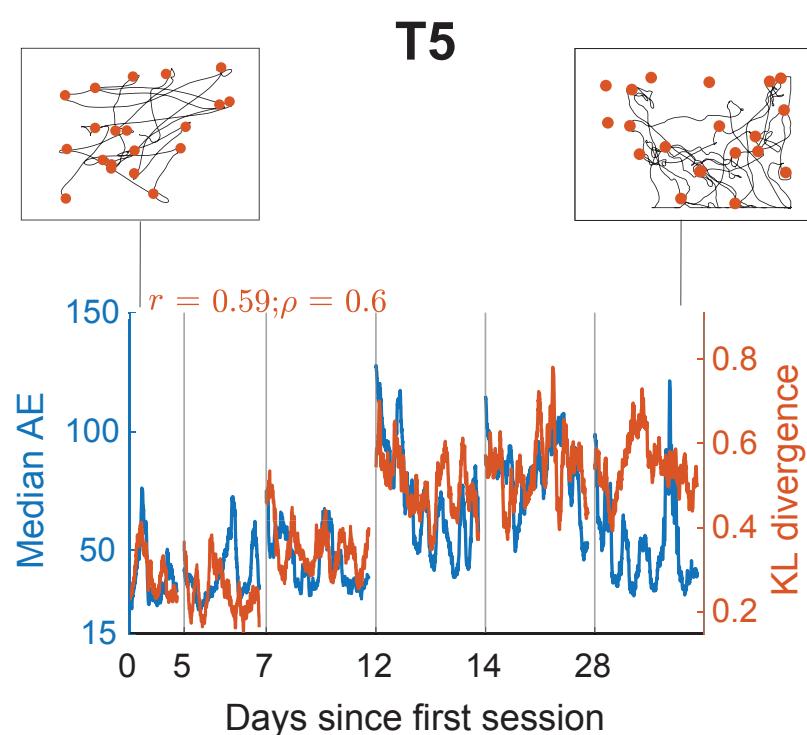
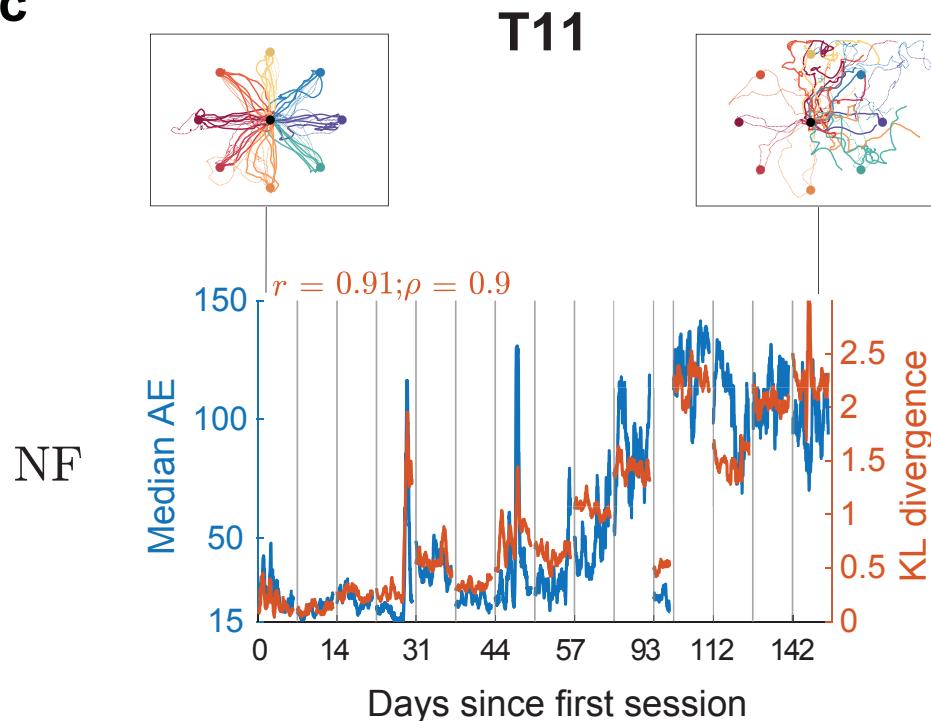
870 **Fig. 2 | Incorporating decoder outputs in the MINDFUL score maintains high correlation with performance**
 871 (a) The KLD (right y-axis) between distributions of decoded directional vectors, \hat{X} , with respect to the first session overlaid onto median AE
 872 (left y-axis in blue) across all recorded sessions for T11 (left panel) and T5 (right panel). Subsequent neural distributions and median AE were
 873 updated every 1 second over a 60-second sliding window. Pearson r , and Spearman rank correlation coefficients ρ , between KLD and median
 874 AE are shown. (b) The KLD between distributions of \hat{X} and \hat{X}_{lag} overlaid onto median AE. (c) The KLD between distributions of the
 875 combination of derived neural features (as shown in Fig. 1c), decoded directional vectors, \hat{X} , and \hat{X}_{lag} , overlaid onto median AE.
 876

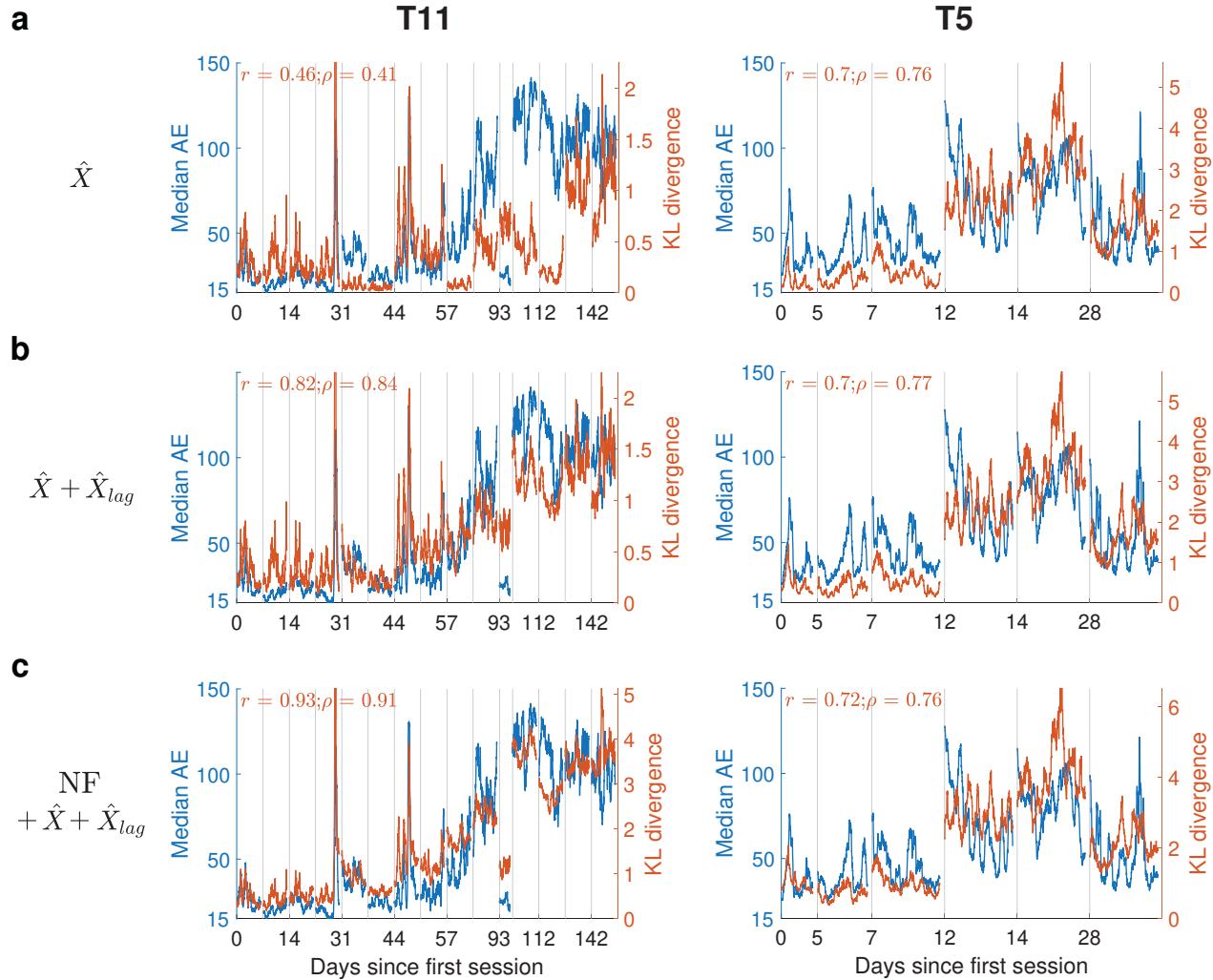
877 **Fig. 3 | Changes in feature tunings across session correlates with KLD**
 878 (a) Changes in preferred directions (PD) and (b) modulation depth (MD) of significantly tuned features used in the decoder (relative to the
 879 tuning of the first day for which the feature was significantly tuned) for T11. Features were ordered by hierarchical clustering to visualize
 880 groups of features with similar tuning behavior changes (see Methods). Gray color indicates features that were not significantly tuned in that
 881 session. Triangle markers correspond to the features presented in (c). (c) Fitted cosine tuning curves for sample units across days for T11
 882 illustrating changes in channel dropout and MD, respectively (color curves); Triangle markers denote PDs for sessions with significant tuning.
 883 (d) Changes in cosine tuning PD and (e) MD for significantly tuned features used in the decoder for T5. (f) Fitted cosine tuning curves for
 884 sample units across days of T5 illustrating changes in MD and PD, respectively. (g) T11 Tuning similarity across days represented by
 885 interpolated Pearson correlations between pairs of tuning maps (see Methods). (h) T5 tuning similarity across session days. (i) T11 mean KLD
 886 of neural distributions between sessions negatively correlates with the tuning similarity (Pearson $r = -0.812$, $p < 10^{-30}$, see Methods). Each
 887 dot corresponds to a pair of sessions with the color indicating the number of days apart. (j) T5 mean KLD of neural distributions between days
 888 negatively correlates with the tuning similarity (Pearson $r = -0.776$, $p < 10^{-4}$).
 889

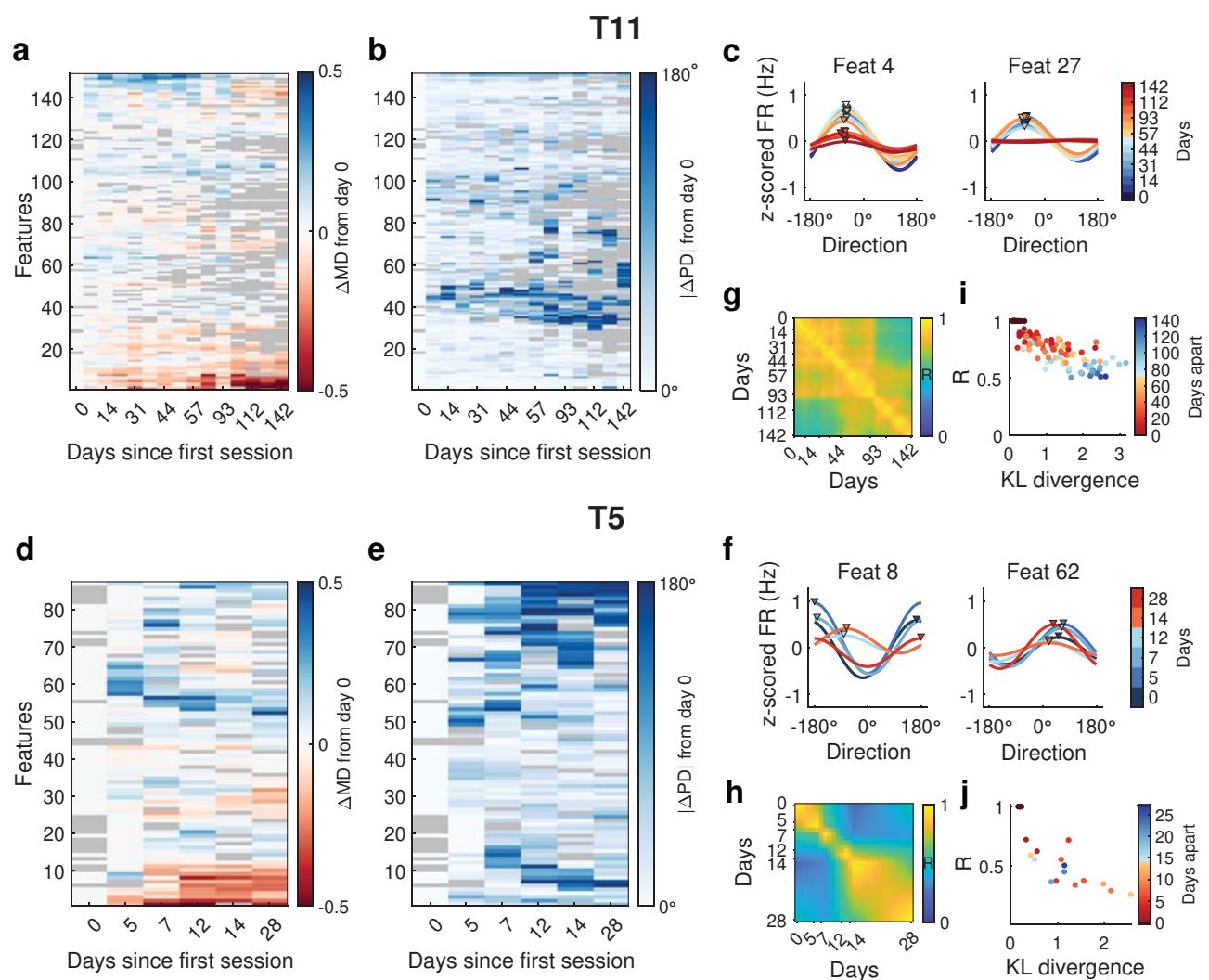
890 **Fig. 4 | Instability reflected in neural latent space**
 891 (a) Projection of neural features of subsequent sessions onto the top two task-dependent PCs latent space of neural features on decoder day 0
 892 using dPCA. Bold solid lines are trial averages per goal directions. Different colors correspond to the goal direction. (b) Projection of neural
 893 features for T5. For comparison simplicity, the random-target task was visualized and colored by discretizing goal directions of each trial into
 894 eight even movement directions as in a center-out-and-back task.
 895

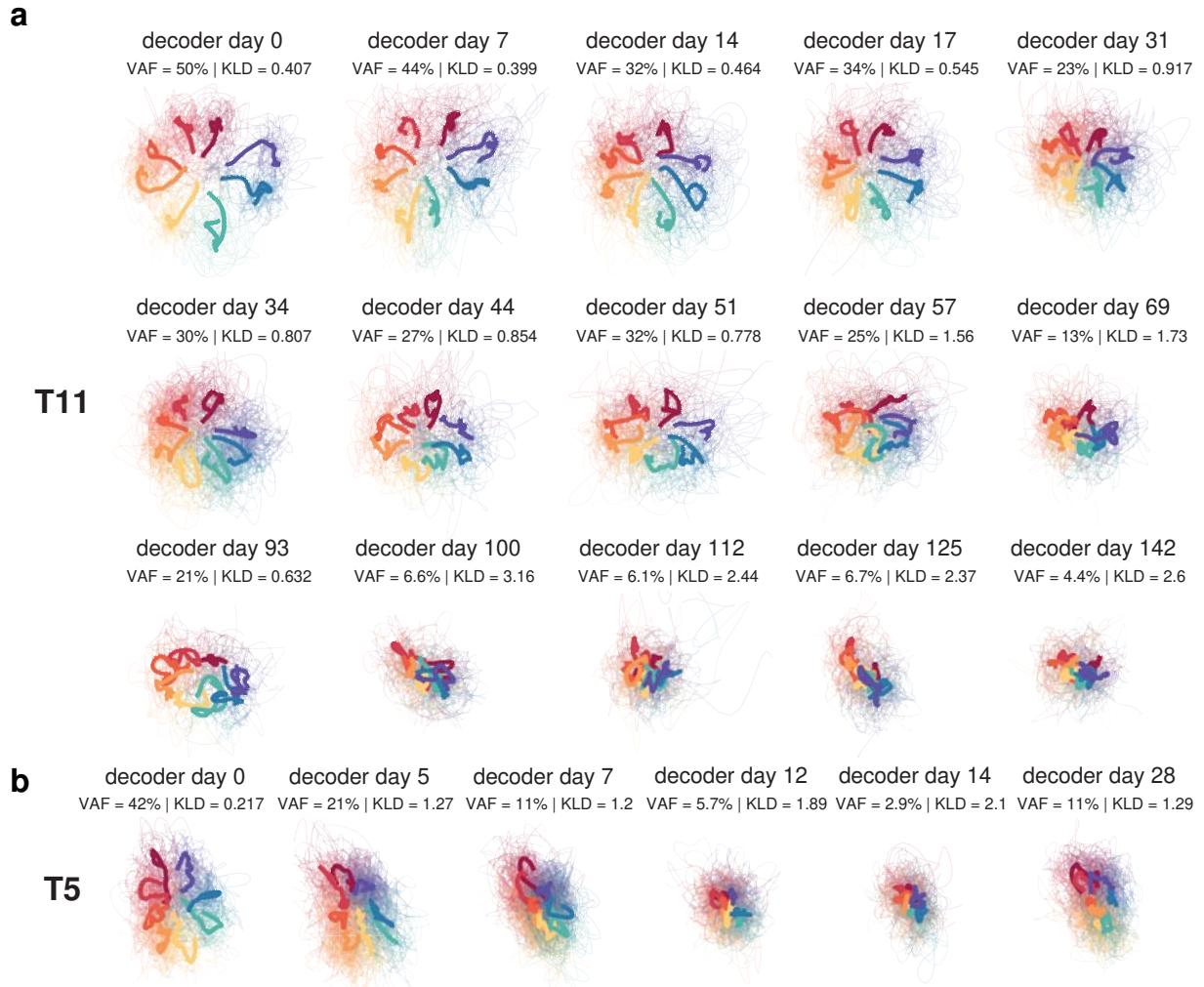
896 **Fig. 5 | Sub-selecting instances of low AE for reference and using longer window length improve MINDFUL correlation to AE** (a)
 897 Spearman correlation coefficients of KLD of NF, \hat{X} , and \hat{X}_{lag} , to AE when sub-selecting time steps with different quantiles of AE as the
 898 reference. Each quantile approximately contains an even number of observations. (b) Spearman correlation coefficients of KLD to AE when
 899 using different window lengths to estimate distributions.
 900

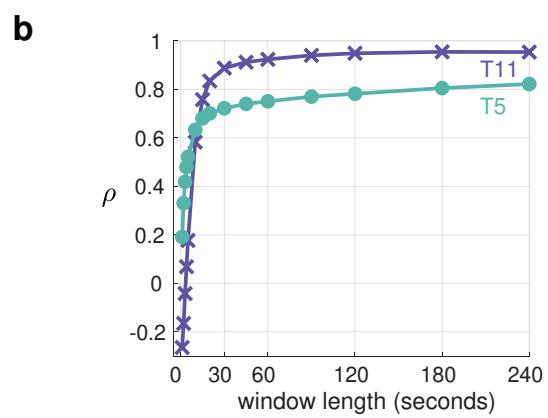
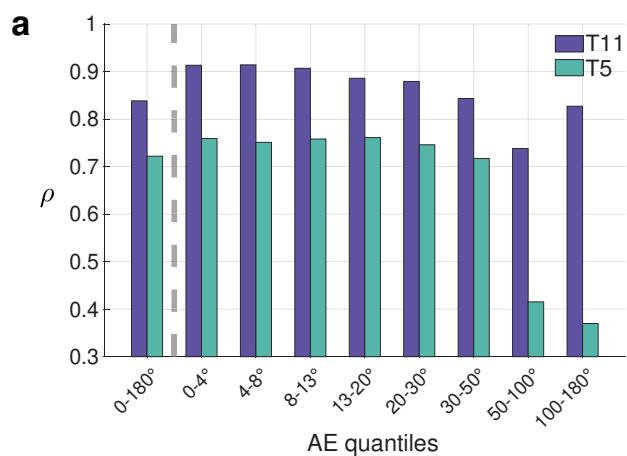
901 **Figure 6 | MINDFUL reference can be applied across multiple cursor control tasks.** Instead of using features from day 0 where participant T11
 902 was performing the same tasks as subsequent days (center-out-and-back), different 2D cursor tasks were selected as reference. In these tasks,
 903 T11 used the same LSTM fixed decoder. The same PCA procedure as described in the methods was applied. No subsampling on angle error was
 904 performed on the reference as it might not be explicitly available during personal use. Features used include all neural features and decoded
 905 kinematics plus its lagged version.
 906 (a) Reference was set to center-out task on day 0, same as in fig. 2c except AE was not sub-selected to be less than 4° .
 907 (b) Reference was set to a block of T11 performing a random target task (5 mins) on day 7, using the same fixed decoder. Task setup is very
 908 similar to T5's data except with varying target sizes. T11 did not perform this task on day 0.
 909 (c) Reference was set to when T11 was using the iBCI for personal desktop use, such as browsing the internet, on day 0. Around 10 mins of active
 910 cursor control period was included.
 911 (d) Reference was set to a combination of the above-mentioned data (concatenating random target task, personal use, and center-out-and-
 912 back task).

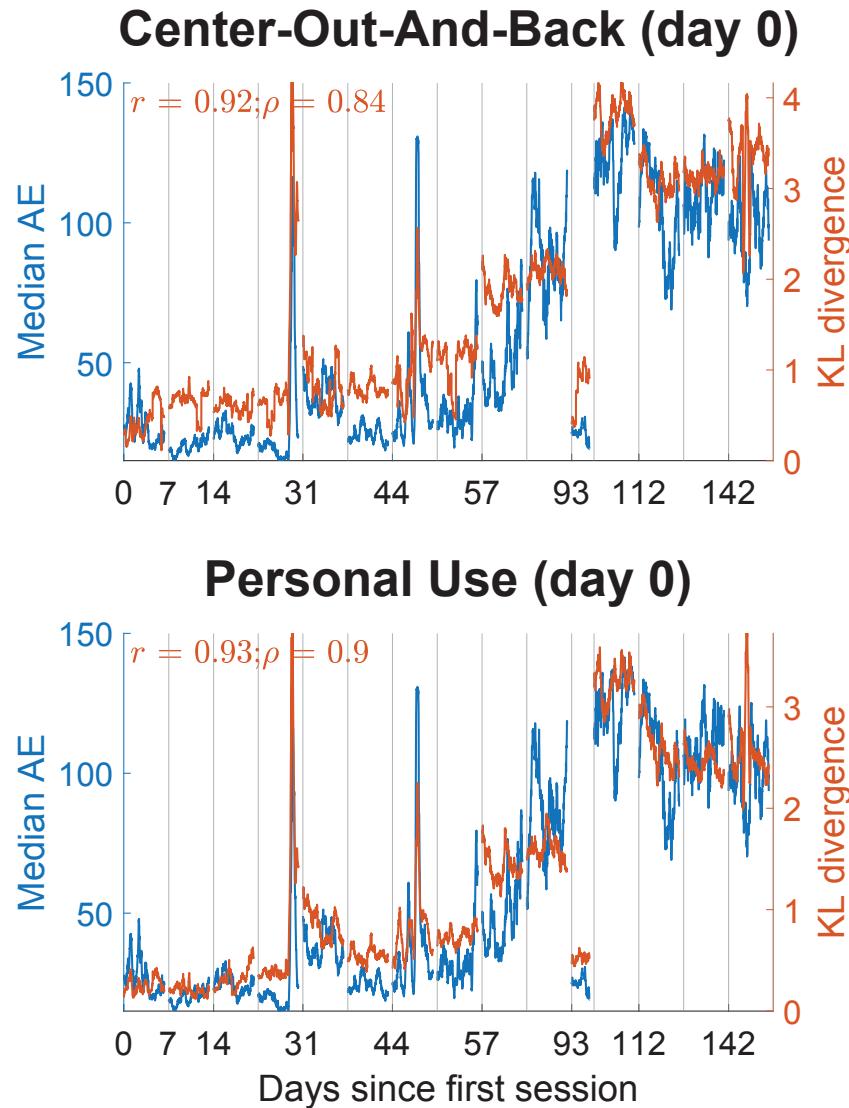
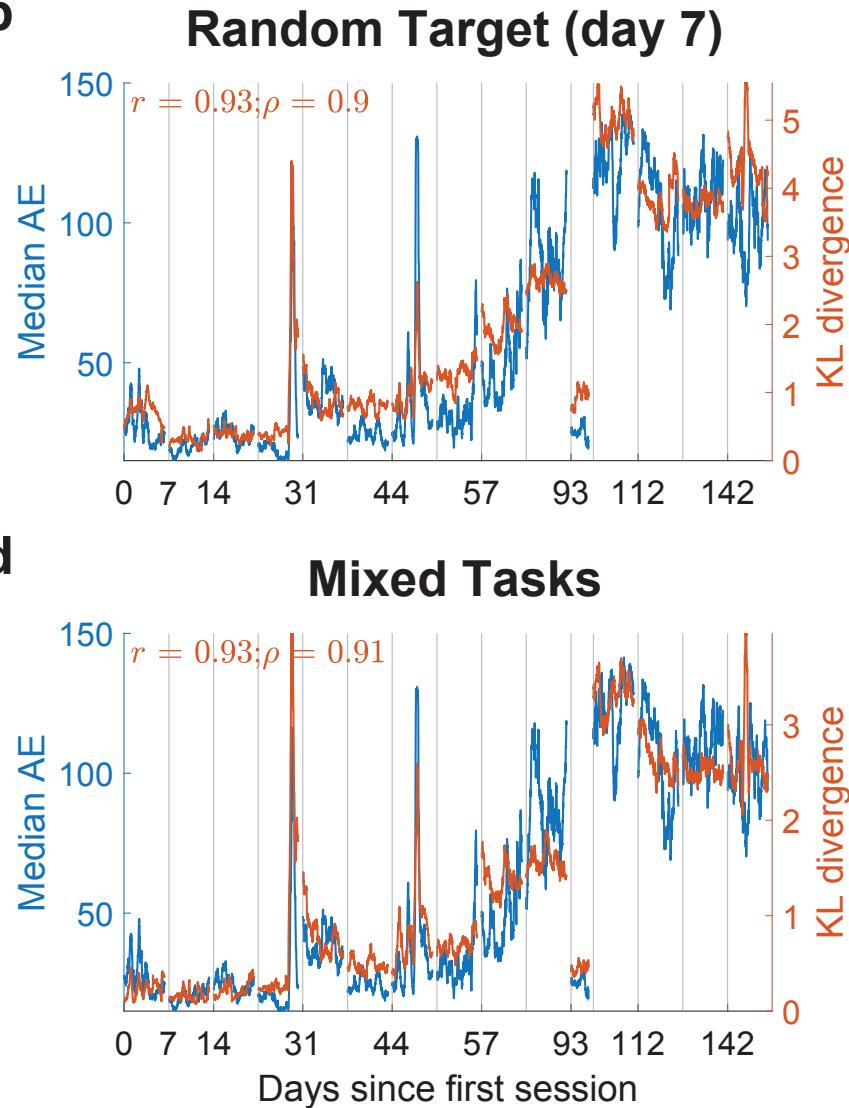
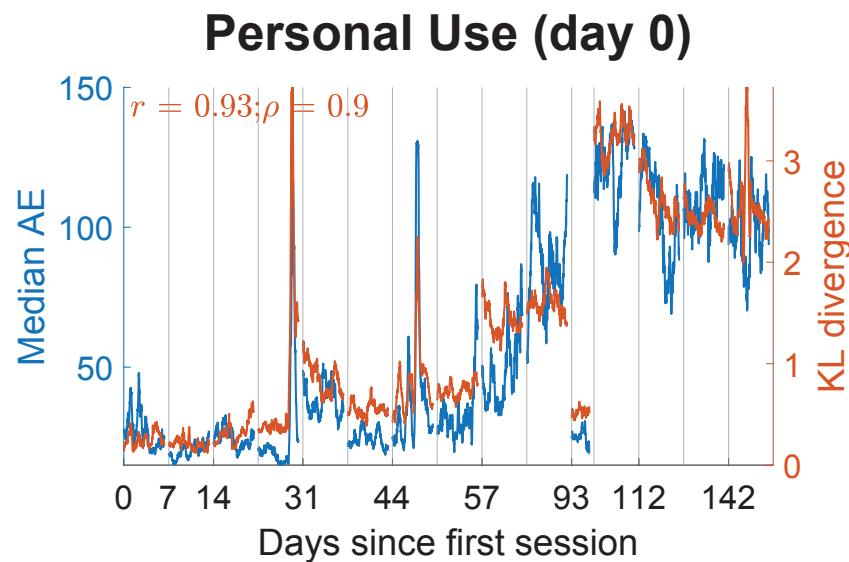
a**b****c**









a**b****c****d**