

Recurrent Processes Emulate a Cascade of Hierarchical Decisions: Evidence from Spatio-Temporal Decoding of Human Brain Activity.

Laura Gwilliams^{1,2} Jean-Remi King^{1,3,4,5}

¹Department of Psychology, New York University, ²NYU Abu Dhabi Institute, ³Frankfurt Institute for Advanced Studies, ⁴Ecole Normale Supérieure, ⁵CNRS

Abstract

Mounting evidence suggests that perception depends on a largely-feedforward brain network. However, the discrepancy between (i) the latency of the corresponding feedforward responses (150-200 ms) and (ii) the time it takes human subjects to recognize brief images (often >500 ms) suggests that recurrent neuronal activity is critical to visual processing. Here, we use magneto-encephalography to localize, track and decode the feedforward and recurrent responses elicited by brief presentations of variably-ambiguous letters and digits. We first confirm that these stimuli trigger, within the first 200 ms, a feedforward response in the ventral and dorsal cortical pathways. The subsequent activity is distributed across temporal, parietal and prefrontal cortices and leads to a slow and incremental cascade of representations culminating in action-specific motor signals. We introduce an analytical framework to show that these brain responses are best accounted for by a hierarchy of recurrent neural assemblies. An accumulation of computational delays across specific processing stages explains subjects' reaction times. Finally, the slow convergence of neural representations towards perceptual categories is quickly followed by all-or-none motor decision signals. Together, these results show how recurrent processes generate, over extended time periods, a cascade of hierarchical decisions that ultimately predicts subjects' perceptual reports.

Keywords: Perceptual decision making, Magneto-encephalography, Recurrent processes

1. Introduction

To process the rich sensory flow emanating from the retina, the brain recruits a hierarchical network originating in the primary visual areas and culminating in the infero-temporal, dorso-parietal and prefrontal cortices [1, 2, 3, 4].

In theory, the feedforward recruitment of this neural hierarchy could suffice to explain our ability to recognize visual objects. For example, recent studies demonstrate that artificial feedforward neural networks trained to categorize objects generate similar activation patterns to those elicited in the infero-temporal cortices [5, 6].

However, feedforward architectures have a fixed number of processing stages, and are thus unable to explain a number of neural and perceptual phenomena. For example, the time it takes subjects to recognize objects considerably varies from one trial to the next [7]. In addition, the neural responses to visual stimuli generally exceed the 200 ms feedforward recruitment of the

visual hierarchy [8, 9].

A large body of research shows that recurrent processing accounts for such behavioral and neural dynamics [9, 10, 11, 12, 13, 14]. In this view, recurrent processing would mainly consist in accumulating sensory evidence until a decision to act is triggered [13].

However, the precise neuronal and computational organization of recurrent processing remains unclear at the system level. In particular, how distinct recurrent assemblies implement series of hierarchical decisions remains a major unknown.

To address this issue, we use magneto-encephalography (MEG) and structural magnetic-resonance imaging (MRI) to localize, track and decode, from whole-brain activity, the feedforward (0-200ms) and recurrent processes (>200 ms) elicited by variably ambiguous characters briefly flashed on a computer screen. We show that the late and sustained neural activity distributed along the visual pathways generates, over extended time periods, a cascade of categorical

41 decisions that ultimately predicts subjects' perceptual
42 reports.

43 2. Results

44 2.1. Subjective reports of stimulus identity are categor- 45 ical

46 To investigate the brain and computational bases of
47 perceptual recognition, we used visual characters as de-
48 scribed in [15]. These stimuli can be parametrically
49 morphed between specific letters and digits by varying
50 the contrast of their individual edges, hereafter referred
51 to as pixels (Fig.1A-B).

52 To check that these stimuli create categorical per-
53 cepts, we asked eight human subjects to provide contin-
54 uous subjective reports by clicking on a disk after each
55 stimulus presentation (Experiment 1. Fig.1A). The radi-
56 us and the angle of the response on this disk indicated
57 the subjective visibility and the subjective identity of
58 the stimulus respectively. We then compared (i) the re-
59 ported angle with (ii) the stimulus evidence (i.e. the ex-
60 pected angle given the pixels) for each morph separately
61 (e.g. 5-6, 6-8, etc). Subjective reports were categori-
62 cal: cross-validated sigmoidal models better predicted
63 subjects' responses ($r=0.49\pm 0.05$, $p=0.002$) than lin-
64 ear models ($r=0.46\pm 0.043$, $p=0.002$, sigmoid>linear:
65 $p=0.017$ Fig.1B-C).

66 We adapted this experimental paradigm for an MEG
67 experiment by modifying three main aspects (Experi-
68 ment 2). First, we used stimuli that could be morphed
69 between letters and digits, to trigger macroscopically
70 distinguishable brain responses in the visual word form
71 area (VWFA) and number form area (NFA) [16, 17].
72 Second, we added two task-irrelevant flankers next to
73 the target stimulus (Fig.1D) to increase our chances of
74 eliciting recurrent processes via crowding [18]. Third, a
75 new set of seventeen subjects reported subjective iden-
76 tity via a two-alternative forced-choice button press.
77 The identity-response mapping was orthogonal to the
78 letter/digit category and changed on every block of 48
79 trials. There were 1920 trials total, 320 of which were
80 presented passively and did not require a response.

81 Perceptual reports followed a similar sigmoidal pat-
82 tern to Experiment 1: performance was worse for
83 more ambiguous trials (65%) as compared to unam-
84 biguous trials (92%, $p<0.001$). In addition, reaction
85 time slightly, and consistently, increased with difficulty.
86 For example, highly ambiguous stimuli were identified
87 within 690 ms, whereas nonambiguous stimuli were
88 identified within 624 ms ($z=-21.68$, $p<0.001$ (Fig.1E-
89 F).

90 2.2. Neural representations are functionally organized 91 over time and space

92 Here, we aimed to decompose the sequence of de-
93 cisions that allow subjects to transform raw visual in-
94 put into perceptual reports. To this aim, we localized
95 the MEG signals onto subjects' structural MRI with dy-
96 namic statistical parametric mapping (dSPM, [19]), and
97 morphed these source estimates onto a common brain
98 coordinate [20, 21]. The results confirmed that the stim-
99 uli elicited, on average, a sharp response in the primary
100 visual areas around 70 ms, followed by a fast feedfor-
101 ward response along the ventral and dorsal visual path-
102 ways within the first 150-200 ms. After 200 ms, the ac-
103 tivity appeared sustained and widely distributed across
104 the associative cortices up until 500-600 ms after stim-
105 ulus onset (Fig.1G and Supplementary Video 1).

106 To separate the processing stages underlying these
107 neural responses, we applied i) mass-univariate; ii) tem-
108 poral decoding and iii) spatial decoding analyses based
109 on the five orthogonal features varying in our study: (1)
110 the position of the stimulus, (2) its identity, (3) its per-
111 ceived category, (4) its difficulty and (5) its correspond-
112 ing button press.

113 First, we aimed to identify when and where low-level
114 visual features would be represented in brain activity.
115 To do so, we estimated, at each time sample separately,
116 the ability of an l2-regularized logistic regression to pre-
117 dict, from all MEG sensors, the position of the stimu-
118 lus on the computer screen (left versus right). Stimulus
119 position was decodable between 41 and 1,500 ms and
120 peaked at 120 ms ($AUC=0.94$; $SEM=0.007$; $p<0.001$
121 as estimated with second-level non-parametric tempo-
122 ral cluster test across subjects, (Fig.2C). These signals
123 peaked in the early visual cortex (mean MNI ($x=27.59$;
124 $y=-74.15$; $z=-1.07$)), and propagated along the ventral
125 and dorsal streams during the first 200 ms (Fig.2A, sup-
126 plementary video). To summarize where stimulus po-
127 sition was represented in the brain, we implemented
128 'spatial decoders': l2-regularized logistic regressions fit
129 across all time samples (0 - 1,500 ms) for each esti-
130 mated brain source separately. Spatial decoding peaked
131 in early visual areas and was significant across a large
132 variety of visual and associative cortices as estimated
133 with a second-level non-parametric spatial cluster test
134 across subjects (Fig.2B), confirming the retinotopic or-
135 ganization of the visual hierarchy [22, 23].

136 Second, we aimed to isolate more abstract representa-
137 tions related to stimulus identity. Stimulus identity can
138 be analyzed either from an objective referential (what
139 stimulus is objectively presented?) or from a subjec-
140 tive referential (i.e. what stimulus did subjects report

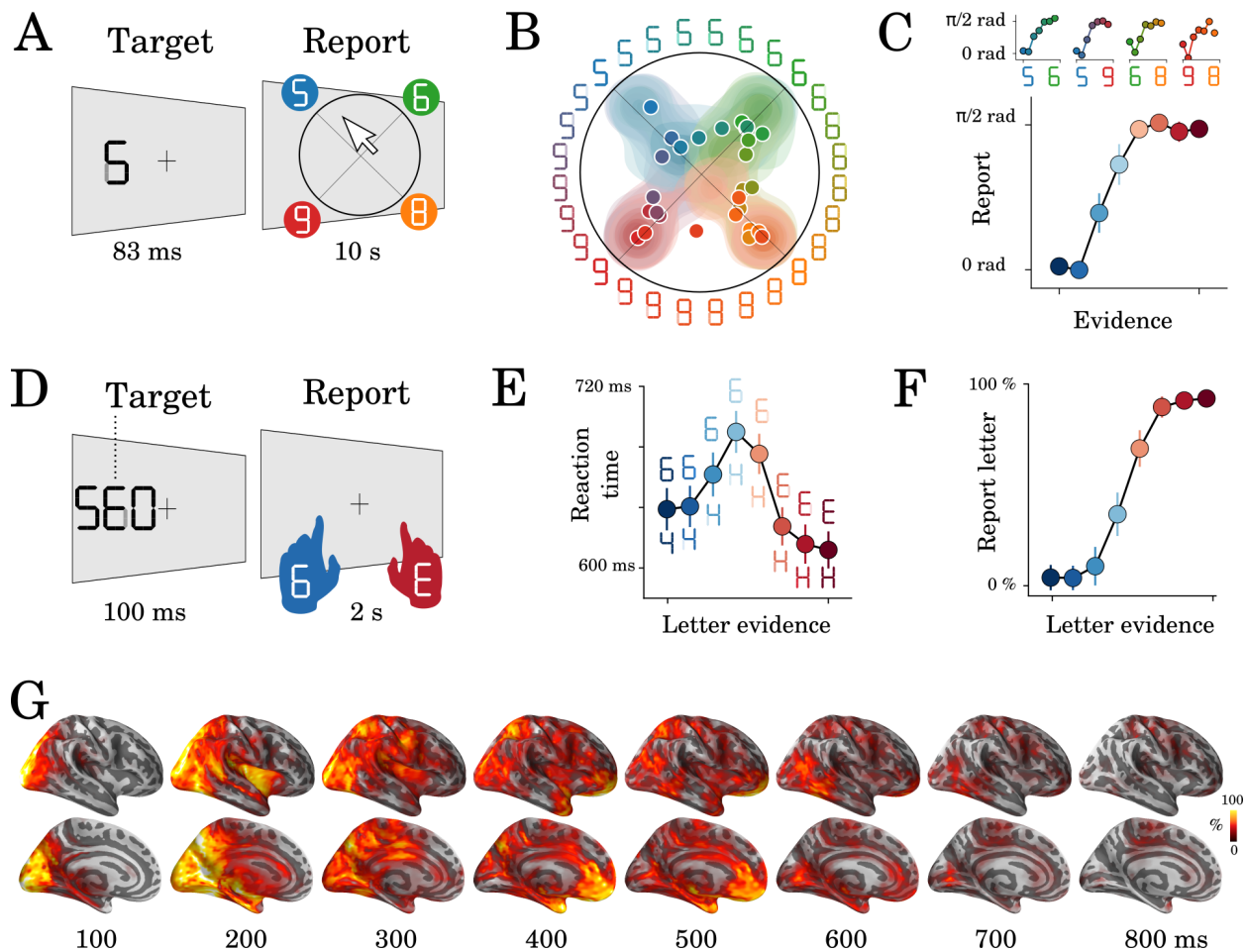


Figure 1: Experimental protocol and behavioral results. Experiment 1: 8 human subjects provided perceptual judgments on variably ambiguous digits briefly flashed at the center of a computer screen (A). Reports were made by clicking on a disk, where (i) the radius and (ii) the angle on the disk indicate (i) subjective visibility and (ii) subjective identity respectively. (B) Distribution (areas) and mean response (dots) location for each color-coded stimulus. (C) Top plots show the same data as B, broken down for each morph set. The x-axis indicates the expected angle given the stimulus pixels (color-coded), hereafter referred to as evidence. The y-axis indicates the angle of the mean response relative to stimulus evidence. The bottom plot shows the same data, grouped across morphs. (D) Experiment 2: 17 subjects categorized a briefly flashed and parametrically manipulated-morph using a two-alternative forced-choice. Stimulus-response mapping changed on every block. (E) Mean reaction times as a function of categorical evidence (the extent to which the stimulus objectively corresponds to a letter). (F) Mean probability of reporting a letter as a function of categorical evidence. (G) Evoked activity estimated with dSPM and estimated across all trials and all subjects. These data are also displayed in Supplementary Video 1. Error-bars indicate the standard-error-of-the-mean (SEM) across subjects.

141 having seen?). We first focus on decoding features of
 142 the stimulus that are not ambiguous, such that subjective
 143 and objective representations are confounded. To
 144 this aim, we grouped stimuli along common continua
 145 (e.g. The eight stimuli along the 4-H continuum belong
 146 to the same morph and are here considered to share a
 147 common identity) and fit logistic regression classifiers
 148 across morphs (i.e. E-6 versus 4-H). The correspond-
 149 ing stimulus identity was decodable between 120 and
 150 845 ms and peaked at 225 ms (AUC=0.59; SEM=0.01;
 151 $p<0.001$). These effects peaked more anteriorly than

152 those of stimulus position (mean MNI: $x=27.75$; $y=-$
 153 62.75 ; $z=-1.55$; $p<0.001$).

154 Third, we aimed to isolate the neural signatures of
 155 subjective perceptual categorization and thus focus on
 156 decoding ambiguous pixels. To this aim, we grouped
 157 stimuli based on whether the subject reported a digit
 158 or to a letter category. Temporal decoders weakly but
 159 significantly classified perceptual category from 150
 160 to 940 ms after stimulus onset and peaked at 370 ms
 161 (AUC=55%; SEM=0.01; $p<0.001$, Fig.2C). The corre-
 162 sponding sources also peaked in the inferotemporal cor-

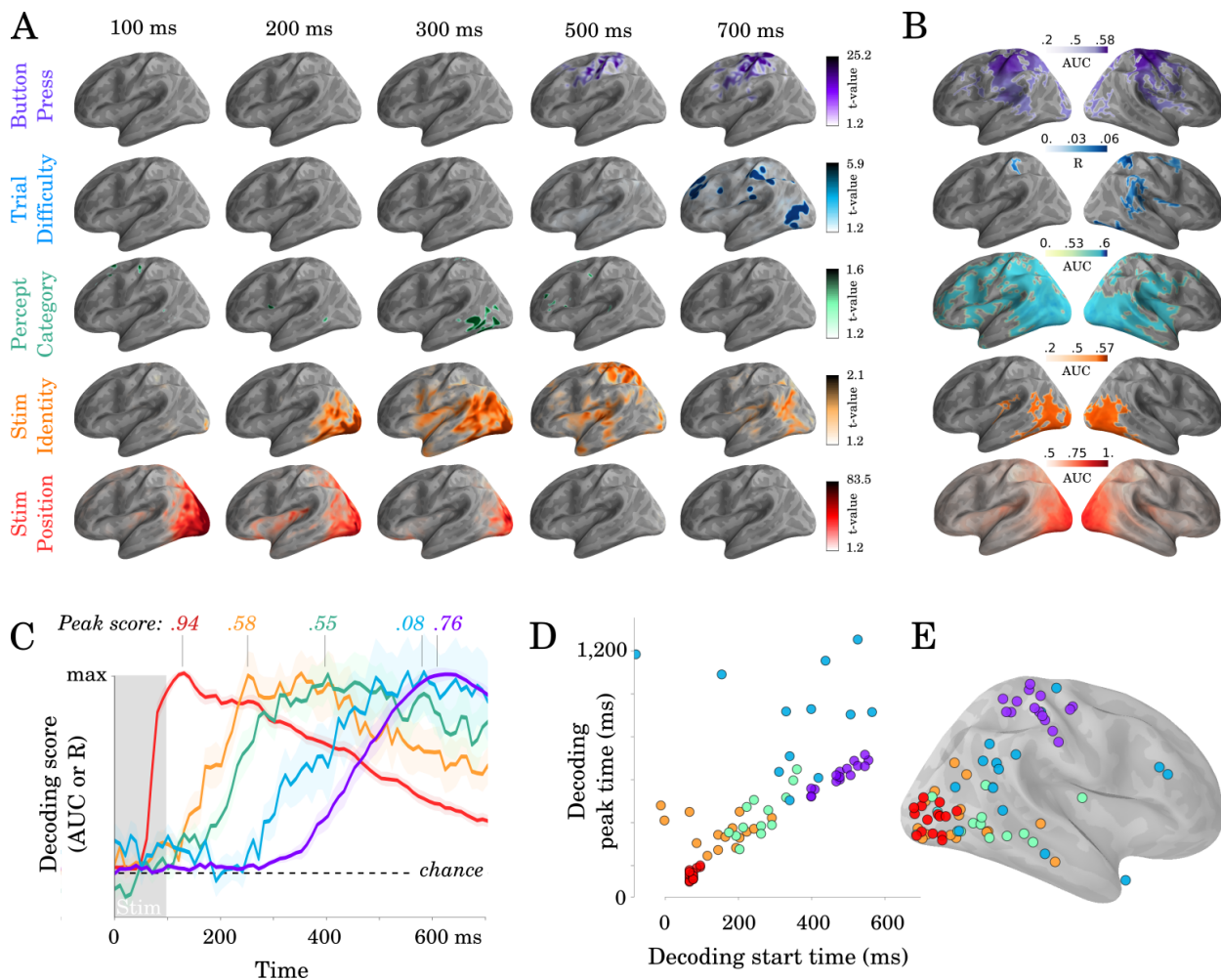


Figure 2: **Spatio-temporal hierarchy.** (A) Mass-univariate statistics. Each row plots the average-across-subjects beta coefficients obtained from regression between single-trial evoked activity and each of the five features orthogonally varying in this study. These results are displayed in Supplementary Video 2. Colors are thresholded based on t-values that exceed an uncorrected $p < .1$. (B) Spatial-decoders, consisting of linear models fit across all time sample for each source separately, summarize where each feature can be decoded. Lines indicate significant clusters of decoding scores across subjects. (C) Temporal-decoders, consisting of linear models fit across all MEG channels, for each time sample separately, summarize when each feature can be decoded. To highlight the sequential generation of each representation, decoding scores are normalized by their respective peaks. Additional statistics are available in Supplementary Figure 1. (D) The peak and the start of temporal decoding plotted for each subject (dot) and for each feature (color). (E) The peak spatial decoding plotted for each subject (dot) and for each feature (color).

163 tex but more anteriorly than stimulus identity ($x=30.89$;
164 $y=-35.64$; $z=21.41$; $p<0.01$). These mass-univariate effects
165 did not survive correction for multiple comparisons (e.g. 210-320 ms: $\bar{t}=1.79$, $p=0.21$). Nonetheless,
166 spatial decoders, which mitigate the trade-off between
167 temporal specificity and the necessity to correct statistical
168 estimates for multiple comparisons, showed that perceptual
169 category was reliably decoded from a large set of brain areas
170 ($\bar{t}=4.82$; $p<0.001$; 594 significant vertices) (Fig.2G).

173 Importantly, when training the classifier on all active
174 trials to distinguish letters (E/H) and digits (4/6),

175 we could significantly (max AUC= 0.55; SEM=0.011;
176 $p<.01$) decode this contrast for different unambiguous
177 tokens (A/C versus 9/8); suggesting that the response
178 is tracking the abstract letter/digit contrast, abstracted
179 from the specific pixel arrangement.

180 Fourth, trial difficulty (i.e. the distance between the
181 presented stimulus and the closest unambiguous character)
182 could be decoded between 270 and 1485 ms and peaked at
183 590 ms (12-regularized regression fit across sensors, $R=0.12$;
184 SEM=0.024; $p<0.01$). Difficulty signals were localized
185 more anteriorly than those of stimulus category ($x=12.58$;
186 $y=-91.44$; $z=-1.23$; $p<0.01$).

187 While spatial decoding led to significant clusters in 237
188 the temporal, parietal and prefrontal areas (Fig.2B), the 238
189 peak location of stimulus difficulty was highly variable 239
190 across subjects and included the dorso-parietal cortex, 240
191 the temporo-parietal junction and the anterior cingulate 241
192 cortex (Fig.2E). 242

193 Finally, temporal decoders of subjects' button press 243
194 (left versus right index fingers) were significant from 244
195 458 ms after stimulus onset and peaked at 604 ms 245
196 (AUC=0.85; SEM=0.011; $p<0.001$). A significant cluster 246
197 of motor signals could be detected around sensorimotor 247
198 cortices between 590 and 840 ms ($\bar{t}=4.98$, 248
199 $p<0.001$, Fig.2A). Response-lock analyses revealed 249
200 qualitatively similar but stronger results. For example, 250
201 temporal decoders were significant from 350 ms prior to 251
202 the response and up to 500 ms after the response reaching 252
203 an AUC of 94% at response time ($p<0.001$). 253

204 Overall, the time at which representations became 254
205 maximally decodable correlated with their peak location 255
206 along the postero-anterior axis (Fig.2D-E) ($r=0.57$, 256
207 $p<0.001$). These results thus strengthen the classic notion 257
208 that perceptual processes are hierarchically organized 258
209 across space, time and function. Importantly however, 259
210 this cascade of representations spreads over more than 260
211 600 ms and largely exceeds the time it takes the feedforward 261
212 response to ignite the ventral and dorsal pathways (Fig.1G and 262
213 Supplementary Video 1). 263

214 2.3. A hierarchy of recurrent layers explains the spatio- 265 215 temporal dynamics of neural representations 266

216 To clarify how a cascade of representations can be 268
217 generated over extended time periods, we propose to 269
218 distinguish feedforward and/or recurrent architectures 270
219 depending on (i) the spatial location, (ii) the timing and 271
220 (iii) the spatio-temporal dynamics of their representations 272
221 (Fig.3). This is done by simulating different architectures, 273
222 and assessing their similarity to the MEG data. In these models, 274
223 we assume that each 'layer' generates new hierarchical features, 275
224 in order to account for the organization of spatial decoders 276
225 (Fig.2E). Furthermore, we only discuss architectures which can 277
226 code for all representations simultaneously, in order to account 278
227 for the overlapping temporal decoding scores (Fig.2C). 279
228 Finally, we only model discrete activations (i.e. a representation 280
229 is either encoded or not) as any more subtle variation can be 281
230 trivially accounted for by signal-to-noise ratio considerations. 282

231 Each model predicts (1) 'source' decoding time courses 283
232 (i.e. what is decodable within each layer) and 284
233 (2) 'temporal generalization' (TG) maps. TG is used to 285
234 characterize the dynamics of neural representations and 286
235 287
236 288

consists in assessing the extent to which a temporal decoder 237
trained at a given time sample generalizes to other time 238
samples [24] (Fig.3D). 239

Our spatial and temporal decoding results can be 240
accounted for by a feedforward architecture that both 241
(i) generates new representations at each layer and 242
(ii) propagates low-level representations across layers 243
(Fig.3 Model 1: 'broadcast'). This architecture predicts 244
that representations would not be maintained within 245
brain areas. This lack of maintenance is not supported 246
by our data. First, the position of the stimulus was 247
decodable in the early visual cortex between 80-320 ms 248
($\bar{t}=5.18$, $p<.001$) and thus longer than the stimulus 249
presentation. Second, most temporal decoders significantly 250
generalized over several hundreds of milliseconds 251
(Fig.4A-B). For example, the temporal decoder trained to 252
predict stimulus position from $t=100$ ms could accurately 253
generalize until ≈ 500 ms as assessed with spatio-temporal 254
cluster tests across subjects (Fig.4A). Similarly, temporal 255
decoders of perceptual category and button-press generalized, 256
on average, for 287 ms (SEM=12.47; $p<.001$) and 689 ms 257
(SEM=30.94; $p<.001$) respectively. Given that the neural 258
activity underlying the decoded representations is partially 259
stable over several hundreds of milliseconds, recurrent 260
connections seem necessary to account our data (Fig.4 261
Model 2-4). 262

Consequently, we then considered a simple hierarchy 265
of recurrent layers, where recurrence only maintains 266
activated units (Fig.3 Model 2: 'maintain'). This architecture 267
predicts strictly square TG matrices (i.e. temporal 268
decoders would be equivalent to one another in terms of 269
their performance) and is thus at odds with the largely 270
diagonal TG matrices observed empirically (Fig.4A). 271
Specifically, the duration of significant temporal decoding 272
(fitting a new decoder at each time sample) was significantly 273
longer than the generalization of a single decoder to 274
subsequent time samples (e.g. 1,239 versus 287 ms for 275
perceptual category ($t=-61.39$; $p<.001$) and 1,215 versus 276
689 ms for button-press ($t=-16.26$; $p<.001$), Fig.4B). 277
These results thus suggest that the decoded representations 278
depend on dynamically-changing activity: i.e. each feature 279
is linearly coded by partially distinct brain activity 280
patterns at different time samples. 281

It is difficult to determine, with MEG alone, whether 282
such dynamic maintenance results from a change of neural 283
activity within or across brain areas. Indeed, Model 1 and 284
Model 3 can equally predict diagonal TG (Fig.3). However, 285
these two models, and their combination (Model 4) diverge 286
in terms of *where* information should be decodable. 287
Specifically, source analyses revealed that both stimulus 288
position and percep-

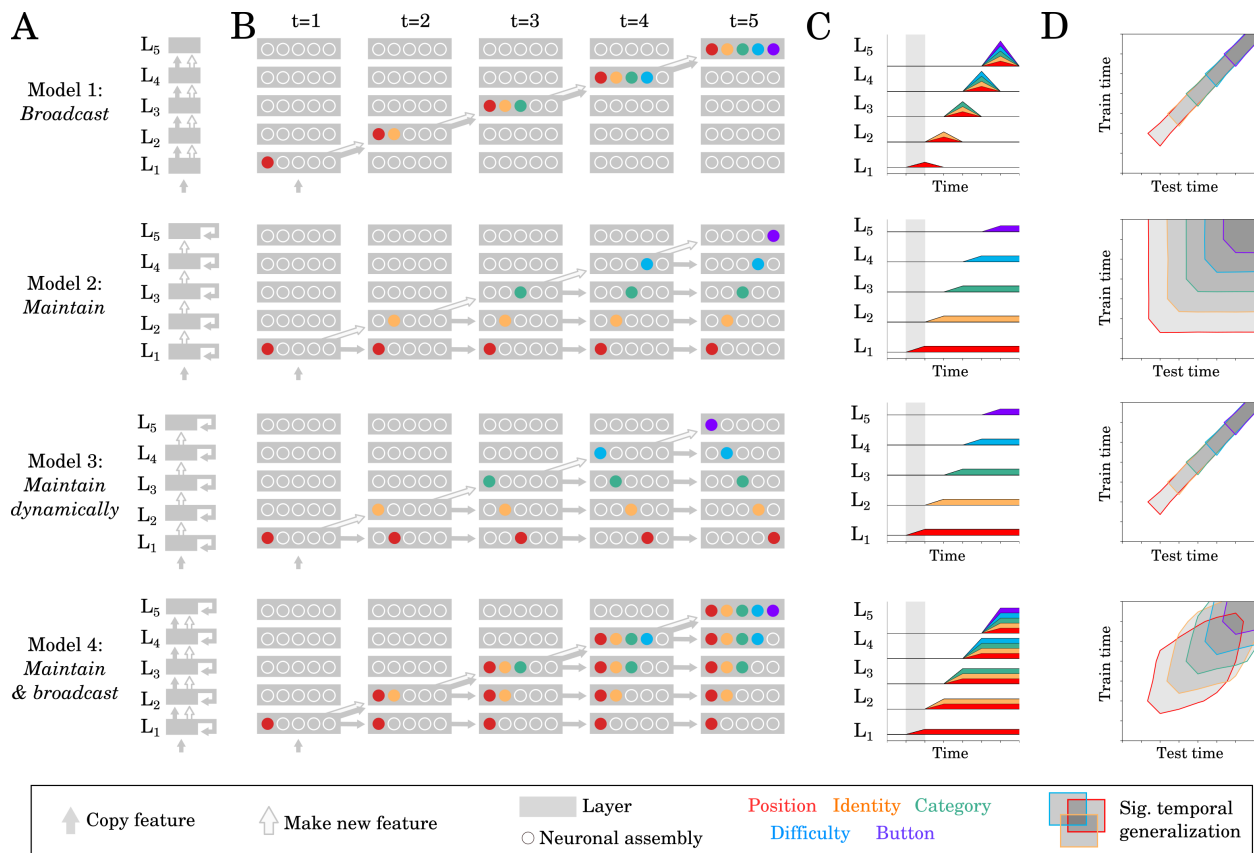


Figure 3: Source and temporal generalization predictions for various neural architectures. (A) Four increasingly complex neural architectures compatible with the spatial and temporal decoders of Fig.2. For each model (rows), the five layers (L1, L2 ... L5) generates new representations. The models differ in their ability to (i) propagate low-level representations across the hierarchy, (ii) maintain information with each layer in a stable or dynamic way. (B) Activations within each layer plotted at five distinct time samples. Dot slots indicate different neural assemblies within the same layer. Colors indicate which feature is linearly represented. For clarity purposes, only effective connections are plotted between different time samples. (C) Summary of the information represented within each layer across time. (D) Expected result for of the temporal generalization analyses, based on the processing dynamics of each model.

289 tual category can be decoded across a wide variety of 305
 290 partially-overlapping brain areas (Fig.2B, Supplemen- 306
 291 tary video 2), similarly to Model 4. Nonetheless, our 307
 292 MEG study remains limited in assessing whether within 308
 293 brain regions dynamics also contribute to the diagonal 309
 294 TG, which would suggest a mixture between models 3 310
 295 and 4. 311

296 Together, source and TG analyses thus suggest that 312
 297 the slow and sequential generation of increasingly ab- 313
 298 stract representations depends on a hierarchy of recur- 314
 299 rent layers that generate, maintain and broadcast repre- 315
 300 sentations across the cortex. 316

301 2.4. Hierarchical recurrence induces an accumulation 317 302 of delays 318

303 Can a hierarchy of recurrent processes account for 320
 304 single-trial dynamics? To address this issue, we hy- 321

305 pothesized that recurrent processes would take variable 306
 307 amounts of time to converge to each intermediary rep- 308
 309 resentation. In this view, (i) each feature is predicted 309
 310 to propagate across brain areas at distinct moments, and 310
 311 (ii) the successive rise of decodable representations is 311
 312 thus predicted to incrementally correlate with reaction 312
 313 times (Fig.5A-E). 313

314 To test this hypothesis, we estimated how the peak 314
 315 of each temporal decoder varied with reaction times. 315
 316 For clarity purposes, we split reaction times into four 316
 317 quantiles, and averaged the time courses of temporal 317
 318 decoders relative to their training time. These anal- 318
 319 yses showed that the latencies of (i) perceptual cate- 319
 320 gory ($r=0.35$; $p=0.006$), (ii) stimulus difficulty ($r=0.37$;
 321 $p=0.004$) and (iii) button press ($r=0.66$; $p<0.001$) in-
 322 creasingly correlated with reaction times (Fig.5F-G).

Overall, these results show that we can track with

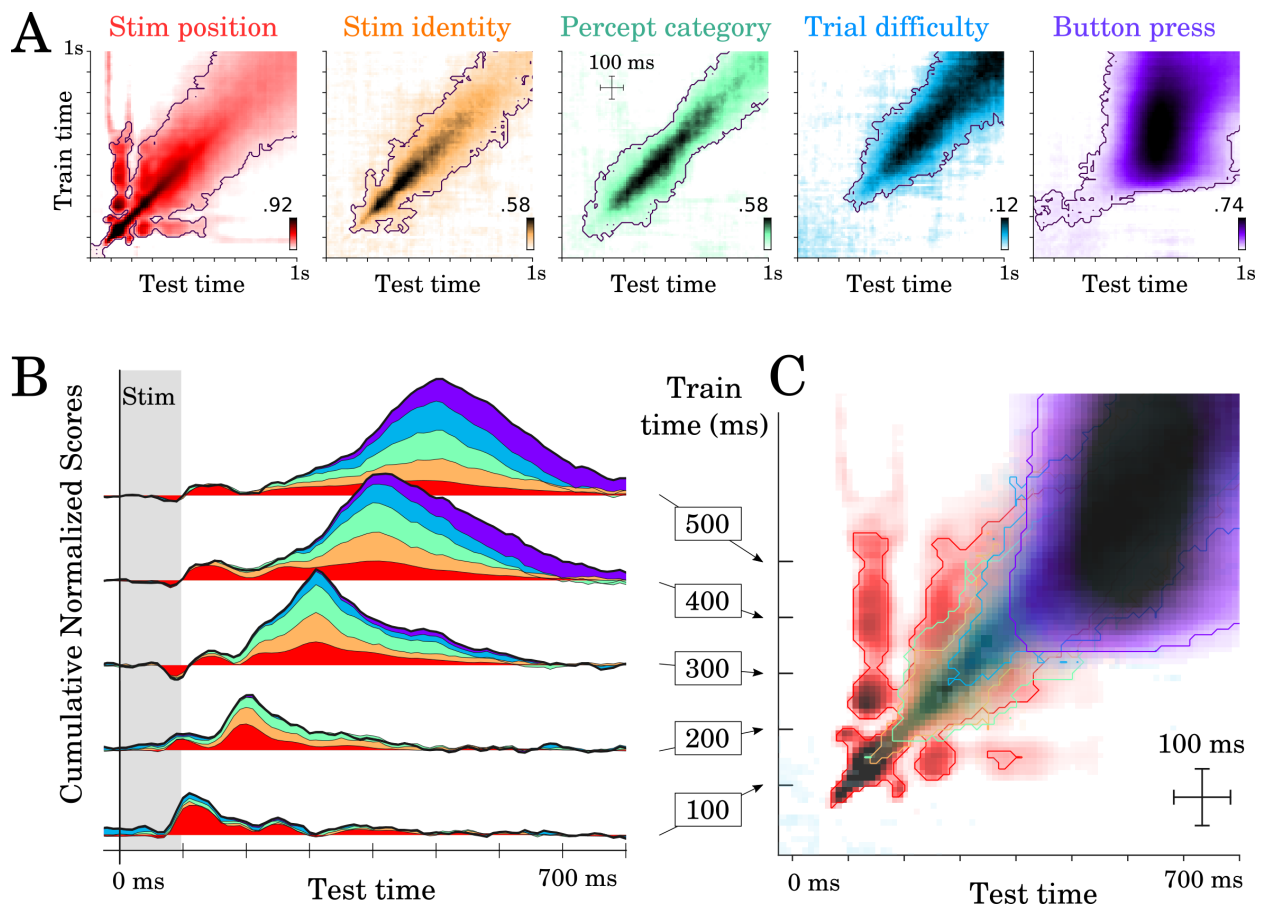


Figure 4: **Temporal generalization results.** (A) Temporal generalization for each of the five features orthogonally varying in our study. Color indicate decoding score (white=chance). Contours indicate significant decoding clusters across subjects. (B) Cumulative temporal generalization scores for the temporal decoders trained at 100, 200, 300, 400 and 500 ms respectively. These decoding scores are normalized by mean decoding peak for clarity purposes. (C) Same data as A but overlaid. For clarity purposes, contours highlight the 25th percentile of decoding performance.

322 MEG, a series of decisions generated by hierarchical
 323 recurrent processes. This neural architecture partially
 324 accounts for subjects' variable and relatively-slow
 325 reaction times.

326 2.5. Hierarchical recurrence implements a series of all- 327 or-none decisions

328 An architecture based on successive decisions pre-
 329 dicta a loss of ambiguous information akin to all-or-
 330 none categorization across successive processing stages
 331 (Fig.6A). To test this prediction, we quantified the ex-
 332 tent to which the decoding of 'percept category' and
 333 of 'motor action' varied linearly or categorically with
 334 (i) categorical evidence and (ii) motor evidence respec-
 335 tively (i.e. the extent to which the stimulus (i) objec-
 336 tively looks like a letter or a digit and (ii) should have
 337 led to a left or right button press given its pixels).

338 The probabilistic decoding predictions of percept cat-
 339 egory correlated linearly with sensory evidence be-
 340 tween 210 and 530 ms ($r=0.38 \pm 0.03$, temporal-cluster
 341 $p<0.001$). The spatial decoders fit from 200 to 400 ms
 342 clustered around the VWFA ($\bar{t}=4.6$; $p=.02$; 224 ver-
 343 tices) (Fig 6H). These results suggest that this region
 344 first represents the stimulus objectively (i.e. in its full
 345 ambiguity).

346 Between 400 and 810 ms, the predictions of 'per-
 347 ceptual category' decoders were better accounted for
 348 by sigmoidal ($r=0.77 \pm 0.03$, $199 p<0.001$) than by linear
 349 trends ($r=0.77 \pm 0.03$, $p<0.001$). Spatial decod-
 350 ing analyses restricted to the 500-700 ms time window
 351 was more distributed ($\bar{t}=4.4$; $p=.022$; 110 vertices). Fi-
 352 nally, ambiguous stimuli (steps 5 and 6 on the contin-
 353 uum) reached maximum decodability 205 ms later than
 354 unambiguous stimuli (steps 1 and 8) ($p<0.001$) (Fig.6J).
 355 The interaction between trend (linear or sigmoidal) and

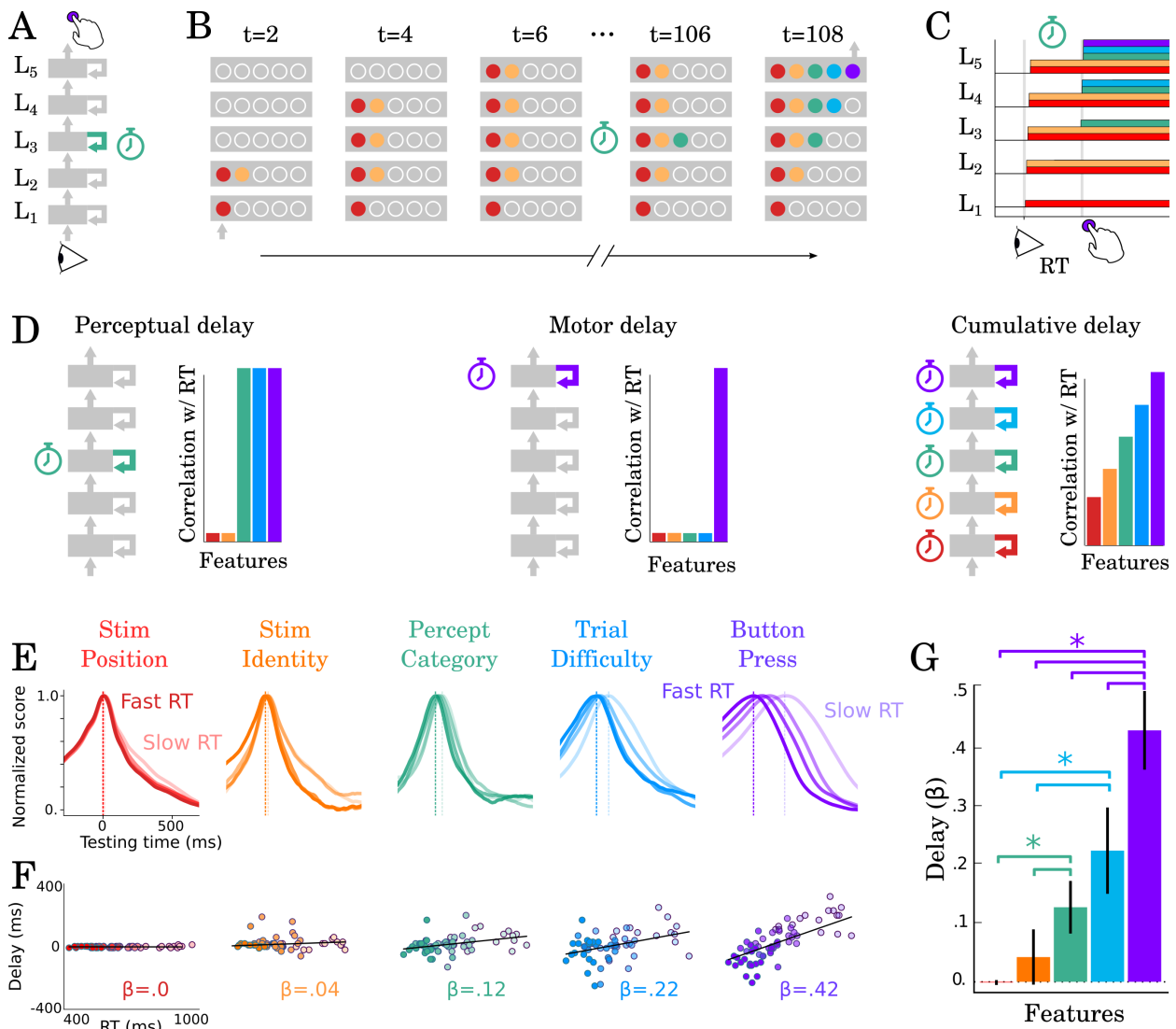


Figure 5: **Correlation between TG peaks and reaction times.** (A, B) Recurrent processing at a given processing stage is hypothesized to take a variable amount of time to generate adequate representations. (C) According to this hypothesis, the rise of the corresponding and subsequent representations would correlate with reaction times. (D, left) Predictions when delays are only induced by the perceptual stage of processing. (D, middle) Predictions when delays are only induced by the motor processing processing stage. (D, right) Predictions when delays are induced by all processing stages. (E) TG scores aligned to training time, split into trials within the fastest and slowest reaction-time quantile and averaged across reaction times bins. Dark and light lines indicate the average decoding performance for trials with fastest and slowest reaction times respectively. (F) Each subject (dot) mean peak decoding time (y-axis) as a function of reaction time (x-axis) color-coded from dark (fastest) to light (slowest). The beta coefficients indicate the average delay estimate. (G) The average slope between processing delay and reaction time for each feature. Error-bars indicate the SEM.

356 window latency was significant across subjects ($r=0.07$; 364
 357 SEM=0.01; $p=0.002$).

358 This progressive categorization of the letter/digit representations contrasts with the all-or-none pattern of 366
 359 motor signals. Specifically, the probabilistic predictions of button-press decoders varied categorically with response 367
 360 evidence from 440 to 1,290 ms (sigmoid > linear cluster, $\bar{t}=3.17$; $p<0.001$). There was also a more tran- 368
 361 362 363

364 sient linear trend from 410 to 580 ms ($\bar{t}=3.69$; $p<0.001$). 365
 366 This suggests that, unlike perceptual category, motor signals largely derive from categorical inputs.

367 Together, delay (Fig.5) and categorization (Fig. 6) 368
 369 analyses thus show that perceptual representations slowly become categorical and are subsequently fol- 370
 370 lowed by all-or-none motor representations.

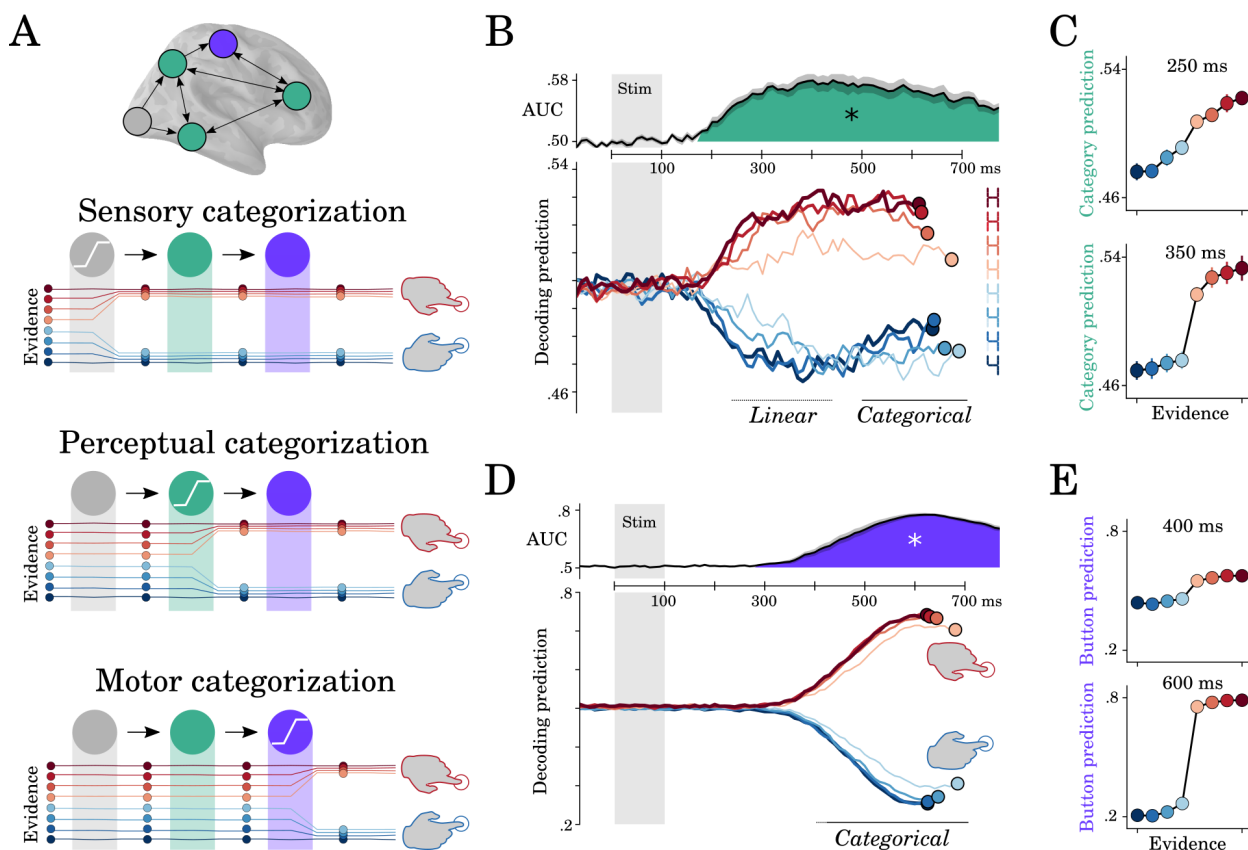


Figure 6: **Motor and perceptual decisions.** (A) Hypothesis space for when responses become categorical: during sensory, perceptual or motor processing. (B, top) Time course of decoding the perceptual decision. (B, bottom) Classifier predictions split into different levels of sensory evidence. (C) Averaging probabilities in different time-windows shows the linear-categorical shift in how information is represented. (D, top) Time course of decoding the motor decision. (D, bottom) Splitting classifier predictions into different levels of difficulty. (E) Different windows of classifier predictions, showing the categorical responses throughout processing.

371 3. Discussion

372 Our results show that briefly-flashed stimuli elicit a
 373 cascade of representations that spread well beyond the
 374 initial feedforward recruitment of the visual pathways.
 375 A hierarchy of recurrent processes that incrementally
 376 build representations best explains this cascade, its ac-
 377 cumulated delays and its all-or-none categorizations.

378 While macroscopic MEG signals advantageously
 379 give a birds-eye view of the cortical correlates of per-
 380 ceptual decision making, it should be stressed that their
 381 source reconstruction remains a coarse approximation.
 382 Consequently, identifying (1) the role of subcortical ar-
 383 eas and (2) the extent to which representations dynam-
 384 ically change within each brain area will necessitate in-
 385vasive brain recordings.

386 Nonetheless, our results bridge three important lines
 387 of research on the neural and computational bases of
 388 visual processing.

389 First, core-object recognition research, generally
 390 based on ≈ 100 ms-long image presentations has repeat-
 391 edly shown that the spiking responses of the inferotem-
 392 poral cortex is better explained by recurrent models than
 393 by feedforward ones [9, 14]. In particular, Kar et al have
 394 recently shown that images that are challenging to rec-
 395 ognize, lead to delayed content-specific spiking activity
 396 in the macaque's infero-temporal cortex [14]. Our find-
 397 ings, based on simpler but highly-controlled stimuli, are
 398 consistent with these results and further highlight that
 399 perceptual representations are not confined to the infero-
 400 temporal cortices, but also reach a large variety of pari-
 401 etal and prefrontal areas [25].

402 Second, the present study makes important contri-
 403 butions to the perceptual decision making literature
 404 [11, 13]. With some notable exceptions (e.g. [26]),
 405 this line of research primarily aims to isolate motor and
 406 supra-modal decision signals in the presence of sus-
 407 tained visual inputs: i.e. neural responses ramping to-

408 wards a virtual decision threshold, independently of the 454
409 representation on which this decision is based [13]. The 455
410 present study complements this approach by tracking 456
411 the representation-specific signals that slowly emerge 457
412 after a brief stimulus. Our results thus open an 458
413 exciting avenue for querying the gating mechanisms of 459
414 successive decisions and clarifying the role of the pre- 460
415 frontal areas in the coordination multiple perceptual and 461
416 supramodal modules [27]. 462

417 Finally, our results constitute an important confirma- 463
418 tion of modern theories of perception. In particular, the 464
419 Global Neuronal Workspace Theory predicts that percep- 465
420 tual representations need to be broadcast to asso- 466
421 ciative cortices via the fronto-parietal areas to lead to 467
422 subjective reports [8]. Yet, at some notable exceptions 468
423 [28, 29], previous studies often fail to dissociate percep- 469
424 tual contents and perceptual reports (e.g. [30, 31]). By 470
425 contrast, the present experimental design allows an un- 471
426 precedented dissection of the distinct processing stages 472
427 that transform sensory input into perceptual represen- 473
428 tations and, ultimately, actions. The generation of let- 474
429 ter and digit representations in the dedicated brain areas 475
430 [32, 17] and their subsequent broadcast to the cortex re- 476
431 inforce the notion that subjective perception relate to the 477
432 global sharing of content-specific representations across 478
433 brain areas [8, 33]. 479

434 4. Method

435 4.1. Target stimuli

436 Using the font designed in [15], the stimuli were 481
437 made from 0, 4, 5, 6, 8, 9, A, C, E, H, O, S, or from 482
438 a linear combination of two of these characters varying 483
439 in a single black bar (hereafter ‘pixel’). The correspond- 484
440 ing ‘morphs’ were created by adjusting the contrast of 485
441 the remaining pixel along eight equally spaced steps be- 486
442 tween 0 (no bar) and 1 (black bar). 487

443 4.2. Experiment 1

444 Eight subjects with normal or corrected vision, seated 484
445 ≈ 60 cm from a 19” CRT monitor (60Hz refresh rate, res- 485
446 olution: 1024x768), performed a stimulus identification 486
447 task with continuous judgements across 28 variably am- 487
448 biguous stimuli generated from digit stimuli. Ten euros 488
449 were provided in compensation for this 1-hour experi- 489
450 ment. 490

451 Subjects performed four blocks of 50 trials, each or- 485
452 ganized in the following way. After a 200 ms fixation, a 486
453 target stimulus, randomly selected from one of the 28 487

stimuli, was flashed for 83 ms on a 50% gray back-
ground to the left or to the right of fixation. The ori-
entation of the reporting disk (e.g 5-6-8-9 versus 5-9-
8-6) was counterbalanced across subjects. Subjects had
then up to 10 seconds to move a cursor on a large disk
to report their percepts. The radius on the disk indi-
cated subjective visibility (center=did not see the stim-
ulus, disk border=max visibility). The angle on the disk
indicated subjective identity (e.g. 5, 6, 8, 9 for the top
left, top right, bottom right, and bottom left ‘corners’
respectively). Inter-trial interval was 500 ms. To ver-
ify that subjects provided meaningful reports, the tar-
get stimulus was absent 15% of the trials. Absent trials
were rated with a low visibility (i.e radius below 5%
of the disk radius) in most cases. Absent trials and trials
reported with a low visibility were excluded from subse-
quent analyses. The report distribution plotted in Fig.1B
were generated with Seaborn’s bivariate Gaussian ker-
nel density estimate function with default parameters.

Modeling categorical reports. To test whether subjec-
tive reports of stimulus identity varied linearly or cat-
egorically with sensory evidence, we analyzed how re-
ports’ angle (i.e. subjective identity) varied with the ex-
pected angle given the stimulus (i.e. sensory evidence).

For each morph (5-6, 5-8, 9-8 and 6-8) separately, we
fit a linear model:

$$\hat{y} \leftarrow \beta_1 x + \beta_0 \quad (1)$$

480 and a sigmoidal model:

$$\hat{y} \leftarrow \frac{1}{1 + \exp(\beta_1 x + \beta_2)} + \beta_0 \quad (2)$$

481 where \hat{y} is the report angle predicted by the model, x
482 is expected angle given the stimulus pixels and β_0 is a
483 free bias parameter.

To minimize the effects of noise, behavioral reports
were first averaged within each level of evidence, sorted
from the stimulus with the least pixels (e.g. 5, in 5-6
morph) to the stimulus with the most pixels (e.g. 6 in
the 5-6 morph). The resulting averages were normal-
ized to range between 0 and 1 within each subject. The
 β parameters were fit with Scipy’s ‘curve_fit’ function
[34] to minimize a mean squared error across trials i :

$$\operatorname{argmin}_{\beta} \sum_i (y_i - \hat{y}_i)^2 \quad (3)$$

484 Because the linear and sigmoidal models have dis-
485 tinct numbers of free parameters, we compared them
486 within a 5-split cross-validation. Specifically, the two
487 models were repeatedly fit and tested on independent

488 trials. A Pearson correlation coefficient r summarised
489 the ability of each model to accurately predict y_{test} given
490 x_{train} , y_{train} and x_{test} . Finally, a Wilcoxon test was ap-
491 plied across subjects to test whether the two models
492 were consistently above chance ($r > 0$) and consistently
493 different from one another ($r_{sigmoid} > r_{linear}$).

494 *Experiment 2.* This experiment was performed at Neu-
495 rospin, Gif sur Yvette, thanks to the support of Stanis-
496 las Dehaene. Seventeen subjects performed a discrete
497 identification task across 22 variably ambiguous stimuli
498 generated from letters and digits inside an Elekta Neu-
499 romag MEG scanner (204 planar gradiometers and 102
500 magnetometers). Seventy euros were provided in com-
501 pensation to the 1-hour experiment and 30 minutes of
502 preparation.

503 Participants' head shape was digitized along with five
504 fiducial points on the forehead and on each aural canal.
505 Five head-position coils were placed on subjects head
506 and localized at the beginning of each block.

507 The trial structure was as follows. A black fixation
508 cross was displayed on a 50% gray background for 300
509 ms followed by a 100ms-long target stimulus presented
510 on the left or on the right of fixation. Two task-irrelevant
511 flankers (e.g. stimulus can be read as an S or a 5) were
512 displayed on the side of this target stimulus to increase
513 our chances of eliciting recurrent processing via crowd-
514 ing [18]. Subjects were given two seconds to report the
515 identity of the stimulus. Reports of stimulus identity
516 were given by pressing a button with the left and right
517 index fingers respectively. The identity-button mapping
518 changed on every block to orthogonalize the neural cor-
519 relates of stimulus identity and the neural correlates of
520 motor actions. For example, in block 1, perceiving an
521 E or a 4 should have been reported with a left button
522 press, whereas in block 2, E and 4 should have been
523 reported with a right button press. The identity-button
524 was explicitly reminded before each block. In addition,
525 a visual feedback was displayed after non-ambiguous
526 trials. Specifically, the fixation turned green for 100ms
527 or red for 300 ms in correct and incorrect trials respec-
528 tively. The brain responses to these feedback stimula-
529 tions are not analyzed in the present study. Inter-trial
530 interval was 1 second. Subjects were provided a short
531 training to ensure they understood the task, and identi-
532 fied non-ambiguous targets at least 80% of the time.

533 A total of 1920 trials, grouped into 40 blocks, were
534 performed by each subject, 320 of which were presented
535 passively at the end of each block – subjects were not
536 required to provide a response. The trial structure was
537 generated by (i) permuting all combinations of stimulus
538 features (e.g. position, identity, response mapping, dif-

539 ficulty), and (ii) shuffling the order of presentation for
540 each subject. The experiment was presented using Psy-
541 chtoolbox [35].

542 All experiments were approved by the local ethics
543 committee. All subjects signed an informed consent
544 form.

545 4.3. Structural MRI

546 For each subject, an anatomical MRI with a resolu-
547 tion of $1 \times 1 \times 1.1$ mm was acquired after the MEG exper-
548 iment with a 3T Siemens scanner. Gray and white mat-
549 ter were segmented with Freesurfer 'recon-all' pipeline
550 [20] and coregistered with each subject's digitized head
551 shapes along with fiducial points.

552 4.4. Preprocessing

553 The continuous MEG recording was noise-reduced
554 using Maxfilter's SSS correction on the raw data,
555 bandpass-filtered between 0.5 and 40 Hz using MNE-
556 Python's default parameters with firwin design [21] and
557 downsampled to 250 Hz. Epochs were then segmented
558 between -300 ms and +1500 ms relative to stimulus on-
559 sets.

560 After coregistering the MEG sensor data with sub-
561 jects' structural MRI and the head position coils, we
562 computed the forward model using a 1-layer (inner
563 skull) boundary element model, for each subject sep-
564 arately and fit a minimum-norm inverse model (signal
565 to noise ratio: 3, loose dipole fitting: 0.2, with normal
566 orientation of the dipole relative to the cortical sheet)
567 using the noise covariance across sensors averaged over
568 the pre-stimulus baseline across trials. Finally, the in-
569 verse model was applied to single-trial data resulting in
570 a dynamic Statistical Parameter Map (dSPM) [19] value
571 for each source at each time sample.

572 4.5. Modeled features

573 We investigated whether single-trial source and sen-
574 sor evoked responses varied as a function of five fea-
575 tures: (1) the position of the stimulus on the computer
576 screen (left versus right of fixation), (2) the morph
577 from which the stimulus is generated (E-6 versus H-4), (3) the
578 category of the stimulus (letter versus digit), (4) the dif-
579 ficulty of the trial (maximum difficulty = stimuli with
580 pixel at 50% contrast; minimum difficult stimuli with
581 pixels at 0% or 100% contrast), and (5) the response
582 button used to report the stimulus (left versus right but-
583 ton). By design, these five features are independent of
584 one another.

585 It is challenging to dissociate brain responses that
586 represent objective sensory information from those that

587 represent perceptual decisions as the two are generally 615
588 collinear. To address this issue, we first fit univariate and 616
589 multivariate models to predict perceptual category: i.e. 617
590 whether the button press indicated a character that be- 618
591 longs to the digit or to the letter category. This feature is 619
592 independent of the button press (e.g. the letter E and the 620
593 digit 4 can be reported with the same button in a given 621
594 block). Furthermore, this feature is not necessary to per- 622
595 form the task (i.e. knowing whether E and H are letters 623
596 is unnecessary to discriminate them). We reasoned that 624
597 if subjects automatically generates letter/digit represen- 625
598 tations during perceptual categorization, then we should 626
599 be able to track the generation of this abstract feature 627
600 from brain activity. 628

601 4.6. Mass univariate statistics 630

To estimate whether brain responses correlated with 631
each of these five features, we first fit, within each sub- 632
ject, mass univariate analyses at each source location 633
and for each time sample with a linear regression: 634

$$\beta = (X^T X)^{-1} X y \quad (4) \quad 635$$

602 where $X \in \mathbb{R}^{n \times f}$ is a design matrix of n epochs by 636
603 $f = 5$ features and $y \in \mathbb{R}^n$ is the univariate brain re- 637
604 sponse at a given source and at given time. The effect 638
605 sizes β were then passed to second-level statistics across 639
606 subjects corrected for multiple comparisons using non- 640
607 parametric spatio-temporal cluster testing (see below). 641

608 4.7. Decoding 644

Decoding analyses consists in predicting each feature 645
from multivariate brain responses. Decoding analyses 646
were performed within a 5-split stratified K-Fold cross- 647
validation using l2-regularized linear models. Classifiers 648
consisted of logistic regressions (with scikit-learn
[36]’s default parameters: $C = 1$):

$$\operatorname{argmin}_{\beta} \sum_i \log(1 + \exp(-y_i \beta^T x_i)) + C \|\beta\|^2 \quad (5) \quad 649$$

609 where $y_i \in \{\pm 1\}$ is the feature to be decoded at trial i 646
610 and x_i is the multivariate brain response. 647

Regressors consisted of ridge regression (with scikit- 648
learn [36]’s default parameters: $\alpha = 1$).

$$\operatorname{argmin}_{\beta} \sum_i (y_i \beta^T x_i)^2 + \alpha \|\beta\|^2 \quad (6) \quad 649$$

611 For each subject independently, decoding perfor- 652
612 mance was summarized across trials, with an area under 653
613 the curve (AUC) and a Spearman r correlation score for 654
614 classifiers and regressors respectively. 655

All decoders were provided with data normalized by 615
the mean and the standard deviation in the training set. 616

Spatial decoding consists in fitting a series of de- 617
coders at each brain source independently, across all 618
1,500 time samples relative to stimulus onset. This anal- 619
ysis results in a decoding brain map that indicates where 620
a feature can be linearly decoded in the brain. These 621
decoding maps were then passed to cluster-corrected 622
second-level statistics across subjects. 623

Temporal decoding consists in fitting a series of de- 624
coders at each time sample independently, across all 625
306 MEG sensors. This analysis results in a decod- 626
ing time course that indicates when a feature can be lin- 627
early decoded from MEG signals. These decoding time 628
courses were then passed to cluster-corrected second- 629
level statistics across subjects. 630

Temporal generalization (TG) consists in testing 631
whether a temporal decoder fit on a training set at time 632
 t can decode a testing set at time t' [24]. TG can be 633
summarized with a square training time \times testing time 634
decoding matrix. To quantify the stability of neural re- 635
presentations, we measured the duration of above-chance 636
generalization of each temporal decoder. To quan- 637
tify the dynamics of neural representations, we com- 638
pared the mean duration of above-chance generaliza- 639
tion across temporal decoders to the duration of above- 640
chance temporal decoding (i.e. the diagonal of the ma- 641
trix versus its rows). These two metrics were assessed 642
within each subject and tested with second-level statis- 643
tics across subjects. 644

645 4.8. Linear versus Categorical

To test whether neural representations varied as a 646
function of (i) reaction times (RTs, split into 4 quan- 647
tiles), (ii) sensory evidence (i.e. the extent to the stim- 648
ulus objectively corresponds to a letter) and (iii) motor 649
evidence (i.e. whether the stimulus should have led to 650
the left button press), we analyzed the extent to which 651
decoders’ predictions covaried with each of these three 652
variables z :

$$f(z, \beta^T X) \quad (7) \quad 653$$

646 where f is a linear or a sigmoidal model, X is the 647
multivariate brain response and β is the decoder’s coef- 648
ficient fit with cross-validation. 649

649 4.9. Statistics

Univariate, decoding and TG models were fit within 650
subjects, and tested across subjects. In case of repeated 651
estimates (e.g. temporal decoding is repeated at each 652
time sample), statistics derived from non-parametric 653
cluster-testing with 10,000 permutations across subjects 654
with MNE-Python’s default parameters [21]. 655

656 *Simulations.* To formalize how distinct neural archi-
657 tectures lead to distinct spatio-temporal dynamics, we
658 modeled discrete linear dynamical systems forced with
659 a transient input U . Specifically:

$$X_{t+1} = AX_t + BU_t \quad (8)$$

660 where X is a multidimensional times series (i.e. neu-
661 rons x time), A is the architecture, and corresponds to
662 square connectivity matrix (i.e. neurons x neurons), B
663 is an input connectivity matrix (i.e. inputs x neurons),
664 and U is the input vector.

665 Distinct architectures differ in the way units are con-
666 nected with one another. For simplicity purposes, we
667 order units in the A matrix such that their row index
668 correspond to their hierarchical levels.

669 In this view, the recurrent, feedforward and skip con-
670 nnections of the architecture A were modeled as a bi-
671 nary diagonal matrix R , a shift matrix F and a matrix
672 S with 1 entries in the last column respectively. These
673 three matrices were modulated by specific weights, as
674 detailed below. The input U was only connected to the
675 first "processing stage", i.e. to the first unit(s) of A , via a
676 matrix B constant across architectures, and consisted of
677 a transient square-wave input, that mimics the transient
678 flash of the stimulus onto subjects' retina.

679 To model multiple features, we adopted the same pro-
680 cedure with multiple units per layer. Each unit within
681 each layer was forced to encode a specific feature.

682 Each architecture was fed an input at $t=1$, and simu-
683 lated for 8 time steps. Finally, temporal generalization
684 analyses based on the architectures' activations were ap-
685 plied for each of the features.

686 5. Acknowledgement

687 This project received funding from the European
688 Union's Horizon 2020 research and innovation program
689 under grant agreement No 660086, the Bettencourt-
690 Schueller Foundation, the Fondation Roger de Spoel-
691 berch, the Philippe Foundation and the Abu Dhabi In-
692 stitute G1001. We are infinitely grateful to Stanislas
693 Dehaene for his support.

- 694 [1] Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology* **160**, 106–154 (1962).
695
696 [2] Maunsell, J. & van Essen, D. C. The connections of the middle temporal visual area (mt) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience* **3**, 2563–2586 (1983).
697
698 [3] Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nature neuroscience* **2**, 1019 (1999).
699
700 [4] DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
701
702 [5] Yamins, D. L. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* **111**, 8619–8624 (2014).
703
704 [6] Schrimpf, M. *et al.* Using brain-score to evaluate and build neural networks for brain-like object recognition. In *Cosyne 19, Date: 2019/02/28-2019/03/03, Location: Lisbon, Portugal* (2019).
705
706 [7] Ratcliff, R. & Smith, P. L. A comparison of sequential sampling models for two-choice reaction time. *Psychological review* **111**, 333 (2004).
707
708 [8] Dehaene, S. & Changeux, J.-P. Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227 (2011).
709
710 [9] Lamme, V. A. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences* **23**, 571–579 (2000).
711
712 [10] Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334 (1989).
713
714 [11] Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annual review of neuroscience* **30** (2007).
715
716 [12] Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of neurophysiology* **86**, 1916–1936 (2001).
717
718 [13] O'connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature neuroscience* **15**, 1729 (2012).
719
720 [14] Kar, K., Kubilius, J., Schmidt, K., Issa, E. B. & DiCarlo, J. J. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience* **22**, 974 (2019).
721
722 [15] King, J.-R. & Dehaene, S. A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130204 (2014).
723
724 [16] Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends in cognitive sciences* **15**, 254–262 (2011).
725
726 [17] Shum, J. *et al.* A brain area for visual numerals. *Journal of Neuroscience* **33**, 6709–6715 (2013).
727
728 [18] Strasburger, H., Rentschler, I. & Jüttner, M. Peripheral vision and pattern recognition: A review. *Journal of vision* **11**, 13–13 (2011).
729
730 [19] Dale, A. M. *et al.* Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *Neuron* **26**, 55–67 (2000).
731
732 [20] Fischl, B. Freesurfer. *Neuroimage* **62**, 774–781 (2012).
733
734 [21] Gramfort, A. *et al.* Mne software for processing meg and eeg data. *Neuroimage* **86**, 446–460 (2014).
735
736 [22] Hagler Jr, D. J. & Sereno, M. I. Spatial maps in frontal and prefrontal cortex. *Neuroimage* **29**, 567–577 (2006).
737
738 [23] Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).
739
740 [24] King, J. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences* **18**, 203–210 (2014).
741
742 [25] Freedman, D. J. & Miller, E. K. Neural mechanisms of visual categorization: insights from neurophysiology. *Neuroscience & Biobehavioral Reviews* **32**, 311–329 (2008).
743
744 [26] Philiastides, M. G. & Sajda, P. Eeg-informed fmri reveals spatiotemporal characteristics of perceptual decision making. *Journal of Neuroscience* **27**, 13082–13091 (2007).
745
746 [27] Sarafyazd, M. & Jazayeri, M. Hierarchical reasoning by neural circuits in the frontal cortex. *Science* **364**, eaav8911 (2019).
747
748 [28] Tong, F., Nakayama, K., Vaughan, J. T. & Kanwisher, N. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**, 753–759 (1998).
749
750 [29] King, J.-R., Pescetelli, N. & Dehaene, S. Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron* **92**, 1122–1134 (2016).
751
752 [30] Sergent, C., Baillet, S. & Dehaene, S. Timing of the brain events underlying access to consciousness during the attentional blink. *Nature neuroscience* **8**, 1391 (2005).
753
754 [31] Van Vugt, B. *et al.* The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* **360**, 537–542 (2018).
755
756 [32] Cohen, L. *et al.* The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307 (2000).
757
758 [33] Lamme, V. A. Why visual attention and awareness are different. *Trends in cognitive sciences* **7**, 12–18 (2003).
759
760 [34] Jones, E., Oliphant, T., Peterson, P. *et al.* SciPy: Open source scientific tools for Python (2001–). URL <http://www.scipy.org/>. [Online; accessed 10day].
761
762 [35] Kleiner, M. *et al.* What's new in psyctoolbox-3. *Perception* **36**, 1 (2007).
763
764 [36] Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *Journal of machine learning research* **12**, 2825–2830 (2011).