

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 **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 activations 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

²¹ visual hierarchy [8, 9].

²² A large body of research shows that recurrent processing accounts for such behavioral and neural dynamics [9, 10, 11, 12, 13, 14, 15, 16, 17]. In this view, recurrent processing would mainly consist in accumulating sensory evidence until a decision to act is triggered [13].

²⁸ 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.

³³ 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,

41 over extended time periods, a cascade of categorical
42 decisions that ultimately predicts subjects' perceptual
43 reports.

44 2. Results

45 2.1. Subjective reports of stimulus identity are categor- 46 ical

47 To investigate the brain and computational bases of
48 perceptual recognition, we used visual characters as de-
49 scribed in [18]. These stimuli can be parametrically
50 morphed between specific letters and digits by varying
51 the contrast of their individual edges, hereafter referred
52 to as pixels (Fig.1A-B).

53 To check that these stimuli create categorical per-
54 cepts, we asked eight human subjects to provide contin-
55 uous subjective reports by clicking on a disk after each
56 stimulus presentation (Experiment 1. Fig.1A). The ra-
57 dius and the angle of the response on this disk indicated
58 the subjective visibility and the subjective identity of
59 the stimulus respectively. We then compared (i) the re-
60 ported angle with (ii) the stimulus evidence (i.e. the ex-
61 pected angle given the pixels) for each morph separately
62 (e.g. 5-6, 6-8, etc). Subjective reports were categori-
63 cal: cross-validated sigmoidal models better predicted
64 subjects' responses ($r=0.49+/-0.05$, $p=0.002$) than lin-
65 ear models ($r=0.46+/-0.043$, $p=0.002$, sigmoid>linear:
66 $p=0.017$ Fig.1B-C).

67 We adapted this experimental paradigm for an MEG
68 experiment by modifying three main aspects (Experi-
69 ment 2). First, we used stimuli that could be morphed
70 between letters and digits, to trigger macroscopically
71 distinguishable brain responses in the visual word form
72 area (VWFA) and number form area (NFA) [19, 20].
73 Second, we added two task-irrelevant flankers next to
74 the target stimulus (Fig.1D) to increase our chances of
75 eliciting recurrent processes via crowding [21]. Third, a
76 new set of seventeen subjects reported subjective iden-
77 tity via a two-alternative forced-choice button press.
78 The identity-response mapping was orthogonal to the
79 letter/digit category and changed on every block of 48
80 trials. There were 1920 trials total, 320 of which were
81 presented passively and did not require a response.

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

91 2.2. Neural representations are functionally organized 92 over time and space

93 Here, we aimed to decompose the sequence of de-
94 cisions that allow subjects to transform raw visual in-
95 put into perceptual reports. To this aim, we localized
96 the MEG signals onto subjects' structural MRI with dy-
97 namic statistical parametric mapping (dSPM, [22]), and
98 morphed these source estimates onto a common brain
99 coordinate [23, 24]. The results confirmed that the stim-
100 ulti elicited, on average, a sharp response in the primary
101 visual areas around 70 ms, followed by a fast feedfor-
102 ward response along the ventral and dorsal visual path-
103 ways within the first 150-200 ms. After 200 ms, the ac-
104 tivity appeared sustained and widely distributed across
105 the associative cortices up until 500-600 ms after stim-
106 ulus onset (Fig.1G and Supplementary Video 1).

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

114 First, we aimed to identify when and where low-level
115 visual features would be represented in brain activity.
116 To do so, we estimated, at each time sample separately,
117 the ability of an l2-regularized logistic regression to pre-
118 dict, from all MEG sensors, the position of the stimu-
119 lus on the computer screen (left versus right). Stimulus
120 position was decodable between 41 and 1,500 ms and
121 peaked at 120 ms (AUC=0.94; SEM=0.007; $p<0.001$
122 as estimated with second-level non-parametric tempo-
123 ral cluster test across subjects, (Fig.2C). These signals
124 peaked in the early visual cortex (mean MNI ($x=27.59$;
125 $y=-74.15$; $z=-1.07$)), and propagated along the ventral
126 and dorsal streams during the first 200 ms (Fig.2A, sup-
127 plementary video). To summarize where stimulus po-
128 sition was represented in the brain, we implemented
129 'spatial decoders': l2-regularized logistic regressions fit
130 across all time samples (0 - 1,500 ms) for each esti-
131 mated brain source separately. Spatial decoding peaked
132 in early visual areas and was significant across a large
133 variety of visual and associative cortices as estimated
134 with a second-level non-parametric spatial cluster test
135 across subjects (Fig.2B), confirming the retinotopic or-
136 ganization of the visual hierarchy [25, 26].

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

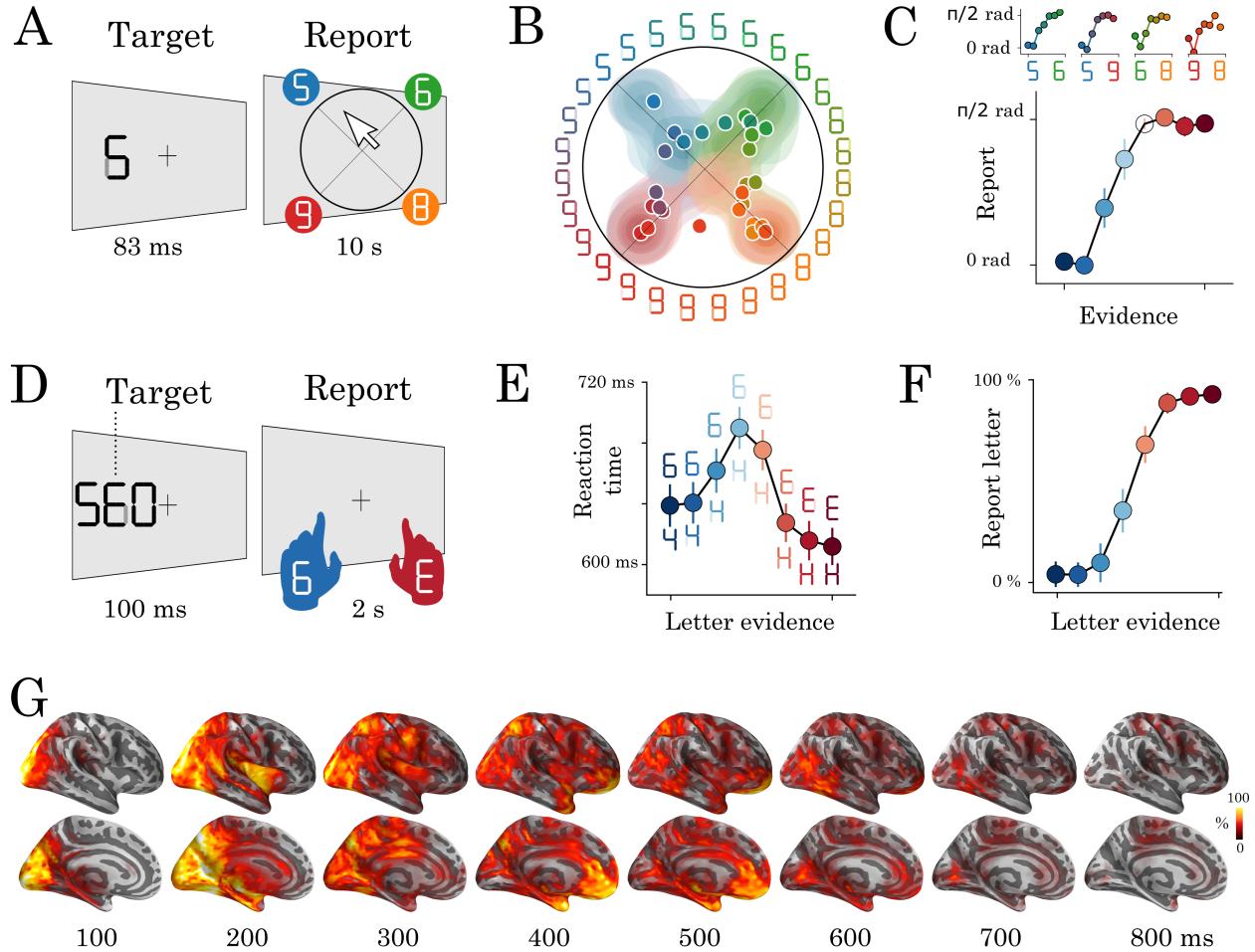


Figure 1: Experimental protocol and behavioral results. Experiment 1: 8 human subjects provided perceptual judgments on variably ambiguous digits briefly flashed at the center of a computer screen (A). Reports were made by clicking on a disk, where (i) the radius and (ii) the angle on the disk indicate (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.

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

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

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

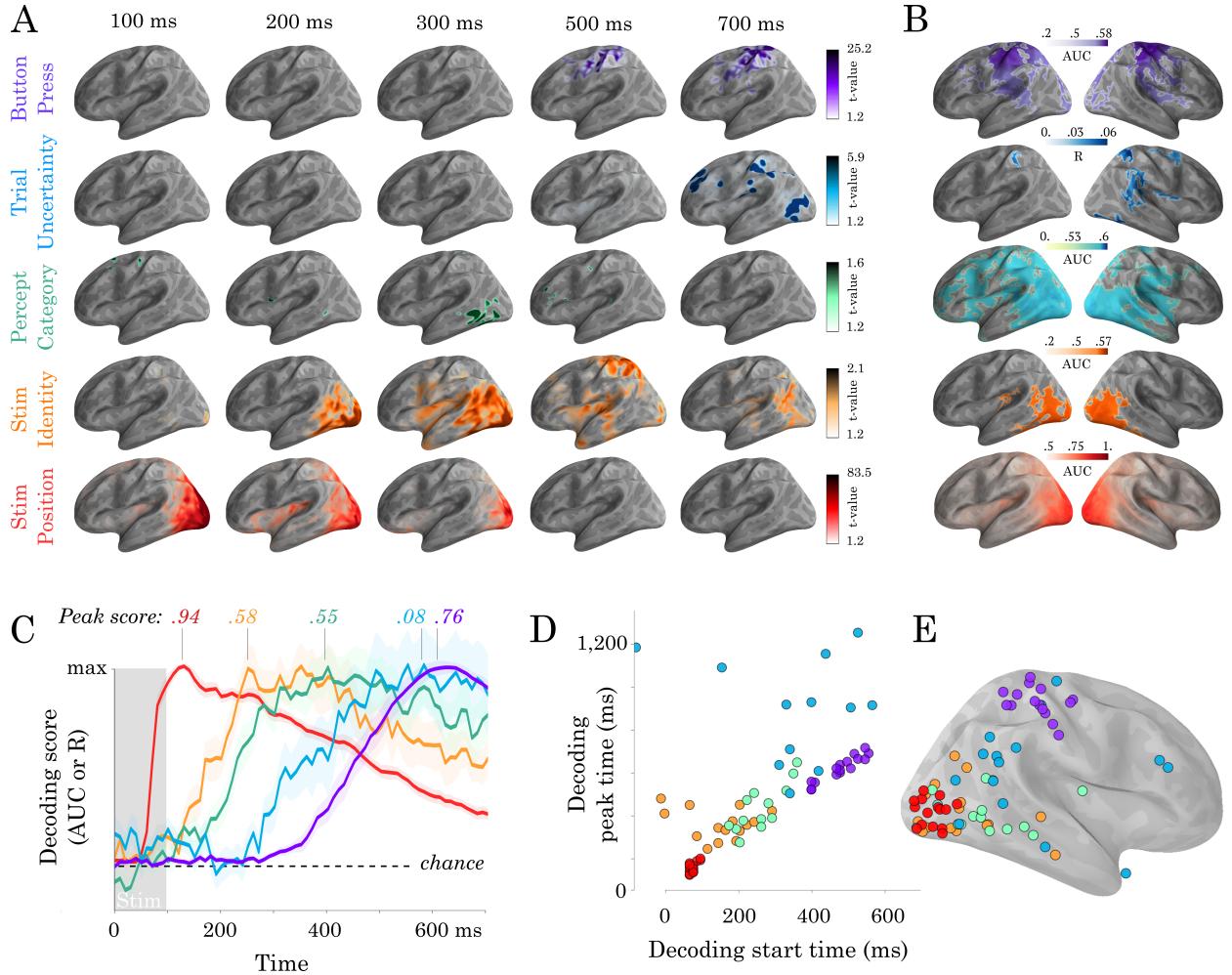


Figure 2: Spatio-temporal hierarchy. (A) Mass-univariate statistics. Each row plots the average-across-subjects beta coefficients obtained from regression between single-trial evoked activity and each of the five features orthogonally varying in this study. These results are displayed in Supplementary 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).

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

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

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

Fourth, trial uncertainty (i.e. the objective distance between the presented stimulus and the closest unambiguous character) could be decoded between 270 and 1485 ms and peaked at 590 ms (l2-regularized regression fit across sensors, $R=0.12$; SEM=0.024; $p<0.01$). Uncertainty signals were localized more anteriorly than those of stimulus category ($x=12.58$; $y=-91.44$; $z=$

188 1.23; $p<0.01$). While spatial decoding led to significant
189 clusters in the temporal, parietal and prefrontal areas
190 (Fig.2B), the peak location of stimulus uncertainty was
191 highly variable across subjects and included the dorso-
192 parietal cortex, the temporo-parietal junction and the an-
193 terior cingulate cortex (Fig.2E).
194

Finally, temporal decoders of subjects' button press
(left versus right index fingers) were significant from
458 ms after stimulus onset and peaked at 604 ms
(AUC=0.85; SEM=0.011; $p<0.001$). A significant clus-
ter of motor signals could be detected around sen-
sorimotor cortices between 590 and 840 ms ($\bar{t}=4.98$,
 $p<0.001$, Fig.2A). Response-lock analyses revealed
qualitatively similar but stronger results. For example,
temporal decoders were significant from 350 ms prior to
the response and up to 500 ms after the response reach-
ing an AUC of 94% at response time ($p<0.001$).
200

Overall, the time at which representations became
maximally decodable correlated with their peak loca-
tion along the postero-anterior axis (Fig.2D-E) ($r=0.57$,
 $p<0.001$). These results thus strengthen the classic no-
tion that perceptual processes are hierarchically orga-
nized across space, time and function. Importantly how-
ever, this cascade of representations spreads over more
than 600 ms and largely exceeds the time it takes the
feedforward response to ignite the ventral and dorsal
pathways (Fig.1G and Supplementary Video 1).
214

215 2.3. A hierarchy of recurrent layers explains the spatio- 216 temporal dynamics of neural representations

To clarify how a cascade of representations can be
generated over extended time periods, we propose to
distinguish feedforward and/or recurrent architectures
depending on (i) the spatial location, (ii) the timing and
(iii) the spatio-temporal dynamics of their representa-
tions (Fig.3). This is done by simulating different ar-
chitectures, and assessing their similarity to the MEG
data. In these models, we assume that each 'layer' gen-
erates new hierarchical features, in order to account for
the organization of spatial decoders (Fig.2E). Further-
more, we only discuss architectures which can code for
all representations simultaneously, in order to account
for the overlapping temporal decoding scores (Fig.2C).
Finally, we only model discrete activations (i.e. a rep-
resentation is either encoded or not) as any more subtle
variation can be trivially accounted for by signal-to-
noise ratio considerations.
233

Each model predicts (1) 'source' decoding time
courses (i.e. what is decodable within each layer) and
(2) 'temporal generalization' (TG) maps. TG is used to
characterize the dynamics of neural representations and
237

238 consists in assessing the extent to which a temporal de-
239 coder trained at a given time sample generalizes to other
240 time samples [27] (Fig.3D).

Our spatial and temporal decoding results can be
accounted for by a feedforward architecture that both
(i) generates new representations at each layer and
(ii) propagates low-level representations across layers
(Fig.3 Model 1: 'broadcast'). This architecture predicts
that representations would not be maintained within
brain areas. This lack of maintenance is not sup-
ported by our data. First, the position of the stimu-
lus was decodable in the early visual cortex between
80-320 ms ($\bar{t}=5.18$, $p<.001$) and thus longer than the
stimulus presentation. Second, most temporal de-
coders significantly generalized over several hundreds
of milliseconds (Fig.4A-B). For example, the tempo-
ral decoder trained to predict stimulus position from
 $t=100$ ms could accurately generalize until ≈ 500 ms as
assessed with spatio-temporal cluster tests across sub-
jects (Fig.4A). Similarly, temporal decoders of percep-
tual category and button-press generalized, on aver-
age, for 287 ms (SEM=12.47; $p<.001$) and 689 ms
(SEM=30.94; $p<.001$) respectively. Given that the neu-
ral activity underlying the decoded representations is
partially stable over several hundreds of milliseconds,
recurrent connections seem necessary to account our
data (Fig.4 Model 2-4).

Consequently, we then considered a simple hierarchy
of recurrent layers, where recurrence only maintains ac-
tivated units (Fig.3 Model 2: 'maintain'). This archi-
tecture predicts strictly square TG matrices (i.e. tem-
poral decoders would be equivalent to one another in
terms of their performance) and is thus at odds with
the largely diagonal TG matrices observed empirically
(Fig.4A). Specifically, the duration of significant tempo-
ral decoding (fitting a new decoder at each time sample)
was significantly longer than the generalization of a sin-
gle decoder to subsequent time samples (e.g. 1,239 ver-
sus 287 ms for perceptual category ($t=-61.39$; $p<.001$)
and 1,215 versus 689 ms for button-press ($t=-16.26$;
 $p<.001$, Fig.4B). These results thus suggest that the de-
coded representations depend on dynamically-changing
activity: i.e. each feature is linearly coded by partially
distinct brain activity patterns at different time samples.

It is difficult to determine, with MEG alone, whether
such dynamic maintenance results from a change of
neural activity within or across brain areas. Indeed,
Model 1 and Model 3 can equally predict diagonal TG
(Fig.3). However, these two models, and their combi-
nation (Model 4) diverge in terms of *where* informa-
tion should be decodable. Specifically, source anal-
yses revealed that both stimulus position and percep-

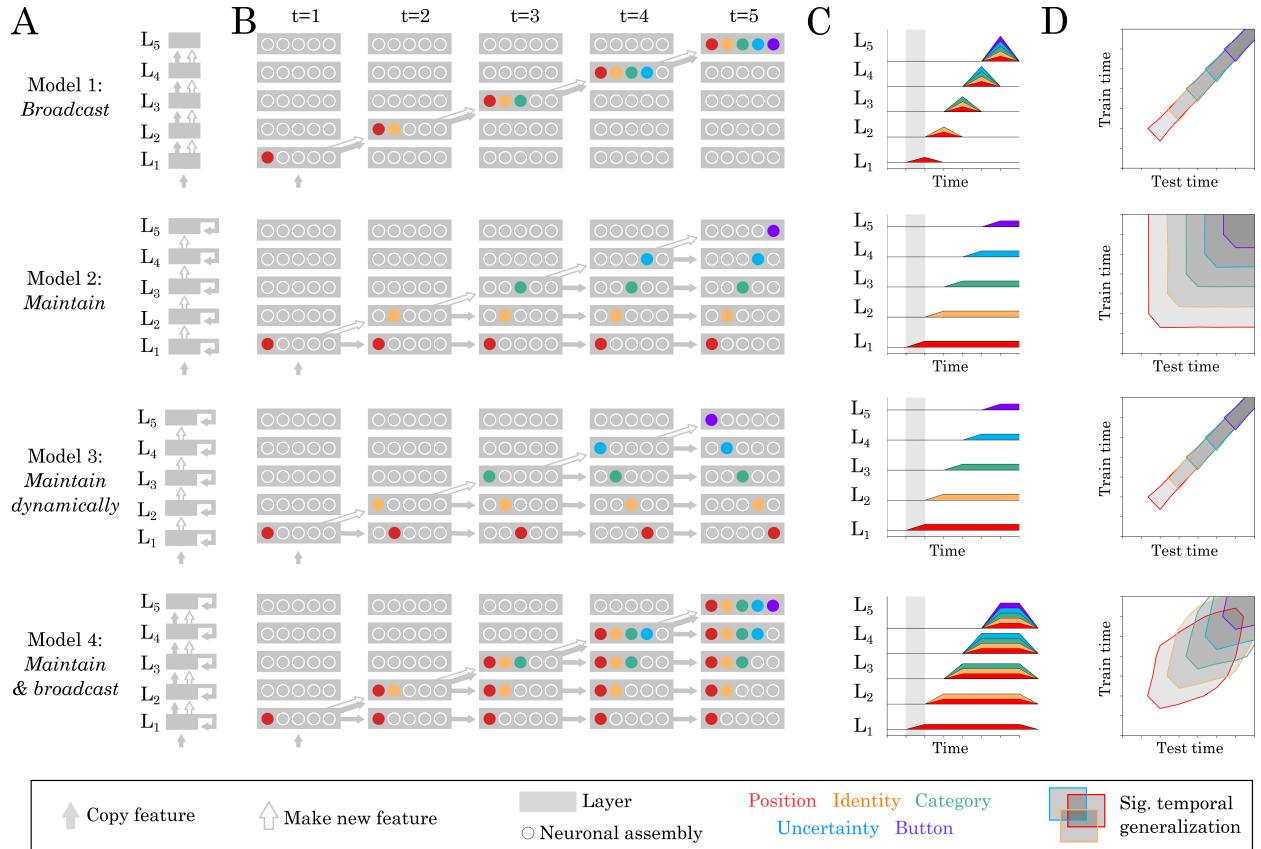


Figure 3: Source and temporal generalization predictions for various neural architectures. (A) Four increasingly complex neural architectures compatible with the spatial and temporal decoders of Fig.2. For each model (rows), the five layers ($L_1, L_2 \dots L_5$) generate new representations. The models 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.

tual category can be decoded across a wide variety of partially-overlapping brain areas (Fig.2B, Supplementary video 2), similarly to Model 4. Nonetheless, our MEG study remains limited in assessing whether within brain regions dynamics also contribute to the diagonal TG, which would suggest a mixture between models 3 and 4.

Together, source and TG analyses thus suggest that the slow and sequential generation of increasingly abstract representations depends on a hierarchy of recurrent layers that generate, maintain and broadcast representations across the cortex.

2.4. Hierarchical recurrence induces an accumulation of delays

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

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

To test this hypothesis, we estimated how the peak of each temporal decoder varied with reaction times. For clarity purposes, we split reaction times into four quantiles, and averaged the time courses of temporal decoders relative to their training time. These analyses showed that the latencies of (i) perceptual category ($r=0.35$; $p=0.006$), (ii) stimulus uncertainty ($r=0.37$; $p=0.004$) and (iii) button press ($r=0.66$; $p<0.001$) increasingly correlated with reaction times (Fig.5F-G).

Overall, these results show that we can track with

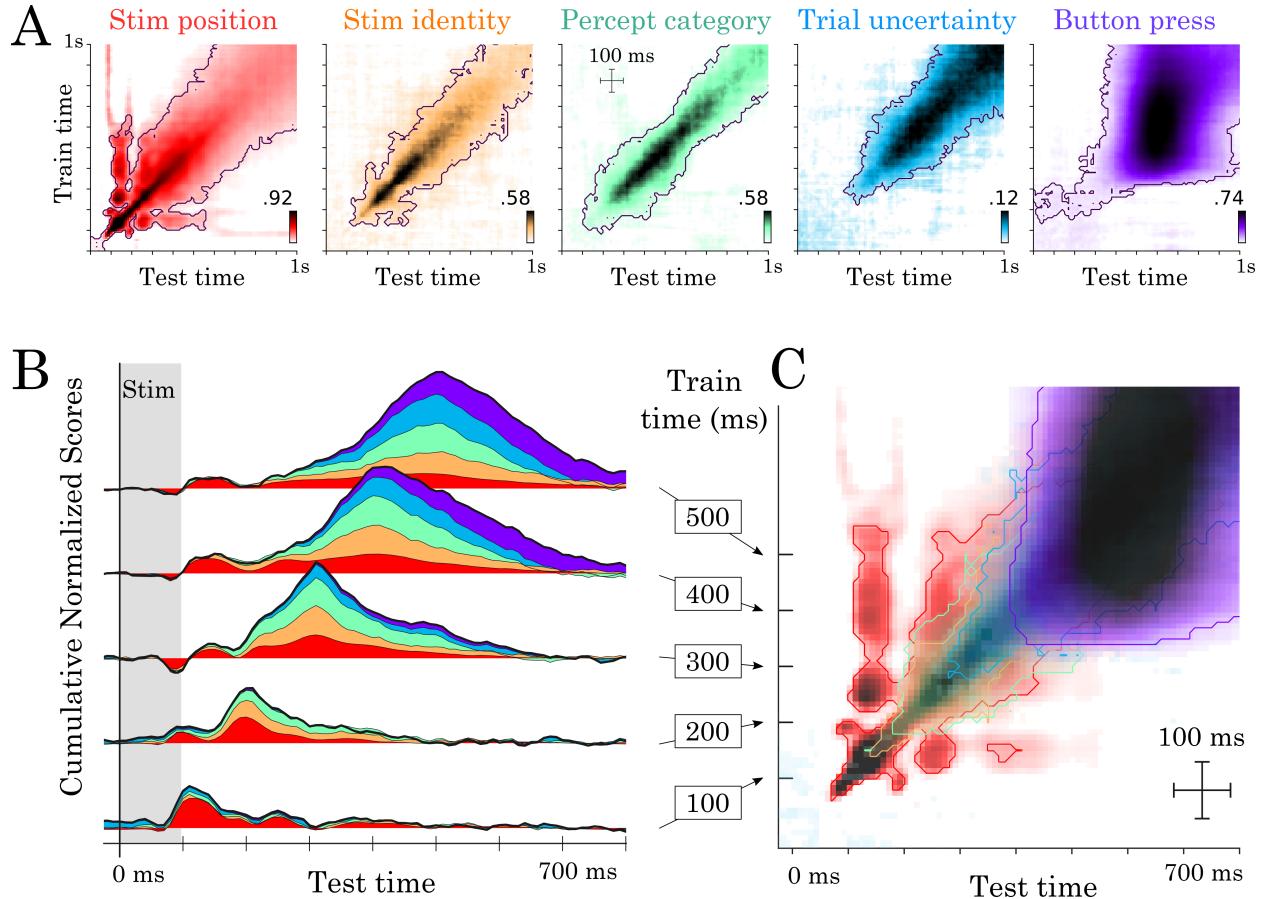


Figure 4: Temporal generalization results. (A) Temporal generalization for each of the five features orthogonally varying in our study. Color indicate decoding score (white=chance). Contours indicate significant decoding clusters across subjects. (B) Cumulative temporal generalization scores for the 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.

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

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

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

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

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

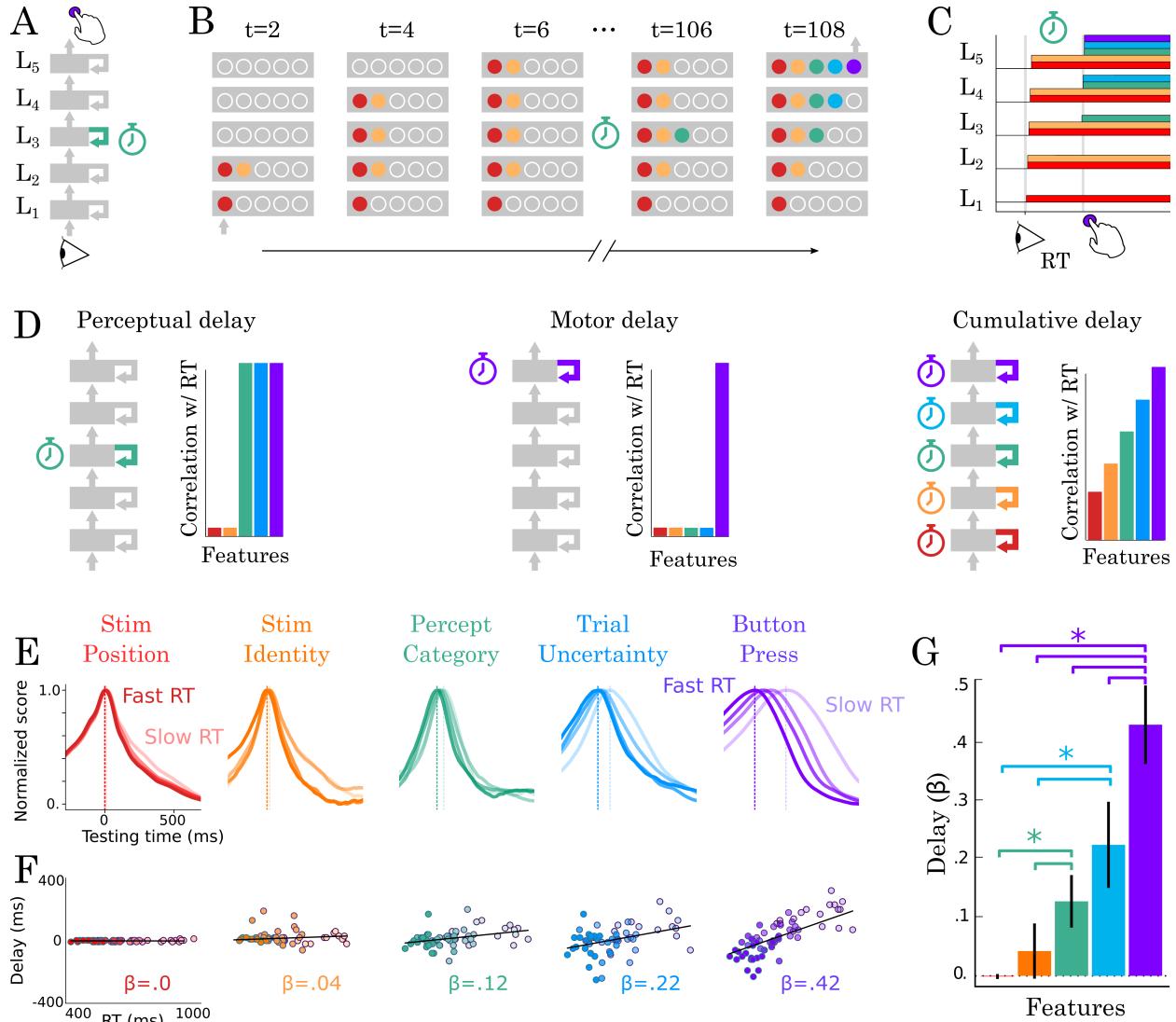


Figure 5: Correlation between TG peaks and reaction times. (A, B) Recurrent processing at a given processing stage is hypothesized to take a variable amount of time to generate adequate representations. (C) According to this hypothesis, the rise of the corresponding and subsequent representations would 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.

357 window latency was significant across subjects ($r=0.07$; 365
358 $SEM=0.01$; $p=0.002$).

359 This progressive categorization of the letter/digit rep- 366
360 resentations contrasts with the all-or-none pattern of 367
361 motor signals. Specifically, the probabilistic predictions 368
362 of button-press decoders varied categorically with re- 369
363 sponse evidence from 440 to 1,290 ms (sigmoid > linear 370
364 cluster, $\bar{t}=3.17$; $p<0.001$). There was also a more tran- 371

sient linear trend from 410 to 580 ms ($\bar{t}=3.69$; $p<0.001$). 365
366 This suggests that, unlike perceptual category, motor 367 signals largely derive from categorical inputs.

368 Together, delay (Fig.5) and categorization (Fig. 6) 369 analyses thus show that perceptual representations 370 slowly become categorical and are subsequently fol- 371 lowed by all-or-none motor representations.

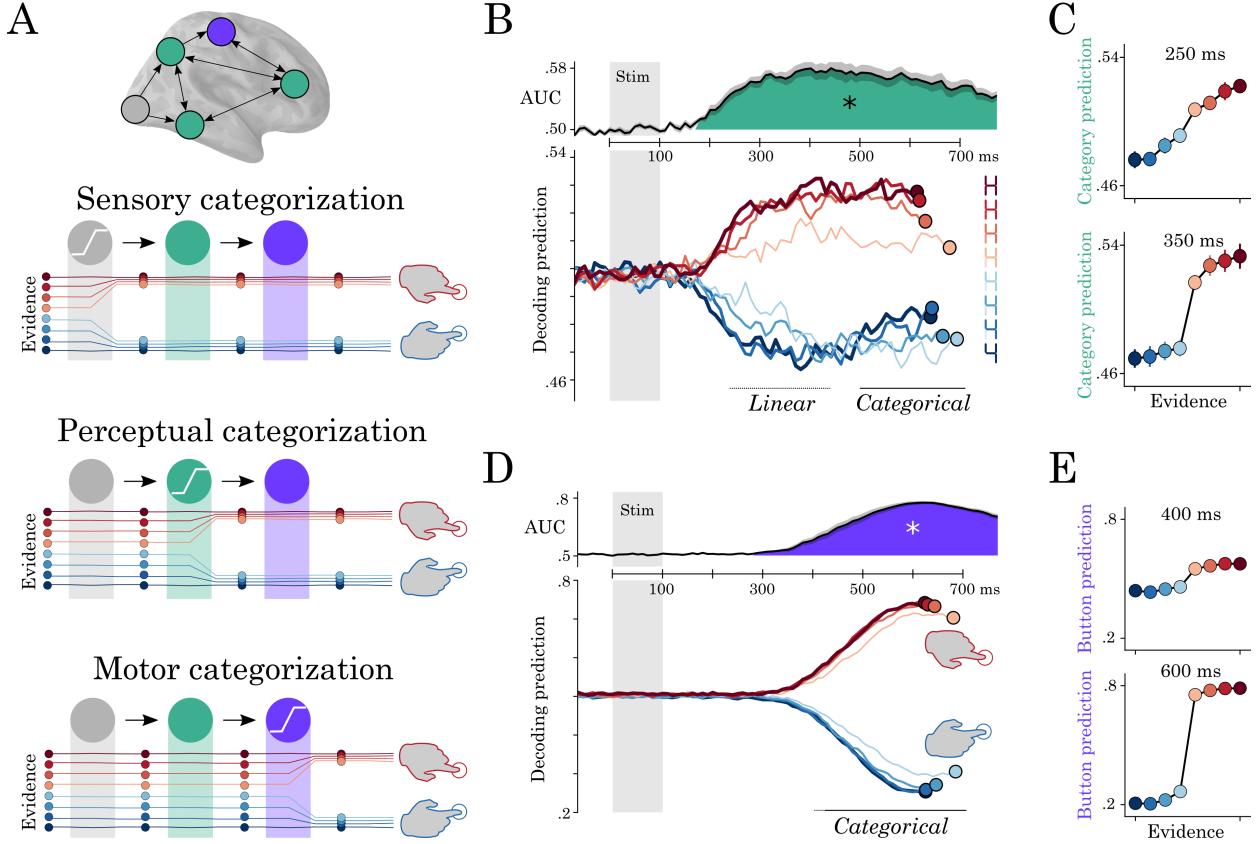


Figure 6: Motor and perceptual decisions. (A) Hypothesis space for when responses become categorical: during sensory, perceptual or motor processing. (B, top) Time course of decoding the perceptual decision. (B, bottom) Classifier predictions split into different levels of sensory evidence. (C) Averaging 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 uncertainty. (E) Different windows of classifier predictions, showing the categorical responses throughout processing.

3. Discussion

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

While macroscopic MEG signals advantageously give a birds-eye view of the cortical correlates of perceptual decision making, it should be stressed that their source reconstruction remains a coarse approximation. Consequently, identifying (1) the role of subcortical areas and (2) the extent to which representations dynamically change within each brain area will necessitate invasive brain recordings.

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

First, core-object recognition research, generally based on ≈ 100 ms-long image presentations has repeatedly shown that the spiking responses of the inferotemporal cortex is better explained by recurrent models than by feedforward ones [9, 15]. In particular, Kar et al have recently shown that images that are challenging to recognize, lead to delayed content-specific spiking activity in the macaque's infero-temporal cortex [15]. Similar evidence for recurrent processes was recently found using MEG [16]. Our findings, based on simpler but highly-controlled stimuli, are consistent with these results and further highlight that perceptual representations are not confined to the inferotemporal cortices, but also reach a large variety of parietal and prefrontal areas [28].

Second, the present study makes important contributions to the perceptual decision making literature [11, 13]. With some notable exceptions (e.g. [29]), this line of research primarily aims to isolate motor and

409 supra-modal decision signals in the presence of sus- 457
 410 tained visual inputs: i.e. neural responses ramping to- 458
 411 wards a virtual decision threshold, independently of the 459
 412 representation on which this decision is based [13]. The 460
 413 present study complements this approach by tracking 461
 414 the representation-specific signals that slowly emerge 462
 415 after a brief stimulus. Our results thus open an ex- 463
 416 citing avenue for querying the gating mechanisms of 464
 417 successive decisions and clarifying the role of the pre- 465
 418 frontal areas in the coordination multiple perceptual and 466
 419 supramodal modules [30]. 467

420 Finally, our results constitute an important confirmation 468
 421 of modern theories of perception. In particular, the 469
 422 Global Neuronal Workspace Theory predicts that per- 470
 423 ceptual representations need to be broadcast to asso- 471
 424 ciative cortices via the fronto-parietal areas to lead to 472
 425 subjective reports [8]. Yet, at some notable exceptions 473
 426 [31, 32], previous studies often fail to dissociate percep- 474
 427 tual contents and perceptual reports (e.g. [33, 34]). By 475
 428 contrast, the present experimental design allows an un- 476
 429 precedented dissection of the distinct processing stages 477
 430 that transform sensory input into perceptual represen- 478
 431 tations and, ultimately, actions. The generation of let- 479
 432 ter and digit representations in the dedicated brain areas 480
 433 [35, 20] and their subsequent broadcast to the cortex re- 481
 434 reinforce the notion that subjective perception relate to the 482
 435 global sharing of content-specific representations across 483
 436 brain areas [8, 36].

4. Method

4.1. Target stimuli

439 Using the font designed in [18], the stimuli were 484
 440 made from 0, 4, 5, 6, 8, 9, A, C, E, H, O, S, or from 485
 441 a linear combination of two of these characters varying 486
 442 in a single black bar (hereafter ‘pixel’). The correspond- 487
 443 ing ‘morphs’ were created by adjusting the contrast of 488
 444 the remaining pixel along eight equally spaced steps be- 489
 445 tween 0 (no bar) and 1 (black bar).

4.2. Experiment 1

446 Eight subjects with normal or corrected vision, seated 487
 447 ≈60 cm from a 19” CRT monitor (60Hz refresh rate, 488
 448 resolution: 1024x768), performed a stimulus identification 489
 449 task with continuous judgements across 28 variably 490
 450 ambiguous stimuli generated from digit stimuli. Ten euros 491
 451 were provided in compensation for this 1-hour experi- 492
 452 ment.

453 Subjects performed four blocks of 50 trials, each or- 493
 454 ganized in the following way. After a 200 ms fixation, a 494
 455 target stimulus, randomly selected from one of the 28 495

stimuli, was flashed for 83 ms on a 50% gray back- 457
 ground to the left or to the right of fixation. The ori- 458
 459 entation of the reporting disk (e.g 5-6-8-9 versus 5-9- 460
 461 8-6) was counterbalanced across subjects. Subjects had 462
 463 then up to 10 seconds to move a cursor on a large disk 464
 465 to report their percepts. The radius on the disk indi- 466
 467 cated subjective visibility (center=did not see the stim- 468
 469 ulus, disk border=max visibility). The angle on the disk 470
 471 indicated subjective identity (e.g. 5, 6, 8, 9 for the top 472
 473 left, top right, bottom right, and bottom left ‘corners’ 474
 475 respectively). Inter-trial interval was 500 ms. To ver- 476
 477 ify that subjects provided meaningful reports, the tar- 478
 479 get stimulus was absent 15% of the trials. Absent trials 480
 481 were rated with a low visibility (i.e radius below 5% 482
 483 of the disk radius) in most cases. Absent trials and trials 484
 485 reported with a low visibility were excluded from subse- 486
 487 quent analyses. The report distribution plotted in Fig.1B 488
 489 were generated with Seaborn’s bivariate Gaussian ker- 490
 491 nel density estimate function with default parameters.

Modeling categorical reports. To test whether subjective reports of stimulus identity varied linearly or categorically with sensory evidence, we analyzed how reports’ angle (i.e. subjective identity) varied with the expected 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:

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

and a sigmoidal model:

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

where \hat{y} is the report angle predicted by the model, x is expected angle given the stimulus pixels and β_0 is a 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 normalized to range between 0 and 1 within each subject. The β parameters were fit with Scipy’s ‘curve_fit’ function [37] to minimize a mean squared error across trials i :

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

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

491 trials. A Pearson correlation coefficient r summarised
492 the ability of each model to accurately predict y_{test} given
493 x_{train} , y_{train} and x_{test} . Finally, a Wilcoxon test was ap-
494 plied across subjects to test whether the two models
495 were consistently above chance ($r > 0$) and consistently
496 different from one another ($r_{sigmoid} > r_{linear}$).
497

498 *Experiment 2.* This experiment was performed at Neu-
499 rospin, Gif sur Yvette, thanks to the support of Stanis-
500 las Dehaene. Seventeen subjects performed a discrete
501 identification task across 22 variably ambiguous stimuli
502 generated from letters and digits inside an Elekta Neu-
503 romag MEG scanner (204 planar gradiometers and 102
504 magnetometers). Seventy euros were provided in com-
505 pensation to the 1-hour experiment and 30 minutes of
506 preparation.
507

508 Participants' head shape was digitized along with five
509 fiducial points on the forehead and on each aural canal.
510 Five head-position coils were placed on subjects head
511 and localized at the beginning of each block.
512

513 The trial structure was as follows. A black fixation
514 cross was displayed on a 50% gray background for 300
515 ms followed by a 100ms-long target stimulus presented
516 on the left or on the right of fixation. Two task-irrelevant
517 flankers (e.g. stimulus can be read as an S or a 5) were
518 displayed on the side of this target stimulus to increase
519 our chances of eliciting recurrent processing via crowd-
520 ing [21]. Subjects were given two seconds to report the
521 identity of the stimulus. Reports of stimulus identity
522 were given by pressing a button with the left and right
523 index fingers respectively. The identity-button mapping
524 changed on every block to orthogonalize the neural cor-
525 relates of stimulus identity and the neural correlates of
526 motor actions. For example, in block 1, perceiving an
527 E or a 4 should have been reported with a left button
528 press, whereas in block 2, E and 4 should have been
529 reported with a right button press. The identity-button
530 was explicitly reminded before each block. In addition,
531 a visual feedback was displayed after non-ambiguous
532 trials. Specifically, the fixation turned green for 100ms
533 or red for 300 ms in correct and incorrect trials respec-
534 tively. The brain responses to these feedback stimula-
535 tions are not analyzed in the present study. Inter-trial
536 interval was 1 second. Subjects were provided a short
537 training to ensure they understood the task, and identi-
538 fied non-ambiguous targets at least 80% of the time.
539

540 A total of 1920 trials, grouped into 40 blocks, were
541 performed by each subject, 320 of which were presented
542 passively at the end of each block – subjects were not
543 required to provide a response. The trial structure was
544 generated by (i) permuting all combinations of stimu-
545 lus features (e.g. position, identity, response mapping,
546

547 uncertainty), and (ii) shuffling the order of presentation
548 for each subject. The experiment was presented using
549 Psychtoolbox [38].
550

551 All experiments were approved by the local ethics
552 committee. All subjects signed an informed consent
553 form.
554

555 4.3. Structural MRI

556 For each subject, an anatomical MRI with a resolu-
557 tion of $1 \times 1 \times 1.1$ mm was acquired after the MEG ex-
558 periment with a 3T Siemens scanner. Gray and white mat-
559 ter were segmented with Freesurfer 'recon-all' pipeline
560 [23] and coregistered with each subject's digitized head
561 shapes along with fiducial points.
562

563 4.4. Preprocessing

564 The continuous MEG recording was noise-reduced
565 using Maxfilter's SSS correction on the raw data,
566 bandpass-filtered between 0.5 and 40 Hz using MNE-
567 Python's default parameters with firwin design [24] and
568 downsampled to 250 Hz. Epochs were then segmented
569 between -300 ms and +1500 ms relative to stimulus on-
570 sets.
571

572 After coregistering the MEG sensor data with sub-
573 jects' structural MRI and the head position coils, we
574 computed the forward model using a 1-layer (inner skull)
575 boundary element model, for each subject sepa-
576 rately and fit a minimum-norm inverse model (signal
577 to noise ratio: 3, loose dipole fitting: 0.2, with normal
578 orientation of the dipole relative to the cortical sheet)
579 using the noise covariance across sensors averaged over
580 the pre-stimulus baseline across trials. Finally, the in-
581 verse model was applied to single-trial data resulting in
582 a dynamic Statistical Parameter Map (dSPM) [22] value
583 for each source at each time sample.
584

585 4.5. Modeled features

586 We investigated whether single-trial source and sen-
587 sor evoked responses varied as a function of five fea-
588 tures: (1) the position of the stimulus on the computer
589 screen (left versus right of fixation), (2) the morph from
590 which the stimulus is generated (E-6 versus H-4), (3)
591 the category of the stimulus (letter versus digit), (4) the
592 uncertainty of the trial (maximum uncertainty = stimuli
593 with pixel at 50% contrast; minimally uncertain stim-
594 ili with pixels at 0% or 100% contrast), and (5) the re-
595 sponse button used to report the stimulus (left versus
596 right button). By design, these five features are inde-
597 pendent of one another.
598

599 It is challenging to dissociate brain responses that
600 represent objective sensory information from those that
601

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

604 4.6. Mass univariate statistics

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

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

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

611 4.7. Decoding

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

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

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

Regressors consisted of ridge regression (with scikit-
 learn [39]’s default parameters: $\alpha = 1$).

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

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

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

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

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

Temporal generalization (TG) consists in testing
 whether a temporal decoder fit on a training set at time
 t can decode a testing set at time t' [27]. TG can be
 summarized with a square training time \times testing time
 decoding matrix. To quantify the stability of neural rep-
 resentations, we measured the duration of above-chance
 generalization of each temporal decoder. To quanti-
 fy the dynamics of neural representations, we com-
 pared the mean duration of above-chance generaliza-
 tion across temporal decoders to the duration of above-
 chance temporal decoding (i.e. the diagonal of the ma-
 trix versus its rows). These two metrics were assessed
 within each subject and tested with second-level statis-
 tics across subjects.

648 4.8. Linear versus Categorical

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

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

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

752 4.9. Statistics

Univariate, decoding and TG models were fit within
 subjects, and tested across subjects. In case of repeated
 estimates (e.g. temporal decoding is repeated at each
 time sample), statistics derived from non-parametric
 cluster-testing with 10,000 permutations across subjects
 with MNE-Python’s default parameters [24].

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

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

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

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

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

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

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

689 5. Acknowledgement

690 This project received funding from the European
691 Union's Horizon 2020 research and innovation program
692 under grant agreement No 660086, the Bettencourt-
693 Schueller Foundation, the Fondation Roger de Spoel-
694 berch, the Philippe Foundation and the Abu Dhabi Insti-
695 tute G1001. We are infinitely grateful to Stanislas De-
696 haene, as well as David Poeppel and Alec Marantz, for
697 their support.

- 698 [1] Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The 699 Journal of physiology* **160**, 106–154 (1962).
 700
 701 [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**, 702 2563–2586 (1983).
 703
 704 [3] Riesenhuber, M. & Poggio, T. Hierarchical models of object 705 recognition in cortex. *Nature neuroscience* **2**, 1019 (1999).
 706
 707 [4] DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain 708 solve visual object recognition? *Neuron* **73**, 415–434 (2012).
 709
 710 [5] Yamins, D. L. *et al.* Performance-optimized hierarchical models 711 predict neural responses in higher visual cortex. *Proceedings of 712 the National Academy of Sciences* **111**, 8619–8624 (2014).
 713
 714 [6] Schrimpf, M. *et al.* Using brain-score to evaluate and build 715 neural networks for brain-like object recognition. In *Cosyne 716 19, Date: 2019/02/28-2019/03/03, Location: Lisbon, Portugal* (2019).
 717
 718 [7] Ratcliff, R. & Smith, P. L. A comparison of sequential sampling 719 models for two-choice reaction time. *Psychological review* **111**, 720 333 (2004).
 721
 722 [8] Dehaene, S. & Changeux, J.-P. Experimental and theoretical 723 approaches to conscious processing. *Neuron* **70**, 200–227 (2011).
 724
 725 [9] Lamme, V. A. & Roelfsema, P. R. The distinct modes of vision 726 offered by feedforward and recurrent processing. *Trends in 727 neurosciences* **23**, 571–579 (2000).
 728
 729 [10] Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory 730 responses in cat visual cortex exhibit inter-columnar synchronization 731 which reflects global stimulus properties. *Nature* **338**, 732 334 (1989).
 733
 734 [11] Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annual review of neuroscience* **30** (2007).
 735
 736 [12] Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual 737 decision in the parietal cortex (area lip) of the rhesus monkey. *Journal 738 of neurophysiology* **86**, 1916–1936 (2001).
 739
 740 [13] O'connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal 741 accumulation-to-bound signal that determines perceptual decisions 742 in humans. *Nature neuroscience* **15**, 1729 (2012).
 743
 744 [14] Spoerer, C. J., McClure, P. & Kriegeskorte, N. Recurrent convolutional neural networks: a better model of biological object 745 recognition. *Frontiers in psychology* **8**, 1551 (2017).
 746
 747 [15] 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 748 neuroscience* **22**, 974 (2019).
 749
 750 [16] Kietzmann, T. C. *et al.* Recurrence is required to capture the 751 representational dynamics of the human visual system. *Proceedings of 752 the National Academy of Sciences* **116**, 21854–21863 (2019).
 753
 754 [17] Spoerer, C. J., Kietzmann, T. C. & Kriegeskorte, N. Recurrent 755 networks can recycle neural resources to flexibly trade speed for 756 accuracy in visual recognition. *bioRxiv* 677237 (2019).
 757
 758 [18] King, J.-R. & Dehaene, S. A model of subjective report and 759 objective discrimination as categorical decisions in a vast 760 representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130204 (2014).
 761
 762 [19] Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends in cognitive sciences* **15**, 254–262 (2011).
 763
 764 [20] Shum, J. *et al.* A brain area for visual numerals. *Journal of 765 Neuroscience* **33**, 6709–6715 (2013).
 766
 767 [21] Strasburger, H., Rentschler, I. & Jüttner, M. Peripheral vision 768 and pattern recognition: A review. *Journal of vision* **11**, 13–13 (2011).
 769
 770 [22] Dale, A. M. *et al.* Dynamic statistical parametric mapping: com- 771 bining fmri and meg for high-resolution imaging of cortical activity. *Neuron* **26**, 55–67 (2000).
 772
 773 [23] Fischl, B. Freesurfer. *Neuroimage* **62**, 774–781 (2012).
 774
 775 [24] Gramfort, A. *et al.* Mne software for processing meg and eeg 776 data. *Neuroimage* **86**, 446–460 (2014).
 777
 778 [25] Hagler Jr, D. J. & Sereno, M. I. Spatial maps in frontal and 779 prefrontal cortex. *Neuroimage* **29**, 567–577 (2006).
 780
 781 [26] Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field 782 maps in human cortex. *Neuron* **56**, 366–383 (2007).
 783
 784 [27] King, J. & Dehaene, S. Characterizing the dynamics of mental 785 representations: the temporal generalization method. *Trends in 786 cognitive sciences* **18**, 203–210 (2014).
 787
 788 [28] Freedman, D. J. & Miller, E. K. Neural mechanisms of visual 789 categorization: insights from neurophysiology. *Neuroscience & 790 Biobehavioral Reviews* **32**, 311–329 (2008).
 791
 792 [29] Philastrides, M. G. & Sajda, P. Eeg-informed fmri reveals spatiotemporal characteristics of perceptual decision making. *Journal 793 of Neuroscience* **27**, 13082–13091 (2007).
 794
 795 [30] Sarafyazd, M. & Jazayeri, M. Hierarchical reasoning by neural 796 circuits in the frontal cortex. *Science* **364**, eaav8911 (2019).
 797
 798 [31] Tong, F., Nakayama, K., Vaughan, J. T. & Kanwisher, N. Binocular 799 rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**, 753–759 (1998).
 800
 801 [32] King, J.-R., Pescetelli, N. & Dehaene, S. Brain mechanisms 802 underlying the brief maintenance of seen and unseen sensory 803 information. *Neuron* **92**, 1122–1134 (2016).
 804
 805 [33] Sergent, C., Baillet, S. & Dehaene, S. Timing of the brain events 806 underlying access to consciousness during the attentional blink. *Nature 807 neuroscience* **8**, 1391 (2005).
 808
 809 [34] Van Vugt, B. *et al.* The threshold for conscious report: Signal 810 loss and response bias in visual and frontal cortex. *Science* **360**, 811 537–542 (2018).
 812
 813 [35] Cohen, L. *et al.* The visual word form area: spatial and temporal 814 characterization of an initial stage of reading in normal subjects and 815 posterior split-brain patients. *Brain* **123**, 291–307 (2000).
 816
 817 [36] Lamme, V. A. Why visual attention and awareness are different. *Trends in cognitive sciences* **7**, 12–18 (2003).
 818
 819 [37] Jones, E., Oliphant, T., Peterson, P. *et al.* SciPy: Open source 820 scientific tools for Python (2001–). URL <http://www.scipy.org/>. [Online; accessed 1 today].
 821
 822 [38] Kleiner, M. *et al.* What's new in psychtoolbox-3. *Perception* **36**, 823 1 (2007).
 824
 825 [39] Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *Journal of machine learning research* **12**, 2825–2830 (2011).