Characterizing the dynamics of mental representations: the temporal generalization method

J-R. King[1,2,3] and **S. Dehaene**[1,2,4,5]

[1]Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale, U992, F-91191 Gif/Yvette, France

[2]NeuroSpin Center, Institute of BioImaging Commissariat à l'Energie Atomique, F-91191 Gif/Yvette, France

[3]Institut du Cerveau et de la Moelle Épinière Research Center, Institut National de la Santé et de la Recherche Médicale, U975 Paris, France

[4]Université Paris 11, Orsay, France

[5]Collège de France, F-75005 Paris, France

Abstract

Parsing a cognitive task into a sequence of operations is a central problem in cognitive neuroscience. We argue that a major advance is now possible owing to the application of pattern classifiers to time-resolved recordings of brain activity [electroencephalography (EEG), magnetoencephalography (MEG), or intracranial recordings]. By testing at which moment a specific mental content becomes decodable in brain activity, we can characterize the time course of cognitive codes. Most importantly, the manner in which the trained classifiers generalize across time, and from one experimental condition to another, sheds light on the temporal organization of information-processing stages. A repertoire of canonical dynamical patterns is observed across various experiments and brain regions. This method thus provides a novel way to understand how mental representations are manipulated and transformed.

Keywords

decoding; EEG; MEG; temporal generalization; serial processing; parallel processing; multivariate pattern analyses

Understanding the organization of processing stages: from behavior to neuroimaging

Understanding how mental representations unfold in time during the performance of a task is a central goal for cognitive psychology. Donders [1] first suggested that mental operations could be dissected by comparing the subjects' response times in different experimental conditions. This 'mental chronometry' was later enriched with several methodological

Corresponding authors: King, J. (jeanremi.king@gmail.com); Dehaene, S. (stanislas.dehaene@cea.fr).

inventions, including the additive-factors method [2] and the psychological refractory period method [3]. Although these behavioral techniques can provide considerable information on the temporal organization of cognitive computations, they remain fraught with ambiguities. For instance, they cannot fully separate serial from parallel processes [4] or processes organized into a discrete series of steps from those operating as a continuous flow or 'cascade' of overlapping stages [5].

More recently, the advent of brain-imaging techniques has provided unprecedented access into the content and the dynamics of mental representations. EEG, MEG, local field potentials, and neuronal recordings can provide a fine-grained dissection of the sequence of brain activations (e.g., see [6,7]). Here we show how the analysis of these time-resolved signals can be enhanced through the use of multivariate pattern analysis (MVPA), also known informally as 'decoding'. We argue that temporal decoding methods offer a vast and largely untapped potential for the determination of how mental representations unfold over time.

Decoding fMRI data identifies the localization and structure of mental representations

MVPA was first introduced to brain imaging in order to refine the analysis of functional MRI (fMRI). Temporal resolution aside, fMRI is an efficient tool to isolate and localize the brain mechanisms underlying specific mental representations. Initially, fMRI was primarily used with binary contrasts that revealed major differences in regional brain activity (e.g., faces versus non-faces in the fusiform cortex [8]). MVPA, however, led to a considerable refinement of such inferences because it proved able to resolve, inside an area, the fine-grained patterns of brain activity that contain detailed information about the stimulus or the ongoing behavior (for reviews, see [9–11]).

Using MVPA, subtle details of mental representations can be decoded from fMRI activity patterns. It is now possible to decode low-level visual features such as orientation [12–15] or color [16] and determine how their cortical representation is changed, tuned, or suppressed according to subjects' goals [13] and prior knowledge [14]. Using distributed activity over the ventral visual pathway, it is possible to reconstruct which static images [17–21], moving objects [22], or movies [23] that the subject is watching. The identification of these mental contents is not limited to sensory information; in the absence of any stimulus, mental objects can still be decoded, for instance during mental imagery [24,25], working memory [25,26], or even dreams [27]. More recently, MPVA has also been used to decode non-visual representations, including auditory [28,29], mnemonic [30], numerical [31], and executive information [32,33]. Although this approach has provided unprecedented details about the structure and brain correlates of various mental representations [11], its temporal resolution remains largely inadequate to describe the way they are dynamically constructed, transformed, and manipulated (although some slow processes can be tracked with fMRI; e.g., [12,13,25]).

Decoding time-resolved signals identifies when mental representations are activated

MVPA applied to fMRI signals does not reveal much about the dynamics with which mental representations are activated. Here we focus specifically on the less explored question of what MVPA may bring to our understanding of the dynamics of information processing in the brain.

Methodologically, MVPA readily applies to EEG, MEG, or intracranial recording data (e.g., multiunit neuronal recordings, local field potentials) where time can be considered as an additional dimension besides the spatial information provided by the multiple sensors. To obtain time-course information, a series of classifiers can be trained, each using as an input a specific time slice of the original data (Box 1). The output is a curve representing decoding performance as a function of time that specifies when a certain piece of information becomes explicitly encoded in brain activity. Multiple such classifiers can be trained to decode distinct features of the trial, thus dissecting a trial into a series of overlapping stages. Recently, for instance, in a simple response-time task, we decoded successively the location of the stimulus, the subject's required response, his actual motor response (which differed on error trials), and whether the brain detected whether the response was correct or erroneous [34].

Even scalp EEG recordings contain a lot of information sufficient to discriminate various processing stages. Low-level attributes such as the position [35,36] and the size [36] of a visual stimulus, as well as the orientation of large visual gratings [37–39], can be reconstructed from EEG and MEG signals. Higher processing stages of perceptual decision making, related to either the objective stimulus or its subjective report, can be similarly tracked [35,40–43]. EEG even permits decoding the covert production of imagined syllables [44]. Motor [45], auditory [46], conceptual and semantic information [47–49], and even music samples [50] have also been decoded from EEG and MEG recordings.

These studies typically decode information from raw sensor data in the time domain. The expansion of brain signals through a time–frequency transform may reveal additional information spread over both frequency and time [44,51,52].

Contrary to what is generally assumed, these studies suggest that the spatial resolution of EEG and MEG may suffice to decode and track various different mental representations. Furthermore, two approaches can be used to provide a more accurate understanding of the spatial origin of the decoded signals. First, source reconstruction analyses may provide an approximate location of the neural generators underlying scalp recordings [53]. Following source reconstruction, decoding may then be restricted to specific regions of interest, thus clarifying which brain regions contain decodable information. Second, intracranial recordings can now be acquired from dozens or even hundreds of different cortical and subcortical sites. When applied to intracranial data, decoding can reveal a rich temporal dynamics of fine-grained codes. For example, in the olfactory bulb of the zebra fish, decoding identified a sequence of neuronal responses to odors, revealing that ambiguous mixtures of odors are initially coded continuously and become categorically represented

later in time [54]. In humans, decoding has been used to show that perceived phonemes are categorically represented in the posterior superior temporal gyrus, with a peak as early as 110 ms following sound onset [55]. Similarly, the production of spoken syllables can be deciphered from the sensorimotor somatotopic cortex, with sequential activation of distinct codes for the initial consonant and the subsequent vowel [56].

Overall, applying decoding techniques to successive time points can identify when, and for how long, a specific piece of information is represented in the brain. The next question we may ask is: how does the underlying neural code evolve in time?

Generalization across time reveals how mental representations are dynamically transformed

The decoding approach detailed above can be extended to ask whether the neural code that supports above-chance decoding is stable or is dynamically evolving (see Figure I in Box 1). The principle is simple: instead of applying a different classifier at each time point, the classifier trained at time t can be tested on its ability to generalize to time t'. Generalization implies that the neural code that was identified at time t recurred at time t'.

Systematically adopting this approach leads to a temporal generalization matrix in which each row corresponds to the time at which the classifier was trained and each column to the time at which it was tested. For instance, when we applied this method to the decoding of novel versus habitual sounds [57], we found that mismatch signals, classically assigned to a single 'mismatch response', actually correspond to an extended sequence of distinct brain activation patterns (Figure 1). Although sound novelty can be decoded over a long time window, each of the classifiers is time specific and does not generalize over a long time period. All of this information is apparent in the temporal generalization matrix, which takes a diagonal form.

The simulations presented in Figure 2 exemplify how the temporal generalization method may distinguish fundamentally different dynamics of brain activity to which traditional sliding-window classifiers can be blind. So far, this method has been used only sporadically in the literature. The few available examples suggest that the diagonal pattern, indicative of a series of processing stages, is only one of several canonical dynamics modes of brain activity (Figure 3). For instance, when we analyze a slightly different response to auditory novelty – the late brain response to frequent versus rare auditory melodies [57] – we find that it obeys a strikingly different dynamics from the mismatch response, with a square generalization matrix indicating a stable neural code (compare Figure 3A and B). A similar square matrix was observed for categories of visual pictures in the inferotemporal cortex during a simple attention task [58]. Interestingly, other experiments [59,60] find that the same area, as well as the prefrontal cortex, may switch to a diagonal pattern during a delayed match-to-sample task (Figure 3F–I). These examples suggest that, in different contexts, either a stable activity pattern or a time-changing code is used by the brain to bridge across a long temporal delay.

Other types of dynamics undoubtedly exist. For instance, the neural code may oscillate as a function of time. Fuentemilla and collaborators [52] trained a multivariate pattern classifier during the presentation of indoor or outdoor stimuli. When testing during the delay period of a working-memory task, they obtained above-chance performance that, interestingly, oscillated at the theta rhythm (4–8 Hz), indicating that a working-memory code recurred cyclically.

Other temporal generalization data suggest that neural codes may reverse over time. For example, Carlson *et al.* trained a classifier to decode the position of a visual stimulus [35]. They showed that, after training a classifier at the onset of the image, the classifier led to below-chance predictions at the time of stimulus offset, suggesting that the neural activity pattern recurred with a reversed polarity. Similar off-diagonal below-chance generalization has been observed by others [35,43,57,61]. Understanding when and why such reversals occur is an interesting question for future research.

Generalization across conditions reveals how information processing is changed

We just showed that the temporal generalization matrix provides detailed information about the sequence of processing stages engaged in a particular task. If we change the experimental conditions, however, some of these processing stages may remain unaffected whereas others may be accelerated, slowed, deleted, or inserted. Can decoding also illuminate such reorganizations? If a classifier is trained in one condition and tested on its ability to generalize to another, the resulting temporal generalization matrix may indicate how information processing has changed (see simulations in Figure 4). For instance, an acceleration of some stages would be perceptible as a displacement of the generalization window outside the diagonal. The method can identify at what time, and for how long, the acceleration occurred. Other changes, such as the deletion or insertion of processing stages or their temporal reordering, can be similarly characterized (Figure 4).

Although the proposed method has rarely been used empirically, it offers great potential for revealing the changing dynamics of processing stages. Indeed, neurophysiological studies indicate that complex temporal reorganization of brain activity may occur. A striking example is provided by the decoding of an animal's spatial location from multiunit neuronal recordings. Capitalizing on the earlier demonstration that rat location could be accurately decoded from hippocampal place-cell activity [62], Louie and Wilson [63] identified a sequence of place-cell firing that reproducibly tagged the animal's movement through a circular track. They then showed that the same sequence was replayed during rapid eye-movement (REM) sleep, at approximately the same speed or slower (also see [64]). Subsequently, Lee and Wilson [65] showed that place-cell activity, which could be used to detect the location and direction of animal movement, was also replayed at 20 times-faster speed during slow-wave sleep, while the animal was, of course, immobile. Accelerated replay was also observed in the awake state [66]. Recently, Pfeiffer and Foster [67] found that fast replay of place-cell activity also occurred when the animal was searching for a goal. When decoded, the neuronal activity unfolded into a goal-directed trajectory that

anticipated, at an accelerated pace, the subsequent movement of the rat on the same trial. Place-cell firing may also recur in reverse order [68], perhaps reflecting a form of goal-based problem solved by means–ends analysis. Those examples prove that temporal accelerations and reversals do occur in brain activity and invite a search for their occurrence in other cognitive domains (e.g., bird song [69]) using the temporal generalization matrix as a telling signature.

Concluding remarks

Since Donders and Sternberg, brain algorithms have been dissected by manipulating experimental factors such as attention, expectation, or instructions that selectively accelerate, slow, remove, insert, or reorder specific processing stages. Behavioral methods of mental chronometry, however, provide only indirect information about such reorganizations. Here we summarized several ideas and empirical studies that suggest that MVPA can provide considerable information about the fine temporal organization of information-processing stages. In the past decade, MVPA has been primarily used to analyze brain-imaging data at a fine spatial scale and to characterize the structure of mental representations [11]. Here we suggest that applying MPVA across space, time, and experimental conditions can reveal not only the time at which information becomes decodable, but also the dynamics with which the underlying representations are processed. Characterizing the building blocks of cognitive processes in time, rather than solely in space, presents many challenges and open issues (Boxes 2 and 3) for which we hope that the temporal generalization method may prove useful.

Acknowledgments

References

1. Donders FC. On the speed of mental processes. Acta Psychol (Amst). 1969; 30:412–431. [PubMed: 5811531]

2. Sternberg S. The discovery of processing stages: extensions of Donders' method. Acta Psychol (Amst). 1969; 30:276–315.

3. Pashler H. Dual-task interference in simple tasks: data and theory. Psychol Bull. 1994; 116:220–244. [PubMed: 7972591]

4. Townsend JT. Serial vs parallel processing: sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. Psychol Sci. 1990; 1:46–54.

5. McClelland JL. On the time-relations of mental processes: an examination of systems of processes in cascade. Psychol Rev. 1979; 86:287–330.

6. Dehaene S. The organization of brain activations in number comparison: event-related potentials and the additive-factors methods. J Cogn Neurosci. 1996; 8:47–68. [PubMed: 23972235]

7. Nishitani N, Hari R. Viewing lip forms: cortical dynamics. Neuron. 2002; 36:1211–1220. [PubMed: 12495633]

8. Kanwisher N, et al. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 1997; 17:4302. [PubMed: 9151747]

9. Quian Quiroga R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci. 2009; 10:173–185. [PubMed: 19229240]

10. Haynes J-DD, Rees G. Decoding mental states from brain activity in humans. Nat Rev Neurosci. 2006; 7:523–534. [PubMed: 16791142]

11. Kriegeskorte N, Kievit RA. Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci. 2013; 17:401–412. [PubMed: 23876494]

12. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 2005; 8:679–685. [PubMed: 15852014]

13. Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009; 458:632–635. [PubMed: 19225460]

14. Kok P, et al. Less is more: expectation sharpens representations in the primary visual cortex. Neuron. 2012; 75:265–270. [PubMed: 22841311]

15. Freeman J, et al. Orientation decoding depends on maps, not columns. J Neurosci. 2011; 31:4792–4804. [PubMed: 21451017]

16. Brouwer GJ, Heeger DJ. Decoding and reconstructing color from responses in human visual cortex. J Neurosci. 2009; 29:13992–14003. [PubMed: 19890009]

17. Gallant, J., et al. Bayesian reconstruction of perceptual experiences from human brain activity. Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. Springer; 2009. p. 393

18. Kay KN, Gallant JL. I can see what you see. Nat Neurosci. 2009; 12:245. [PubMed: 19238184]

19. Miyawaki Y, et al. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron. 2008; 60:915–929. [PubMed: 19081384]

20. Thirion B, et al. Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage. 2006; 33:1104–1116. [PubMed: 17029988]

21. Gramfort, A., et al. 2012 Second International Workshop on Pattern Recognition in NeuroImaging; IEEE Computer Society; 2012. p. 13-16.

22. Van Gerven MAJ, et al. Dynamic decoding of ongoing perception. Neuroimage. 2011; 57:950–957. [PubMed: 21609771]

23. Nishimoto S, et al. Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol. 2011; 21:1641–1646. [PubMed: 21945275]

24. Lee S-H, et al. Disentangling visual imagery and perception of real-world objects. Neuroimage. 2012; 59:4064–4073. [PubMed: 22040738]

25. Albers AM, et al. Shared representations for working memory and mental imagery in early visual cortex. Curr Biol. 2013; 23:1427–1431. [PubMed: 23871239]

26. Christophel TB, et al. Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci. 2012; 32:12983–12989. [PubMed: 22993415]

27. Horikawa T, et al. Neural decoding of visual imagery during sleep. Science. 2013; 340:639–642. [PubMed: 23558170]

28. Formisano E, et al. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science. 2008; 322:970–973. [PubMed: 18988858]

29. Giordano BL, et al. Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex. 2013; 23:2025–2037. [PubMed: 22802575]

30. Rissman J, et al. Detecting individual memories through the neural decoding of memory states and past experience. Proc Natl Acad Sci USA. 2010; 107:9849–9854. [PubMed: 20457911]

31. Knops A, et al. Recruitment of an area involved in eye movements during mental arithmetic. Science. 2009; 324:1583–1585. [PubMed: 19423779]

32. Cole MW, et al. Multi-task connectivity reveals flexible hubs for adaptive task control. Nat Neurosci. 2013; 16:1348–1355. [PubMed: 23892552]

33. Zhang J, et al. Choosing the rules: distinct and overlapping frontoparietal representations of task rules for perceptual decisions. J Neurosci. 2013; 33:11852–11862. [PubMed: 23864675]

34. Charles L, et al. Decoding the dynamics of action, intention, and error-detection for conscious and subliminal stimuli. J Neurosci. 2014; 34:1158–1170. [PubMed: 24453309]

35. Carlson TA. High temporal resolution decoding of object position and category. J Vis. 2011; 11:1–17.

36. Isik L, et al. The dynamics of invariant object recognition in the human visual system. J Neurophysiol. 2013; doi: 10.1152/jn.00394.2013

37. Garcia JO, et al. Near-real-time feature-selective modulations in human cortex. Curr Biol. 2013; 23:515–522. [PubMed: 23477721]

38. Ramkumar P, et al. Feature-specific information processing precedes concerted activation in human visual cortex. J Neurosci. 2013; 33:7691–7699. [PubMed: 23637162]

39. Duncan KK, et al. Identifying spatially overlapping local cortical networks with MEG. Hum Brain Mapp. 2010; 31:1003–1016. [PubMed: 19998365]

40. Philiastides MG, et al. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J Neurosci. 2006; 26:8965–8975. [PubMed: 16943552]

41. Ratcliff R, et al. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci USA. 2009; 106:6539–6544. [PubMed: 19342495]

42. Philiastides MG, Sajda P. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex. 2006; 16:509–518. [PubMed: 16014865]

43. Carlson T, et al. Representational dynamics of object vision: the first 1000 ms. J Vis. 2013; 13:1–19.

44. Deng S, et al. EEG classification of imagined syllable rhythm using Hilbert spectrum methods. J Neural Eng. 2010; 7:046006. [PubMed: 20551510]

45. Waldert S, et al. Hand movement direction decoded from MEG and EEG. J Neurosci. 2008; 28:1000–1008. [PubMed: 18216207]

46. King J-R, et al. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. Neuroimage. 2013; 83C:726–738.

47. Chan AM, et al. Decoding word and category-specific spatiotemporal representations from MEG and EEG. Neuroimage. 2011; 54:3028–3039. [PubMed: 21040796]

48. Sudre G, et al. Tracking neural coding of perceptual and semantic features of concrete nouns. Neuroimage. 2012; 62:451–463. [PubMed: 22565201]

49. Guimaraes MP, et al. Single-trial classification of MEG recordings. IEEE Trans Biomed Eng. 2007; 54:436–443. [PubMed: 17355055]

50. Schaefer RS, et al. Name that tune: decoding music from the listening brain. Neuroimage. 2011; 56:843–849. [PubMed: 20541612]

51. Schyns PG, et al. Cracking the code of oscillatory activity. PLoS Biol. 2011; 9:e1001064. [PubMed: 21610856]

52. Fuentemilla L, et al. Theta-coupled periodic replay in working memory. Curr Biol. 2010; 20:606–612. [PubMed: 20303266]

53. Hämäläinen M, et al. Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys. 1993; 65:413–497.

54. Niessing J, Friedrich RW. Olfactory pattern classification by discrete neuronal network states. Nature. 2010; 465:47–52. [PubMed: 20393466]

55. Chang EF, et al. Categorical speech representation in human superior temporal gyrus. Nat Neurosci. 2010; 13:1428–1432. [PubMed: 20890293]

56. Bouchard KE, et al. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013; 495:327–332. [PubMed: 23426266]

57. King J-R, et al. Two distinct dynamic modes subtend the detection of unexpected sounds. PLoS ONE. 2014; doi: 10.1371/journal.pone.0085791

58. Zhang Y, et al. Object decoding with attention in inferior temporal cortex. Proc Natl Acad Sci USA. 2011; 108:8850–8855. [PubMed: 21555594]

59. Meyers EM, et al. Dynamic population coding of category information in inferior temporal and prefrontal cortex. J Neurophysiol. 2008; 100:1407–1419. [PubMed: 18562555]

60. Stokes MG, et al. Dynamic coding for cognitive control in prefrontal cortex. Neuron. 2013; 78:364–375. [PubMed: 23562541]

61. Nikoli D, et al. Distributed fading memory for stimulus properties in the primary visual cortex. PLoS Biol. 2009; 7:e1000260. [PubMed: 20027205]

62. Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. Science. 1994; 265:676–679. [PubMed: 8036517]

63. Louie K, Wilson MA. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. Neuron. 2001; 29:145–156. [PubMed: 11182087]

64. Poe GR, et al. Experience-dependent phase-reversal of hippocampal neuron firing during REM sleep. Brain Res. 2000; 855:176–180. [PubMed: 10650147]

65. Lee AK, Wilson MA. Memory of sequential experience in the hippocampus during slow wave sleep. Neuron. 2002; 36:1183–1194. [PubMed: 12495631]

66. Davidson TJ, et al. Hippocampal replay of extended experience. Neuron. 2009; 63:497–507. [PubMed: 19709631]

67. Pfeiffer BE, Foster DJ. Hippocampal place-cell sequences depict future paths to remembered goals. Nature. 2013; 497:74–79. [PubMed: 23594744]

68. Foster DJ, Wilson MA. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature. 2006; 440:680–683. [PubMed: 16474382]

69. Dave AS, Margoliash D. Song replay during sleep and computational rules for sensorimotor vocal learning. Science. 2000; 290:812–816. [PubMed: 11052946]

70. Shepard RN. Toward a universal law of generalization for psychological science. Science. 1987; 237:1317–1323. [PubMed: 3629243]

71. Haynes J-D, et al. Reading hidden intentions in the human brain. Curr Biol. 2007; 17:323–328. [PubMed: 17291759]

72. Kriegeskorte N, et al. Information-based functional brain mapping. Proc Natl Acad Sci USA. 2006; 103:3863–3868. [PubMed: 16537458]

73. Schurger A, et al. Reducing multi-sensor data to a single time course that reveals experimental effects. BMC Neurosci. 2013; 14:122. [PubMed: 24125590]

74. DiCarlo JJ, et al. How does the brain solve visual object recognition? Neuron. 2012; 73:415–434. [PubMed: 22325196]

75. Li N, DiCarlo JJ. Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. Neuron. 2010; 67:1062–1075. [PubMed: 20869601]

76. Chang C, Lin C. LIBSVM: a library for support vector machines. Computer. 2001; 2:1–30.

77. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 2004; 27:712–719. [PubMed: 15541511]

78. Heinzle J, et al. Topographically specific functional connectivity between visual field maps in the human brain. Neuroimage. 2011; 56:1426–1436. [PubMed: 21376818]

79. Freiwald WA, et al. A face feature space in the macaque temporal lobe. Nat Neurosci. 2009; 12:1187–1196. [PubMed: 19668199]

80. Brincat SL, Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci. 2004; 7:880–886. [PubMed: 15235606]

81. Mineault PJ, et al. Hierarchical processing of complex motion along the primate dorsal visual pathway. Proc Natl Acad Sci USA. 2012; 109:E972–E980. [PubMed: 22308392]

82. Rust NC, Dicarlo JJ. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J Neurosci. 2010; 30:12978–12995. [PubMed: 20881116]

83. Williams MA, et al. Only some spatial patterns of fMRI response are read out in task performance. Nat Neurosci. 2007; 10:685–686. [PubMed: 17486103]

Box 1

**Methodological issues in decoding from time-resolved brain-imaging data**

The ability of MVPA to extract information from a complex multidimensional dataset has rendered this statistical technique indispensable to the neuroimaging community [11]. By concentrating the information distributed in several data points, MVPA increases the signal-to-noise ratio and facilitates single-trial decoding. These advantages are particularly useful when signals present a high interindividual variability that prevents spatial averaging. MVPA simplifies such complex datasets by presenting the results in an information space of direct relevance [11,70] in which data from multiple subjects can readily be averaged.

Here we specifically emphasize the usefulness of MVPA applied to time-resolved intracranial, MEG, and EEG recordings. MVPA projects such multidimensional data into a smaller-dimensionality space whose axes are the cognitive codes of interest (e.g., stimulus features, subject response). The output axes are shared across individuals, but the projection is optimized for each subject and each time point. In MEG, for example, this approach allows maximizing the spatial resolution of the signals without necessarily addressing the difficult issue of source localization.
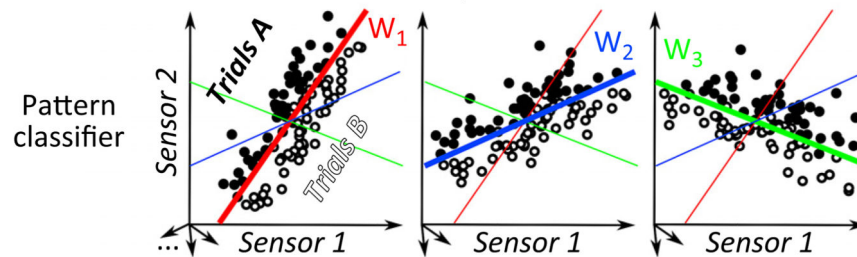
In fMRI, MVPA is often used in combination with a searchlight method [71,72], which comprises sliding a small spatial window over the data to detect which segments contain decodable information. By analogy, with time-resolved signals, a sliding time window can be used to detect periods of optimal decodability. Furthermore, this approach can be extended by testing whether a classifier is able to generalize to other moments in time. Not only can decoding determine when and for how long a given piece of information is explicitly present in brain activity, but with generalization it becomes able to characterize whether this information recurs in time, and when.

When applied systematically, the method yields a full temporal generalization matrix where classifiers are trained and tested at all available time points (Figure I.). Various methods can be used to build such a matrix, including support vector machines (SVMs) (e.g., [57,61]), linear discriminant analyses (e.g., [35,43]), or even simple linear regression (e.g., [58,59,73]) (see [36] for empirical comparisons). To avoid data over-fitting, MVPA must be nested inside a cross-validation loop, such that performance is measured only on novel left-out data samples (for both diagonal and off-diagonal time points). To evaluate classifier performance, 'criterion-free' estimates should be preferred over mean accuracy, because the latter may lead to systematic biases during generalization (i.e., all trials may be classified in the same category). When dealing with a two-class problem, we favor using the area under the curve (AUC), a sensitive, nonparametric criterion-free measure of generalization.
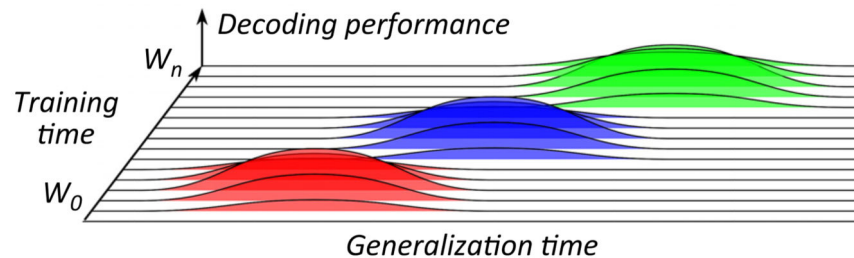
1. A differential brain activity pattern is recorded at each time point.

t_1 t_2 t_3

Differential brain activity pattern (**A**-Ⓑ)

Sensors
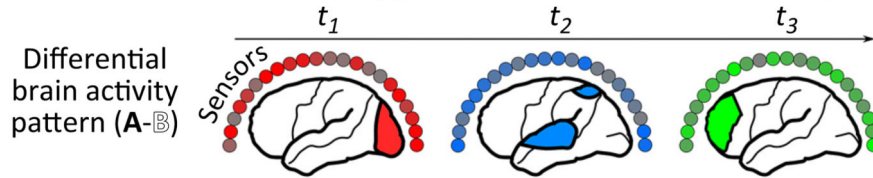
2. A classifier is trained at each time point.

Pattern classifier

Trials A W_1
Sensor 2
Trials B W_2 W_3

...   Sensor 1   Sensor 1   Sensor 1

3. Each classifier is tested on its ability to generalize to all time points.

Decoding performance

W_n

Training time

W_0

Generalization time

TRENDS in Cognitive Sciences

Figure I.

The principles underlying temporal decoding and temporal generalization. Multivariate pattern analysis (MVPA) applied to time-resolved signals allows tracking mental representations across time by detecting and comparing the patterns of neural activity recorded across time. Specifically, at each time point t, a multivariate pattern classifier $w(t)$ is trained to identify the linear combination of measurements [e.g., weights of electroencephalography (EEG) electrodes, magnetoencephalography (MEG) sensors, individual neurons] that best discriminates two or more experimental conditions (e.g., A and B). This combination may vary with time as the underlying pattern of activity changes. Classifier performance is assessed not only at the time used for training [e.g., classifier $w(t_1)$ is tested at t_1, $w(t_2)$ is tested at t_2,…, hereafter referred to as 'diagonal decoding'], but also on data from other time samples [e.g., classifier $w(t_1)$ is tested at all times t_1, t_2, t_3,…, hereafter referred to as 'off-diagonal decoding']. The outcome is a 'temporal generalization matrix' representing the decoding performance of each classifier at each time point. By convention, we depict this matrix with training time on the y-axis and generalization time on the x-axis. With this convention, horizontal 'slices' through

the matrix show the time course of activation of the brain systems identified by a given classifier.

Box 2

Does the brain operate as a decoder of its own signals?

MVPA is a powerful method. Could it also be considered a useful metaphor for some brain operations? Consider the problem of identifying a face from retinal inputs. The primary visual area contains all the information needed to identify a flashed face, yet this information is 'entangled' in a complex manifold of firing rates [74,75]. A central goal of cognitive neuroscience is thus to understand how the hierarchy of brain areas in the ventral visual stream 'disentangles' visual information to make it explicitly coded in the firing rate of a neuronal population – that is, easily decodable by other regions.

For inferotemporal neurons, learning to become sensitive to a specific face may comprise learning a classification function that, given the input firing rates, separates instances of this face from any other stimulus. Similarly, training on a psychophysical experiment may imply that the prefrontal and parietal areas learn to weigh evidence from the relevant sensory neurons and disregard uninformative cells. Each brain area may thus be confronted with a multivariate classification problem similar to that of the neuroscientist attempting to sift through a pile of recorded data.

Asking whether a brain area acts as a decoder of its afferent inputs begs the question of what type of decoding algorithm it uses. First, is the brain confronted with issues of over-fitting and, if so, does it use regularization or penalization schemes, as some pattern-classification algorithms do (e.g., SVM [76])? Does it reject outlier data? Does it suffer from a 'curse of dimensionality' and, if so, does the pyramidal neuron's limited dendritic span provide a solution by drastically reducing the data under consideration?

Second, does the brain use linear or nonlinear decoders? Some theories suggest that once the information is properly encoded in stochastic firing rates, a linear combination of spiking inputs suffices to approximate complex Bayesian operations [77]. Furthermore, the spontaneous fMRI activity in the extrastriate cortex has been modeled as a linear combination of the signals from the primary visual cortex (V1) and secondary visual cortex (V2) [78], thus revealing the structure of corticocortical projections. Recently, the activity of single neurons in the inferotemporal cortex [79,80] or medial–superior temporal cortex [81] has also been modeled as a weighted sum of individual responses to simple lines or local motion at predefined locations. In this case, however, the addition of nonlinear terms improved the model significantly for many neurons. An interesting suggestion is that nonlinearities may enter at only the input and output stages, thus leaving the learning problem itself essentially linear [81]. Such a nonlinear input and output transformation may have the advantage of providing increased selectivity for feature combinations by changing the summation of inputs into a product [79–81].

Understanding which classification algorithm, if any, applies inside the brain is an important goal for learning theory. Ultimately, MVPA should aim to decipher precisely those signals that the brain uses for its internal computations. MVPA should not use exceedingly powerful nonlinear algorithms; otherwise, even abstract information could

be decoded from V1. Rather, MVPA should aim to attain the same performance as subjects' brain and behavior (e.g., [75,82]) and fail on exactly the same trials [83].

Box 3

Outstanding questions

- What factors make certain cerebral operations appear as 'diagonal' and others as 'square' in temporal generalization matrices?

- What other dynamical patterns may appear in brain activity? Does the brain rely on a limited repertoire of canonical dynamical patterns that recur in different types of computation?

- Does each brain region possess a characteristic repertoire of dynamical patterns (e.g., serial flow in the ventral stream, evidence accumulation in the parietal cortex, all-or-none maintenance in the prefrontal cortex)?

- Is there a systematic correspondence between dynamical patterns at the single-neuron and macroscopic (ECOG, EEG, and MEG) levels?

- Does sensory processing comprise a discrete series of stages (as suggested by the labels of EEG components: e.g., P1, N1, P2) or does it involve a continuous cascade of events?

- Why do we often see a reversal of generalization, leading classifiers to perform below chance level when tested away from their initial training time?

- How do experimental factors such as expectation, task relevance, or task difficulty affect the dynamics of neuronal encoding and decoding? Do they systematically modulate, delay, or reorganize brain activity?
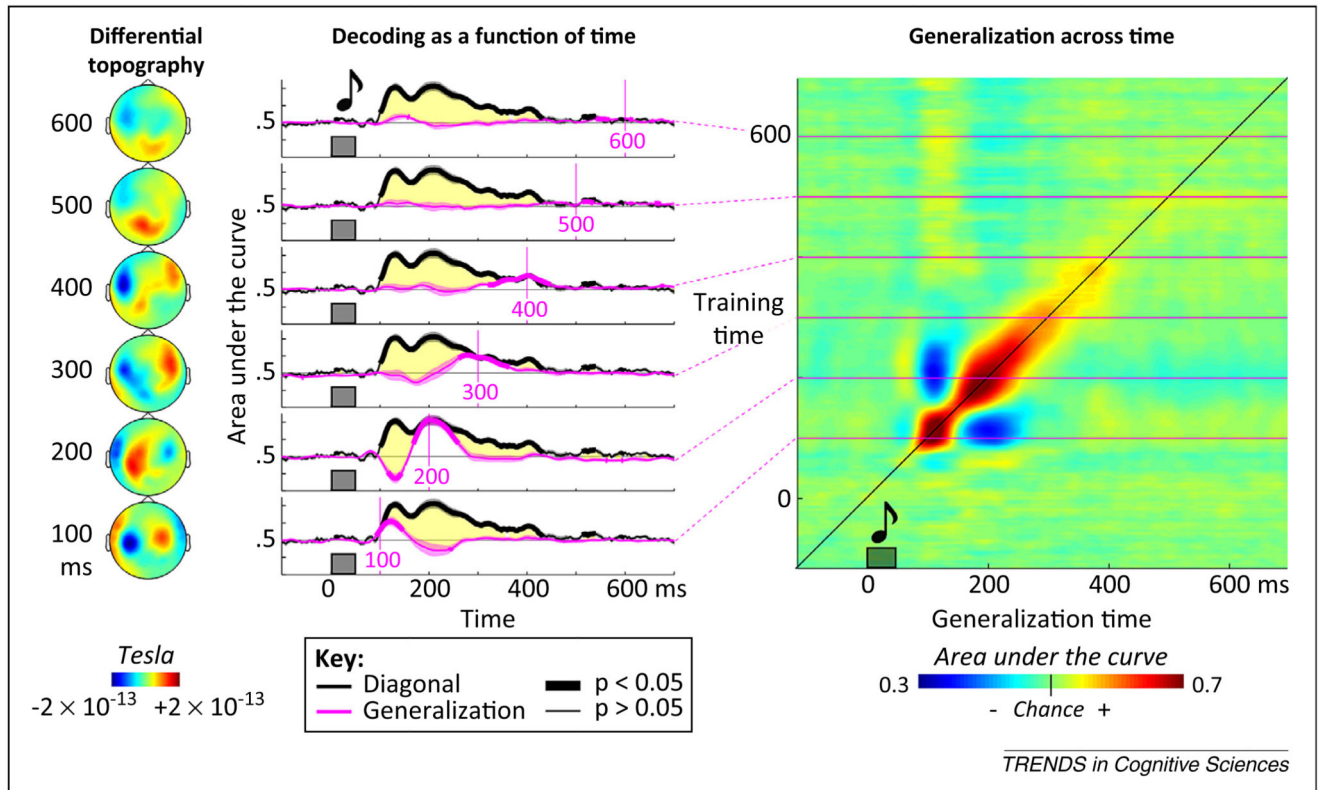
Figure 1.

Example of generalization across time. In an auditory mismatch magnetoencephalography (MEG) experiment, subjects were presented with four identical sounds followed by either the same sound or a different sound (thus generating a mismatch response). Classifiers were trained to determine, at each time point, whether the sound was repeated or deviant. Decoding performance was assessed by the area under the curve, a nonparametric measure of effect size derived from signal-detection theory. Between approximately 100 and 400 ms after sound onset, decoding was above chance (thick black lines). However, each classifier generalized over a transient time period only, as indicated by the diagonal generalization matrix and the fact that diagonal performance (thick black line) was systematically superior to off-diagonal temporal generalization (six pink lines, each indicating a classifier trained at the indicated time). The changing topographies on the left confirm that the differences in magnetic field evolve over time. These findings suggest that the same stimulus (an unexpected sound) elicits a series of distinct patterns of brain activity over time. Adapted from [57].
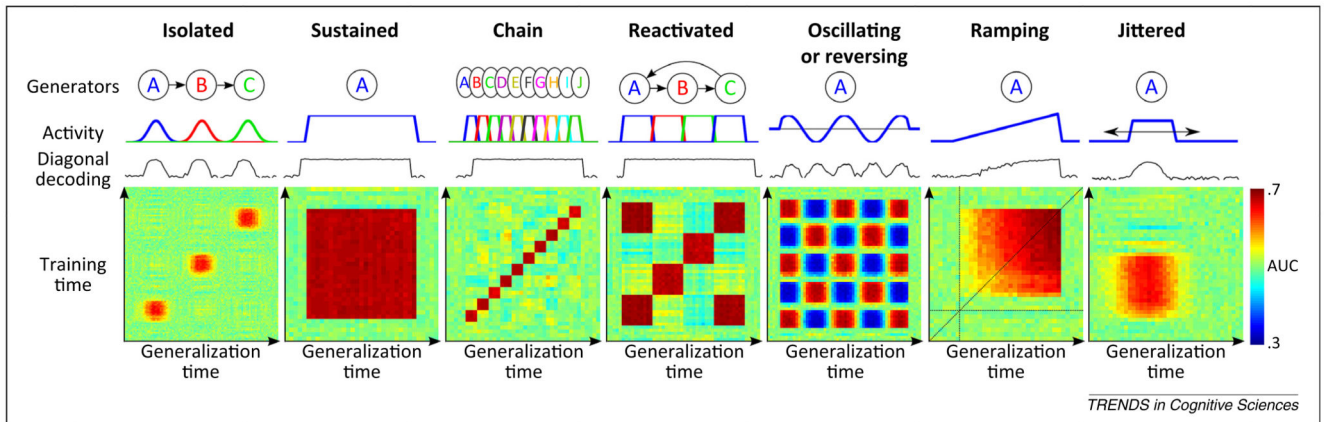
Figure 2.

Generalization across time: principles and possibilities. The temporal generalization matrix (see Figure I in Box 1) contains detailed information about the underlying brain processes. For illustration, we simulated seven different temporal structures. For each one, the generators (neural generators A, B, C, and so on), their time course, the diagonal decoding performance summarized with the area under the curve (AUC), and the full temporal generalization matrix are displayed. Isolated: Three simulated brain regions are differentially activated at three distinct times, leading to three isolated patterns of above-chance decoding performance. Sustained: Analysis of a single process maintained over time leads to a square-shaped decoding performance. Chain: Decoding a chain of distinct generators leads to a diagonal-shaped decoding performance, because each component generalizes over a brief amount of time only. Reactivated: A given generator reactivates at a later time, leading to transient off-diagonal generalization. Note that the maintained, chain, and reactivated conditions are indistinguishable from their diagonal performance, but are easily separated by their temporal generalization matrices. Oscillating: An oscillatory or reversing component leads to transient below-change performance. Ramping: Slowly increasing activity leads to a subtle asymmetry; temporal generalization is higher when the classifier is trained with high signal-to-noise data and tested with noisier signals than in the converse condition. Jittered: Temporal jitter in activation onset smoothes the generalization matrix both horizontally and vertically.
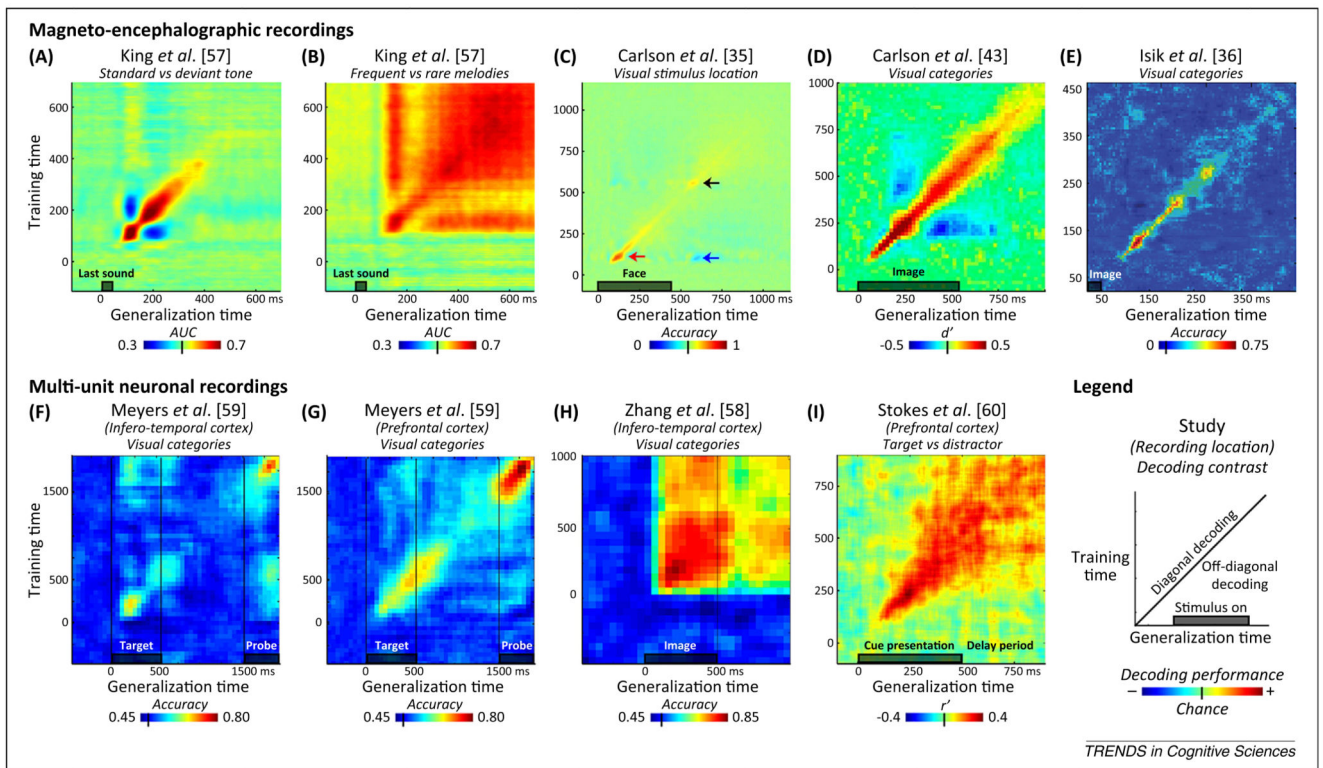
Figure 3.
Examples of empirical findings. Temporal generalization has been empirically tested on both noninvasive magnetoencephalography (MEG) data [35,36,43,57] and invasive multiunit neuronal recordings [58–60] using different decoding categories [e.g., decoding auditory regularities (**A,B**), decoding the position of a visual stimulus (**C**)], different types of classifier (e.g., support vector machine, linear discriminant analysis), different decoding performance metric [e.g., accuracy, area under the curve (AUC)], and different chance levels. The temporal generalization matrices are reproduced from each of these nine studies (**A–I**). The results illustrate some of the major dynamic structures postulated in Figure 2: diagonal chain (e.g., **A,C,D,E,G**), reactivation at the time of a second stimulus (**F,G**) or at stimulus offset (**C**), sustained activity (**B,H**), or transition from diagonal to sustained (**I**). Reversing brain patterns are visible as below-chance generalization performance (below-chance patches in panels **A, C,** and **D**). Interestingly, across studies the same brain region (e.g., the inferotemporal cortex) seems capable of generating very distinct dynamics (compare panels **F** and **H**). Furthermore, similar types of stimulus can lead to very different types of dynamics (compare panels **A** and **B**).
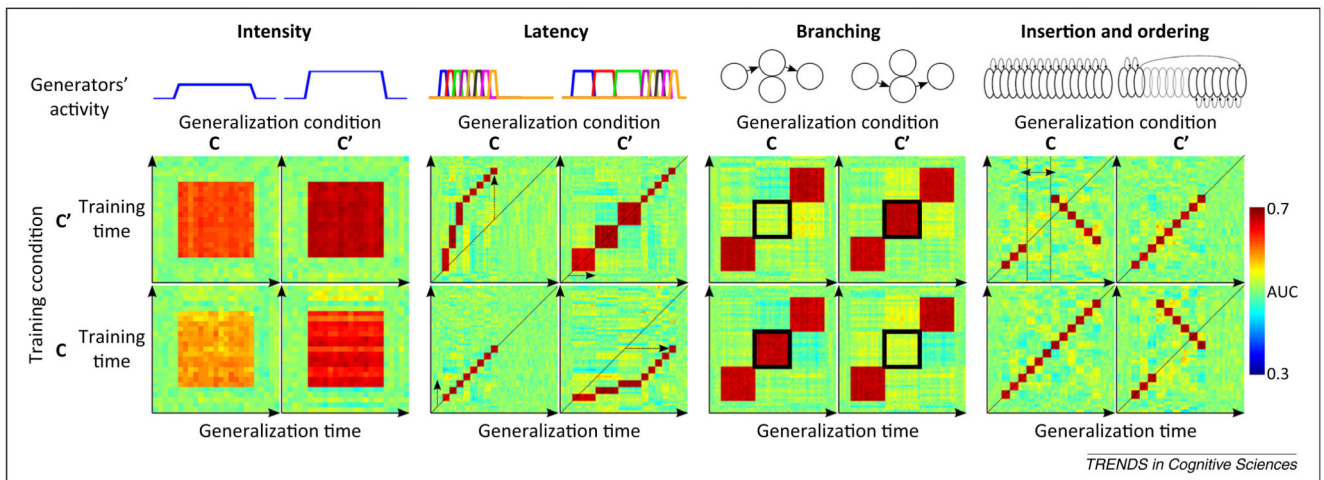
Figure 4.
Generalization across time and experimental conditions: principle and possibilities. When data are available from two experimental conditions (C and C'), the classifiers can be trained in one condition and tested on their ability to generalize to the other. The resulting pattern of generalization across time and conditions provides diagnostic information about the organization of brain processes. The four simulations, using similar principles to the ones detailed in Figure 2, indicate how changes in activation intensity, latency, branching, or ordering each lead to a distinct cross-condition generalization matrix. Note that a change of signal intensity leads to asymmetric generalizations across conditions because it can be more efficient to train a decoder with relatively high signal-to-noise ratio (SNR) data and generalize it to low SNR data rather than vice versa. Changes in the speed of information processing (latency) and in the selection (branching) and order of processing stages can be tracked by comparing the temporal generalization matrices across conditions.