

# Sequence Coding and Learning

Gilles Laurent

## Abstract

The topic pertaining to “sequence coding and learning” is deceptively wide ranging. Sequences or activity patterned in time in the brain can be linked to the nature of the physical world, to the nature and needs of active motion in the world, and to the fact that brains are dynamic systems. Coding, as it pertains to brains, is by contrast a rather fuzzy, ill-defined, and plastic notion. In the context of this discussion, it refers to the idea that information is contained in patterned activity. This chapter discusses some of the assumptions that go into such a statement, and reasons why we should tread carefully when addressing the issue. For instance, circuit dynamics might exist and be useful, even if the spike patterns they generate are never explicitly decoded. Linking sequences to codes is thus far from trivial. While learning requires no definition here, it is useful to be reminded of the likely coevolution of learning and representation systems in the brain: perception and memory depend on one another; this implies that the rules underlying brain “codes” must be constrained, at least in part, by those underlying learning (and vice versa). Given this, how do we go about testing the idea that learning helps generate stable dynamic attractors for perception and action?

## Sequences

I can think of at least three groups of reasons to introduce the importance of dynamics in brain activity. Some are a reflection of temporal structure in the world. Others are linked to the constraints of action in a physical world. The third set is internal and linked to the dynamic nature of neurons and circuits (Gerstner 1995; Theunissen and Miller 1995). This division is somewhat artificial and used here solely for the purpose of description. Most likely, all features, causes, and consequences of brain dynamics have been influenced by one another throughout the evolution of neural systems.

## Temporal Structure in the World

Life, as seen from an individual’s own perspective, could be described as a sequence of interactions with the world that occurs between birth and death.

This sequence is locally nonrandom, due mainly to the laws of physics—where this cat stands now strongly predicts where it will be some time from now; the earth-day has a certain periodicity, etc.—and therefore contains large amounts of information (e.g., correlations) which are of great value for survival. Such correlations occur over a variety of timescales (subseconds to years); not surprisingly, brains evolved to detect and take advantage of them. This can be seen, for example, in motion sensitivity, in the deciphering of speech/song sound patterns, in circadian entrainment, and in many other pattern-sensitive attributes of sensory systems (Carr 1993). This definition extends to sensory experience which is itself the result of sequences of motor action, as in the emergence of place fields in rodent hippocampus and in their sequential activation during behavior or certain phases of rest and sleep (O’Keefe 1993; Mehta et al. 2000).

### **Action as Ordered Sequences of Muscle Recruitment**

Conversely, brains generate motor actions through the ordered recruitment of muscle groups. This, too, is greatly constrained by the physics of the world (air, ground, water), of the plant generating the action (skeleton—when there is one—tendons, muscles, etc.), and of the interactions between them. Not surprisingly, therefore, action is often the expression and consequence of highly optimized motor “programs” encoded as sequential muscle recruitments in specialized circuits (Georgopoulos et al. 1986; Ekeberg et al. 1991; Llinás 1991). The mapping between motor program and action is, however, not always simple. Insect flight, for instance, can be generated by the nervous system in two main ways. First, motoneuron firing frequency determines the wing-beat frequency in a one-to-one manner. In such species (most large flying insects), wing-beat frequencies are usually low. Second, motoneurons serve simply to bring the myo-skeletal system in a resonant regime; in this way, wing-beat frequency can far exceed neural driving frequencies (e.g., flies, mosquitoes). Neuronal patterning is thus adapted to the physics of the plant and to a number of complex features, such as inertial and viscous forces generated by motion in a viscous medium. Sound production by animals obeys the same basic rules and constraints. Most motor programs (speech/song production, locomotion, skilled action) require refinement through learning and thus must involve some comparison between a measure of execution and a kind of template or desired output. Some programs are executed in an open-loop configuration, but even those often require calibration through learning. Here, as well, the timescale of such programs of sequential activity is widely variable (subseconds to hours), implying learning rules with wide dynamic ranges, or many learning rules to cover this dynamic range well.

## Brains as Dynamic Systems

A third source of dynamics in the brain is purely internal. Brains are groupings of neurons and ordered couplings between them. Because neurons and synapses are themselves endowed with dynamic properties (FitzHugh 1961; Connors and Gutnick 1990; Izhikevich 2001)—due, for example, to the capacitive and resistive (usually nonlinear) nature of membranes, to diffusion and depletion, to channels and receptors kinetics—neural circuits are complex dynamic systems, with preferred behaviors and collective emergent properties (Brunel 2000; Gerstner et al. 1996; Maass et al. 2002; Rabinovich et al. 2008; van Hemmen et al. 1990). Some of them are relatively easy to detect and describe: oscillations or limit cycles (von der Malsburg 1981/1994; Gray et al. 1989; Eckhorn et al. 1988; Singer and Gray 1995; Engel et al. 1991; Laurent 1996, 2002), propagating waves (Lubenov and Siapas 2009). Others are not: complex spatiotemporal patterns of activity lacking periodicity or obvious large-scale spatial order (Diesmann et al. 1999; Stopfer et al. 2003; Mazor and Laurent 2005). While the kinds of patterning described in the previous two sections are linked causally to interactions with the external world (e.g., think of both the sensation and the production of sound), the kinds of sequences and patterning alluded to here are linked to the external world in only an indirect way. The sequences of activity seen in the early olfactory system in response to odors, for example, are not a result of sequences in a time-varying input; they are a consequence of, and therefore reflect (possibly encode), a particular, time-invariant, odorant signal; they are solely generated by and located in the nervous system. Generally speaking, this describes cases where the dynamic attributes of the representation are not a simple reflection of the dynamics of the input that drives activity (Rabinovich et al. 2008). Oscillatory patterns of activity (in the retina, in visual cortex, in olfactory systems, in hippocampus) are one form of such temporally ordered activity. A looming question, therefore, is whether those dynamics are useful and if so, for what?

## Coding

### Linking Neural Dynamics to Coding Is Difficult

I would expect that most attendees at this Forum would agree that in our field of experimental systems neuroscience, there exist at least some incontrovertible examples of spatiotemporal patterning, indicative of some form of collective order (e.g., synchronization, periodicity, waves of activity, seemingly random yet deterministic sequences of firings). Clearly, more experimental work is needed to better assess the brain's dynamic landscape, to characterize the properties and behavior of specific systems, and to classify the observed diversity of patterning. I would argue, however, that single clear examples are

sufficient to prove the existence of such ordered phenomena in neural circuits and justify the study of their causes, consequences, and significance. Where this is accomplished (i.e., with which circuits, brain regions, or animal species) has no real importance to me at this point, other than practical considerations.

What is less clear, in my opinion, is whether we can say with equal confidence that such brain dynamics have been clearly linked to coding in the above examples. I can see at least two reasons for this hesitation.

First, coding (as in, “neural codes”) is a rather fuzzy notion in neuroscience. In sensory neuroscience, it is now conventional to assess coding (as in, these retinal ganglion cells code for X) on the basis of (for instance) a measure of mutual information between an input and a response set (or distributions) (Bialek and Rieke 1992; Berry et al. 1997). Yet, neural codes are not codes in the sense of Shannon’s (1948) information theory: brains are not channels optimized for information propagation. A brain selects, prunes, “computes,” and generates a highly selective, adapted, and specialized output—not a copy of its input. A brain is also rife with complex recurrent paths, impossible to map simply onto a source-receiver model of information flow. However, the information theory approach is undoubtedly valuable: it defines what can be extracted by an observer from a spike train. The problem, in the end, lies with defining the observer (can or should a neuron be considered to be equivalent to an ideal observer?) and its approach to read-out (e.g., spatiotemporal inputs, initial and time-varying states, nonlinearities). For practical reasons, I would thus define neural codes very locally and base those definitions on the properties of the “decoder.”

Second, proving that temporal structure in neural activity has any functional consequence is an extremely difficult task. It is difficult experimentally because selective manipulation (in space and in time) of spiking has, until recently, been impossible on a large and distributed scale. (While still lacking high spatial specificity, recent optogenetic techniques now enable the temporal manipulation of cell populations.) Proving the relevance of temporal patterning is difficult as well because brain activity patterns are the expression of a collective behavior, itself prone to adaptive rearrangement: to take a crude analogy, the loss of a midfielder in a soccer team can be rapidly overcome when the other players assume new roles and behaviors. This compensation is not solely an expression of redundancy; it reveals an ability of the system to change adaptively. The criteria traditionally applied to experimental data (necessity and sufficiency) to assign functional significance may thus often be inadequate. We need to be open to different/additional assays and criteria.

### **Decoupling “Encoding” and “Decoding”**

An argument often advanced by those who doubt the relevance of temporal codes is that they are difficult to read out. While this argument is disputable, I would like to introduce the notion that dynamics and patterning do not

necessarily mean codes: more precisely, what we often describe as “encoding” and “decoding”—the two sides of the “coding” coin—may not always be coupled to one another. Stated differently, dynamics and temporal patterning may be useful even in cases in which the patterns themselves are not decoded.

Imagine that a particular sensory input,  $S$  (e.g., an visual scene), causes, in a particular part of the brain, a complex spatiotemporal pattern of activity,  $PS$ . In purely descriptive terms, one could say that  $PS$  represents or “encodes”  $S$ . This is technically true from our perspective as outside observers. It might be, however, that the  $PS$  is nothing more than the expression of the brain’s physical response to  $S$ , as would be that of many other similarly stimulated dynamic systems. Furthermore, it is possible that the decoding of  $PS$  by downstream neurons or networks involves no sequence decoding per se. This does not mean that  $PS$ , as a dynamical pattern, is not relevant. It may, however, reveal that its value is only implicit in the code.

### *Analogy*

If you have learned a racquet sport such as tennis at one point in your life, you will have learned the value of the “follow-through”—the seemingly stylistic exercise of keeping the racquet in motion toward some imaginary position after the racquet has hit the ball. This, of course, appears completely nonsensical to a beginner: “Why does it matter what I do after I have hit the ball?” Trivially, the answer is that what I do after the time of contact is at least in part a consequence of what I did before, including at the time of contact; what I do after the hit is a consequence of the unfolding of a sequence of actions that preceded it: I cannot have one without the other because it is part of the physics of the system (the outside world, the ball, the body and, possibly, the brain itself). As far as the outcome of the game is concerned, however, the decoding of this action (i.e., the quality and accuracy of the hit) is transitory: the action sequence is implicitly contained in the hit, but it is not decoded as a sequence of movements. To summarize, the action sequence (“encoding”) is necessary, but the read-out (“decoding”) is done only transiently, at some appropriate moment throughout the motion. Hence, the dynamics, as the expression of a system’s physical properties, are indispensable but do not embody the code itself.

### *Example*

The system that my lab has studied for 15 years or so, the insect olfactory system, is dynamically rich. When an odor is presented, a population of some 1,000 projection neurons (PNs) in the antennal lobe (the functional analogs of mitral cells in the vertebrate olfactory bulb) becomes transiently synchronized, forming a distributed oscillatory pattern ( $\sim 20$  Hz) in which different subsets of PNs fire at each oscillation cycle. An odor representation is thus a specific time series of PN-activity vectors, updated at each oscillation cycle; these sequences

of PN-activity vectors can also be imagined as odor-specific trajectories in the phase space defined by this system. These trajectories are stimulus specific and, provided that the stimulus is sustained for a long enough duration ( $>1.5$  s), they contain three epochs: (a) an on-transient that leaves the baseline state, (b) a fixed-point attractor, and (c) an off-transient that leads back to baseline along a path different from the on-transient. The transients and the fixed point are all odor specific, but analysis of pairwise distances between the trajectories that correspond to different stimuli shows that distances are greatest when the system is in motion (during both on- and off-transients), not between steady-state attractors corresponding to the different odors. Interestingly, the targets of these PNs, called mushroom-body Kenyon cells (KCs), are active mainly when PN activity is in motion—not when PN steady-state has been reached. In fact, KC responses are highest when separation between PN vectors is greatest (i.e., during the initial phase of the on- and the off- transients). This leads to two conclusions: First, while circuit dynamics seem to serve an important function (here, decorrelation of representations), they do not appear to serve a coding purpose in the traditional sense; that is, there is no evidence that downstream areas decode the sequences of PN output. Second, a significant portion of a trajectory, its fixed-point attractor, does not even appear to be read out by targets (Mazor and Laurent 2005). This begs the question of the use, if any, of fixed attractors in this system. Whatever this use may be, it does not seem to be in the embodiment of a representation.

In conclusion, the existence of dynamics in a representation does not, in and of itself, prove that the dynamics are part of the “code,” or a message to be decoded. The dynamics may be the result of properties of the system, and they may even be useful (e.g., in optimizing representations). However, they are not necessarily a feature that requires explicit decoding.

## Learning

### Learning and Perception Are Not Separable

The practicalities and sociology of neuroscience as a human endeavor are such that the subareas of learning/memory and perception/coding overlap only occasionally. Attendance at any large neuroscience meeting will often confirm this impression. Yet, perception is difficult to imagine without learning, and vice versa. Perception and recognition would not be what they are without templates/representations stored for classification and comparison, or without some trace of recent history. Conversely, the ability to learn seems pointless if it did not serve future comparisons between immediate input and a bank of memory traces—comparisons that are needed for perception, recognition, and adaptive action. It follows then that the mechanisms underlying sensation and perception should, somewhere, express the requirements imposed by

learning, storage, and recall; this is simply because circuits must have evolved with these coexisting constraints. When we talk about neural codes (ignoring for now the ambiguity of the term), therefore, we should not forget that their formats may be optimized not for coding per se, but for learning and recall as well. In other words, the attributes of biological learning rules, presumably adapted to the statistical structure of the world and to the intrinsic properties of the brain, should be interpreted as an added constraint on the formats of sensory and motor codes. This is particularly relevant to our thinking about “neural codes,” especially if those (a) are at least partly dynamical in nature or (b) have substrates that express dynamical properties. I am reminded of this every time I use my bike’s lock, a perfect example of procedural memory: while the lock’s combination escapes me now, it springs back to me reliably every time I start spinning the rotating face of the lock. In conclusion, thinking about codes and representations can only benefit from including the constraints and necessities of learning and recall.

### Learning Rules and Sequence Coding

Recent experimental results on the mechanisms of plasticity (long-term potentiation, asymmetric Hebbian rules, spike timing-dependent plasticity) have allowed realistic links to be drawn between learning and network activity (Hebb 1949; Herz et al. 1989; Bliss and Collingridge 1993; Markram et al. 1997; Bell et al. 1997; Bi and Poo 1998; Cassenaer and Laurent 2007). In a typical example, an externally imposed sequence of activity leads, through a learning phase, to the selective modification of synaptic weights via application of a learning rule with an appropriate time window; thus, the training sequence can, after learning is terminated, be recalled or replayed in the absence of the external drive, very much as in a feedforward “synfire” chain. The practical consequence can be the recall of a full sequence (as a path in physical space or as in procedural learning), a bias for recall of a particular pattern upon appropriate seeding (as in the recall of selective memories). In all cases, the main idea is that the pattern has become a stable attractor for activity. The theoretical questions that these ideas raise are many and complex:

- What are the best dynamical models to explain such behavior? What do they predict?
- How stable can such dynamic attractors be, given the known biophysical properties of neurons and synapses (and their non-fully deterministic behavior)?
- Similarly, what is the tolerance to external noise for such representation mode?
- What is the gain on memory capacity (relative to classical fixed-point attractor models) for such systems?

- What are the constraints that such rules impose on representation density/sparseness?
- Do such representation modes require (or work best with) a discrete timeframe?
- Do they allow time compression/warping?

As a dedicated experimentalist, I would argue that many of the critical answers (or hints to those answers) will be given by experiments and offer the following as examples of crucial issues to be addressed:

- What are promising experimental systems to study these issues?
- What are the criteria needed to establish firmly the existence of spatio-temporal patterning in circuits? In other words, if I were given a neuronal system with dynamical attractors, would I even detect its nature?
- What are the criteria needed to establish the functional relevance of spatiotemporal patterning?
- What are the different forms that this relevance could take?
- What constraints would the existence of such attractors impose on the learning rules we know?
- Assuming that many of the learning rules we know are used mainly for homeostatic regulation (e.g., on weight distributions, timing), how plastic must they be to underlie the formation of stable dynamical attractors?