through dynamical network modeling that explicitly allows for a hierarchy of spatiotemporal levels.

Conclusion

The contributions to this volume differ in their theoretical outlook and in their methodology, which ranges from neuroanatomy, electrophysiology, and functional imaging, through computational analysis and behavioral experimentation, to philosophical analysis. As such, they offer a wide variety of perspectives on the issues about which the authors are in agreement: the dynamical nature of phenomenal experience, and the need to understand and model it as such. Given the profundity and the difficulty of the fundamental questions that arise in consciousness studies, their eventual resolution will require a concerted, interdisciplinary effort on the part of the entire community of researchers. We hope that this book will help generate the right kind of dynamics in this most exciting field of philosophical and scientific endeavor, and look forward to the phenomenal experience of watching this dynamics play out.

Shimon Edelman

Tomer Fekete

Neta Zach

February 2012

Time after time

Temporality in the dynamic brain

Dan Lloyd
Department of Philosophy and Program in Neuroscience, Trinity College, Connecticut, USA

A useful theory linking dynamical systems to phenomenal experience will be a story thrice told. It will involve some description of phenomenal experience, which should be true. It will also involve some sort of dynamical model (second). But (third) the model will have to be at least plausibly implementable in human beings – that's where the theory becomes useful. Finally, once all three stories are told, they must align. It should be evident to all that the phenomenal story, the dynamical story, and the implementation story are really one story, about one entity, described in three different ways, akin to one story as it might be told in three different languages. A theory of consciousness then is an exercise in translation, somewhat like deciphering the Rosetta Stone. This chapter outlines a possible alignment with respect to a foundational, structural property of experience, namely, time.

1. Introduction: Ubiquitous time

Philosophers in the phenomenological tradition have always stressed the importance of temporality in all experience. For them, our awareness of objects and other people, as well as our awareness of our own psychological states, is not just an awareness of the here and now. Rather, our sensory manifold is embedded in the conscious yet non-sensory context of the past and future, as if the objects of the sensory world each had a time line or history projecting from them. Beginning in 1905, Edmund Husserl (Husserl 1966 (1928)) offered the most elaborate modern account of temporality in perception, an account that has been elaborated but not overthrown in Husserl's successors: Heidegger, Sartre, and Merleau-Ponty, among others. In Husserlian terminology, an "act" of consciousness has three aspects. There is its sensory facet, the information streaming right now from the sensed objects in my immediate perceptual field, which Husserl called the "primal impression." Looking at a cup of coffee, the primal impression comprises sensations and apprehensions of the immediate here and now. And there is the "primary memory" of the immediate past of every component of awareness. Husserl called this "retention" – that is my awareness of the recent history of my cup of coffee. And there is a future-facing "protention," which is my immediate anticipation of what will or may happen next. If I grasp and lift the coffee cup, I expect it to have a certain weight and to move freely with my hand. If I turn it over, I expect the coffee to pour out. If the cup turns out to be empty, my surprise reveals the anticipation that was built into my awareness of the cup.

Phenomenologists construe temporality as a structural feature of all states of consciousness, part of the "infrastructure of reality" (Zahavi 1999). Every moment of consciousness is temporal, and every element of conscious awareness has its temporal extensions into the immediate future and immediate past. Protention and retention are folded into the present, at every instance of consciousness. In effect, for Husserl time comprises two dimensions. The vertical structure of protention, primal impression, and retention defines an inner time dimension, and this whole tripartite structure is in continuous flux, a horizontal flow of time in which the three phases of experience are constantly changing. As time flows, protentions are continually drawing near to the primary impression where they will be fulfilled, or not, and primal impressions are continually sliding back into immediate retentional memory. To make things more complicated, the entire tripartite structure is itself a temporal object in one's awareness. The retention that I experience now of the moment just passed will itself be retained, a retention of a retention. This recursive elaboration is continuous, and extends in anticipation and retention to an indefinite temporal horizon. Over that horizon our temporal awareness is mediated by distinct psychological processes of recollection and explicit anticipation (as would be involved in prediction or planning).

This way of thinking about perception is quite different from the approach taken in mainstream cognitive science. In cognitive science and perceptual psychology, the problem of perception is usually understood as the process of interpreting the immediate sensory field, the primal impression. In David Marr's classic, *Vision,* for example, visual perception is entirely conceived as a process of recovering three dimensional objects from the flat retinal image (Marr 1982). The perception of the temporal evolution of perceived objects plays no role in this process. Temporality is still largely ignored in contemporary cognitive neuroscience. As in cognitive science overall, neuroscience assumes that perception is driven by the immediate environment, and that the brain is locked or clamped to that environment. So, for example, if I gaze steadily at the coffee cup, my brain enters a steady state of cuppishness, which changes to something different when I look away. But then when I return my gaze to the cup, I return to that same state of cuppishness. Cognitive neuroscience assumes that the brain's responses to the world are "stationary", that is, always the same when external conditions are the same and unchanging due to the passage of time. The assumption that the same perceptual object is correlated with the same state of the brain supports almost every experiment in cognitive neuroscience. In contrast, phenomenal temporality contradicts this assumption of environmental dependency. As I stare fixedly at the cup, my temporal experience of it is continually changing, as I am aware of the ever-changing duration of my looking as a continual elaboration of the retention of my perception of the cup. I'm seeing it for one second, now two, now three, and so forth, and I'm aware of the growing duration of the cup in this location, its history of stability. Eventually I'll achieve the temporal condition of boredom. And if I look back at the cup, my second look is not the same as my first, even if the cup has not changed. I'm aware that I'm seeing the cup *again*, and this is different from seeing for the first time.

Thus cognitive neuroscience omits the consideration of a salient dimension of human cognition and consciousness. Its methods and assumptions are blind to time. But phenomenological temporality is also not without its problems. Husserl describes the flow of temporal experience somewhat like the flow of a river. He acknowledges, however, that temporal experience includes internal distinctions analogous to the experience of spatial objects. For example, interest and attention can highlight some aspects of the temporal field, yielding a foreground and background analogous to figure and ground in visual space. But in general Husserl's writings on time do not offer more than suggestions about how to think about the articulation of subjective time within the foundational structure of protention/primal impression/retention. Phenomenology thus faces a problem that is the converse of that faced by cognitive neuroscience. Against the background of ever-changing time, we nonetheless perceive stable objects and scenes, and recognize their re-occurrence. This too seems like a fundamental aspect of experience.

Both phenomenology and cognitive neuroscience thus leave out important dimensions of human experience, but neither is wholly false. It may be helpful to regard these not as binary opponents but rather as marking two ends of a spectrum of possible models of cognition. Between these antitheses, then, we could look for synthesis, some intermediate analysis that would respect the fact that we do experience a stable world in which objects and situations fall in categories we can recognize over time, while also respecting the fact that time itself generates ceaseless change. If temporality really is a ubiquitous feature of all conscious experience, then the brain must somehow embody temporality.

2. A middle way: Dynamical systems

A dynamical system, like a thrice-divorced heiress, is a system with a past. But unlike persons with baggage, that loaded past is entirely determinative of the present (and thus the future). To call something a dynamical system, accordingly, is to choose a kind of description rather than to discover some fact of nature, since any system can be dynamically characterized. In this respect the "dynamical systems perspective" is like computationalism, which also describes every system, seen through its lens (Edelman 2008). The DS perspective highlights certain features of systems, and these can be useful. Among these is the idea of a state space, an abstract expanse of many dimensions, one for each variable of interest in the system. Seething vectors condense to wandering points. Geometrical intuitions apply in the state space to a certain extent. For example, distance in its comfortable Euclidean sense can characterize similarity and dissimilarity among dynamical states. But the point analogy is misleading as well. Naked points (or positions) in the physical world are usually uninformative, or in other words points seem not to be capable of encoding complex information. However, in state space, as in the real estate business, location is everything. Magnitude along each dimension of the space is a concrete value encoded in every point.

I look at a coffee cup, and appropriate regions of the brain are excited by its shape, proximity, possibilities, and more. Cognitive neuroscience meets dynamical systems theory when it surmises that each regional activation is a dimension of a brainy state space. The percept of a cup is a point in that space. Phenomenal temporality sets that point in motion. Thus, there is no point corresponding to coffee cups or even this coffee cup. Instead, there is a region in brain space which accommodates the trajectory of the coffee cup. Within that subspace, the cup ten seconds ago is distinguished from the cup now. Furthermore, phenomenology subordinates the Now to retention and protention. Neither of these Not-Nows is part of the current instantaneous environment, so the content of the current point must in some sense encode both the past and future. But how? One way to answer this question is to look and see.

3. Brainspace

The dynamical systems perspective offers some very accommodating metaphors for temporality, but can the trajectories of time be detected in biological brains? The body of this chapter comprises a re-analysis of fMRI data in a particular experiment, a case study affording a look at the methods and data that may inform a scientific neurophenomenology. Can the data of brain science, neuroimaging in particular, be reinterpreted to discover the neural traces of Husserlian temporality? Subjective time should be an aspect of every experience, and thus every brain, provided we use the right interpretive techniques.

Here the target is one neuroimaging study, conducted at the Olin Neuropsychiatric Research Center in Hartford, Connecticut, and I thank the researchers there for their willingness to share their many gigabytes of data.[1] In this experiment, participants drive a simulated automobile through a virtual landscape, controlling their route with a steering wheel, and brake and gas pedals. The experiment is divided into three runs, and each run has the same structure of three parts: First, subjects see a black screen with a plus sign, a "fixation cross," in the middle, for 30 seconds. Then the virtual reality simulation begins, and subjects drive through the simulated landscape, for 90 seconds. Then, subjects move through the simulated landscape again, but it is as if they are passengers in the automobile, merely observing but not driving. This lasts for 60 seconds. This same sequence of Fixate, Drive, and Observe repeats three times without intermission, and is followed with a final thirty seconds of fixation. The whole session, then, comprises this sequence:

Fixate – Drive – Observe – Fixate – Drive – Observe – Fixate – Drive – Observe – Fixate

With three runs, the experiment lasts about nine minutes. While the subjects are driving, a number of behavioral variables are recorded. This includes the current position of the car in the virtual landscape, steering, accelerating, and braking, and others. Each session yields 375 whole-brain images, collected every 1.5 seconds. Although no one would confuse the experience in the scanner with actual driving, it is nonetheless a compelling and immersing perceptual and behavioral task. Complex and open-ended, the task comes closer to simulating a natural interaction with the world than the host of button-pushing stimulus-response experiments designed to probe specific narrowly defined facets of perception or cognition. In addition, each session in the scanner comprises three very different subjective states, Fixation, Driving, and Observing, which interweave and repeat.

In their phenomenology the three conditions are certainly diverse: One of them requires action, while two are passive. One presents a nearly blank screen, while two are visually rich and dynamic. Facing this experiment, the standard neuroimaging approach would seek correlations between activity in specific regions

---

[1] Publications from this study include: Carvalho, K.N., G.D. Pearlson et al. (2006). Simulated driving and brain imaging: Combining behavior, brain activity, and virtual reality. *CNS Spectr* **11**(1): 52–62. Calhoun, V.D., J.J. Pekar et al. (2004). Alcohol intoxication effects on simulated driving: Exploring alcohol-dose effects on brain activation using functional MRI. *Neuropsychopharmacology* **29**(11): 2097–2017, Calhoun, V.D., K. Carvalho et al. (2005). Using virtual reality to study alcohol intoxication effects on the neural correlates of simulated driving. *Appl Psychophysiol Biofeedback* **30**(3): 285–306.

of the brain and particular configurations of stimuli and responses, ultimately to explain the behaviors elicited by the experiment as the joint production of functionally specific brain regions. The phenomenology of temporality, however, demands something more than the parceling of brain states into bins to match the experimental task. For starters, the repeating experimental conditions cannot be combined in an analysis of temporality. The first, second, and third runs retain their distinctness in experience. The time line of the experiment must at least be adumbrated to this extent:

Fixate(1) – Drive(1) – Observe(1) – Fixate(2) – Drive(2) – Observe(2) – Fixate(3) – Drive(3) – Observe(3) – Fixate(4)

Moreover, within this articulation temporality demands a continual awareness of passing time. Thus, each moment of the experiment is further articulated by temporality itself, the awareness of elapsed or anticipated durations, embedded in consciousness as a structural feature of each moment of awareness:

Fixate(1.1) – Fixate(1.2) – Fixate(1.3) – Fixate(1.4) –...

At this point the divergence of cognitive neuroscience and phenomenology is sharply drawn. Phenomenology says to neuroscience, show me how Fixate(1.1) is different from Fixate(1.2) and (1.3), etc. and show me how these are different from Drive(1.1), and furthermore how Drive(1) differs from Drive(2). The neuroscientist says back to the phenomenologist, show me how Fixate(1.1), (1.2) and (1.3), are the same, and likewise how the three drive epochs share a common functional basis, and how the three experimental conditions differ.

Wrestling with these questions conversationally leads to dead ends. Thinking of state spaces, on the other hand, leads directly to a reconciliation. Two states are similar if they are nearby in state space, and different to the extent that they are mutually distant. This sliding distinction allows sameness and difference to cohabit the same data. In the present case, we imagine a state space containing three broad territories, mapping the three experimental conditions. But within each territory, we anticipate a spatial subdivision corresponding to the three instances of each condition. So, Drive(1) and Drive(2) can occupy their own counties within the territory shared by the driving condition. Furthermore, the fine-grained temporal distinctions, the ticking of the subjective clock, can be captured as well, as baby-steps within the overall space of each condition in the experiment.

State spaces are capacious, accommodating as many dimensions as there are variables of interest. In the brain, the variable of interest is commonly assumed to be the neuron (or perhaps the synapse) – these have been satisfying basic entities in small scale explanations of signal detection, learning, and behavioral coordination, not to mention their fit in biological explanation overall. The state space of neurons in a typical human brain is high dimensional indeed. We can't track that many variables and even if we could, it's not clear what would be gained – this is the "curse of dimensionality" (Fekete & Edleman 2011). A brain equation of one hundred billion variables would be of slight practical use. Accordingly, with high dimensions comes the impulse to simplify, that is, to combine or eliminate dimensions to build a useful subspace.

With the present experiment, data reduction occurs twice over. First, fMRI is both indirect and blurry in its detection of neurons in action. It's indirect, insofar as it tracks the blood oxygen level dependent (BOLD) response to neuronal activity. It's blurry, in that every voxel is an amalgam of millions of neurons smeared over 1.5 seconds (Logothetis 2002). Despite that formidable funneling, each image in the driving dataset comprises 150,000 voxels, still an unwieldy multitude. The second wave of simplification, then, is Independent Component Analysis (ICA) (Calhoun, Adali et al. 2002). The "components" discovered by ICA are ensembles of voxels that activate and deactivate in unison, "temporally coherent networks" of correlated activity. These voxel teams can be localized but are often distributed across two or more regions of the brain. Each component sings with one voice, but is uncorrelated with the other components. Around twenty components can retain nearly 90% of the variance in most brain data sets.

So billions of neurons appear to the researcher as a temporally and spatially blurred vector of just twenty elements. That is, the state space in this analysis has a mere twenty dimensions. At this point, a sensible person would doubt the possibility of a fine grained analysis of the sort needed to identify the trajectory of time. Nonetheless, the spark of curiosity invites some exploratory data analysis. To begin, we can represent the brain space in even fewer dimensions, using Multi-Dimensional Scaling (MDS), as in Figure 1. Here, a two dimensional map preserves the distances between the points in 20-d brain space, for ten subjects in the driving experiment. Each time point is mapped, and coded for the big distinction between Fixate, Drive, and Observe conditions.[2] Inspection reveals a promising regional distinction among the three conditions, which has been crudely marked – one could draw the boundaries many different ways. The map conforms to a very basic intuition about this experiment, namely, that brains are in different global states during the three conditions. Cognitive neuroscience and phenomenology converge to this extent at least.

---

[2] MDS was implemented using the mdscale function in MATLAB. The analysis was based on the 20-element vectors of ICA magnitude at each of the 375 time points in the experiment. The ten subjects were concatenated, so that each time point was represented as a 200-element vector. From this Euclidean interpoint distances were caclulated, and non-metric MDS performed.

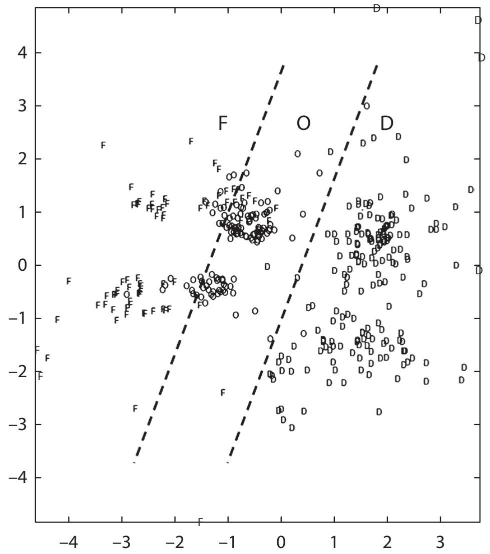

Figure 1. State space of 375 images of the driving experiment, combined across ten subjects. Twenty dimensions have been reduced to two with multidimensional scaling. F: Fixate conditions; O: Observe conditions; D: Drive conditions. For the most part, similar conditions are similarly located in brainspace, as suggested by the dashed dividers

Although the MDS map in Figure 1 is helpful for visualizing brain space, it is also quite inaccurate, since high dimensional inter-point distances must be squeezed and stretched to force 375 points onto a flat surface. "Stress," the term for map distortion in MDS, is high. Cluster analysis, another exploratory method, is not so stressful. We can confirm our map intuitions by clustering the nearest points in the high-dimensional space. Carving brain space into three clusters shows that the distinctions are indeed sharp. Figure 2 is a cluster diagram of a slightly non-traditional format. Clusters are distinguished along the y-axis, and the time points of the experiment pass from left to right. Plusses stand for each point in state space, each assigned to its cluster.

Cluster analysis is more than a convenience for data visualization, however. Cluster membership and boundaries demarcate the capacities of a system for making distinctions. Fekete and Edeleman write:

> The basic means of realizing conceptual structure is *clustering* of activity: a representational system embodies concepts by parceling the world (or rather experience) into categories through the discernments or distinctions that it induces over it. As it gives rise to experience… activity should possess no more and no less detail than that found in the corresponding experience….. [A]ctivity space must divide itself intrinsically into compartments, structured by the requisite within- and between-concept similarity relations.
>
> (Fekete & Edelman 2011, p. 5)

Figure 2 confirms that the gross distinction between conditions in this experiment are represented in the global brain space. The two stages of fMRI observation and ICA data simplification have unwoven the curse of dimensionality without losing the conceptual structure we expect of this cognitive system.

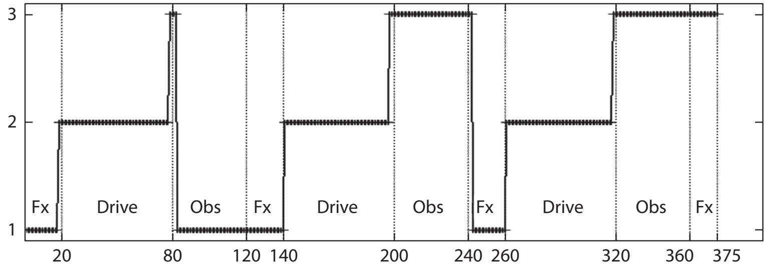

Figure 2. Brainspace in the driving experiment divided into three clusters. Each image is represented by + at its time point, assigned to its cluster. (Clusters are numbered (arbitrarily) on the *y* axis.) Clusters are based on Euclidean distance and partitioned with Ward's method. Each condition is assigned to its own cluster, with the exception of the first Observe and last Fixate conditions

Brain state, cognitive science, and phenomenology converge on the common sense carving of the state space of the experiment. Images from the Drive condition are clustered with 99% accuracy. Overall, the clustering is 83% correct.

Once we go beyond three clusters, however, we can begin to explore brain-based categories, to probe whether further divisions are driven by time. By constraining the analysis to *n* clusters, we examine the dynamical brain at different levels of granularity (Fekete & Edleman 2011), depending on *n.* Temporality is displayed if at some granularity temporally adjacent time-points are clustered together. Figure 3 displays cluster anatomy at 5, 8, 12, and 20 clusters, successively zooming in on the fine structure of brainspace. Figure 3A, with five clusters, already displays the temporal separation of the first occurrences of the three experimental conditions. Overall, the state space segregates into early, short segments and later, longer segments. This repeats at different scales. The Drive epochs also resolve into repeating short-long units (panel B, with eight clusters). Twelve clusters separate all three conditions into separate spaces (panel C). Finally, with twenty clusters available, the Observe blocks separate into their temporal parts, while the Drive epochs further subdivide. Overall, the largest territories are categorical, reflecting the distinctions between the experimental conditions, but these resolve into temporally contiguous subgroups. Time appears to be the modifier of all three conditions. Finally, Figure 4 carves brainspace into 60 territories.

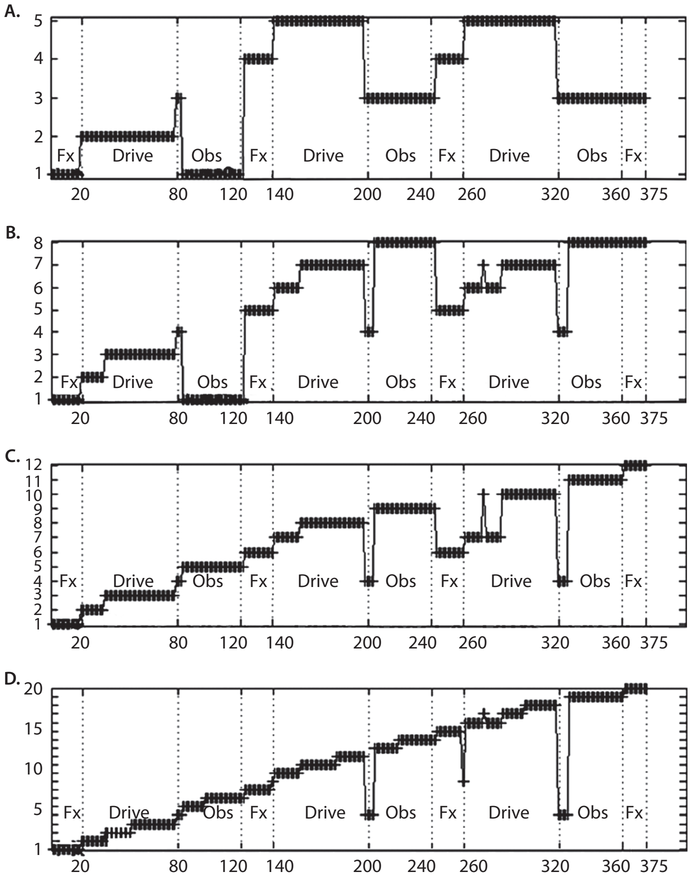

Figure 3. Cluster analyses at different levels of subdivision. A. Five clusters (numbered arbitrarily on the *y* axis) show separation of early and later epochs along with segregation by conditions. B. Eight clusters show within-condition subdivisions, also based on time. C. Twelve clusters almost completely separate same conditions by time. D. Twenty clusters show groups almost entirely segregated by temporal sequence

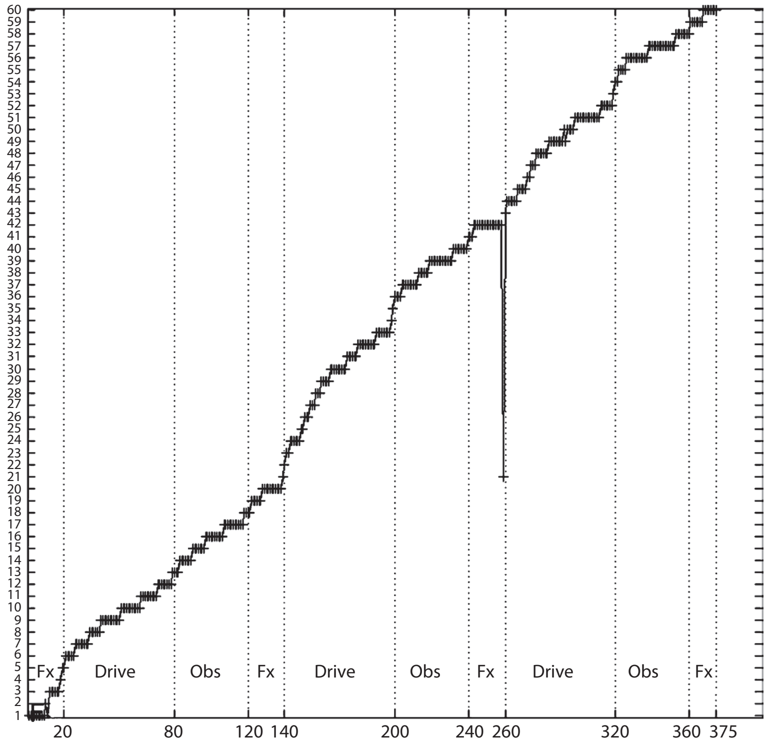

Figure 4. Cluster analysis of brainspace trajectory across sixty clusters. Temporal divisions dominate. At this grain, velocity through brainspace becomes visible as time spent within each cluster. Cluster jumps are more frequent at condition boundaries, for example

At this level, aspects of the flow of time become visible. Note the boundaries of the Drive conditions, at which the brain moves rapidly through distinct states. The overall "velocity," or the number of cluster jumps over time, varies with the different conditions. Brain activity seems to change most quickly at the boundaries of conditions. Within conditions, Drive and Observe clusters pass more quickly (compared to the expected cluster size with a uniform distribution). Here we can see the rapid jumping from state to state in the eventful Drive sequences. In contrast, time drags during the boring Fixation conditions. But looking at these episodes retrospectively changes how they seem. The many events of the Drive periods might make those sections seem longer in retrospect, in contrast to the Fixation sections, which in retrospect might seem shorter than their actual duration.

Considered together, then, as the view through brainspace zooms, both traditional phenomenology and the assumptions of cognitive neuroscience are accommodated. At the coarsest level, the three very different conditions of the experiment push the brain into three different distributed states, just as cognitive science sees it. But within each of those territories, brains wander about. Nonetheless, the wanderings keep time points distinct and ordered, just as the phenomenologists insist.

4. Temporality now!

Cluster analysis of brain space in this experiment reveals different facets of brain function at different magnifications. At a coarse grain, the brains analyzed here are categorical engines, switching from one global state to another as the task demands. But at a finer grain, within each global category, brains are time-keepers – notwithstanding the fact that subjects in this experiment were not required to notice the passing of time. Initially, therefore, we observe a correlate to structural, ubiquitous time.

Clocks also change with passing time, but clocks show the time without knowing the time. Clock time is not temporal awareness for the clock. Apparently, temporal information is encoded in state space, but this is not yet to establish that this information is encoded in the brain itself. What more is needed? Phenomenological temporality embodies a three way distinction between protention, primal impression, and retention. These are three co-occurring aspects of every state of consciousness. Accordingly, their brainy implementation must not only tag the current time, but encode past and future as well. With this requirement, temporality pulls the brain still farther from the stationary, task-clamped organ described in standard cognitive neuroscience. Each brain state not only encodes its current location in a temporal landscape, but also its past and future positions. It not only has a trajectory but somehow encodes that trajectory as well.

Trajectory through brainspace is at the heart of the proposed computational theory of experience of Fekete and Edelman (2011), and motivated in part by the phenomenology which is also central to this chapter:

> A model in which qualia are equated to instantaneous independent states would be hard pressed to explain various mundane cognitive phenomena, not the least of which is the experience of time. (p. 9)

Indeed, but the phenomenology does not merely imply that dynamical brain configurations evolve along a path over time but that instantaneous states

themselves represent this trajectory. That's the Husserlian bottom line, the vertical dimension of temporality that shadows the horizontal flow of time. This instantaneous synchronic representation of trajectory (which may be both gappy and inaccurate) is distinct from the diachronic trajectory itself.

What empirical evidence might be enlisted to confirm this temporal extension? Here we take an indirect route, inspired by work in machine learning and Support Vector Machines (SVMs) in particular (Cox & Savoy 2003; LaConte, Strother et al. 2005). SVMs learn to map a set of input patterns to a corresponding set of output patterns. Their learning is confirmed when an additional, unlearned, set of input patterns is correctly associated with target outputs. Successful association on the novel inputs implies that they must themselves encode information in their distributed activity to enable the SVM to project the desired target. This is the consideration that we will exploit to search for encoded time in patterns of brain activity.

Specifically, we can use SVMs to probe whether a pattern of brain activity at time t_n can predict activity at t_{n+i} where i ranges into the near future, and likewise retrodict/recover past states. In the driving experiment, can SVMs take the brain state at (for example) 22 seconds into the experiment, and correctly anticipate experimental conditions at 23, 24, 25... seconds, as well as correctly reconstruct past patterns at 21, 20, 19... seconds. As in other SVM paradigms, here we test accuracy with a "leave one out" cross validation strategy. During training, one vector of activations is omitted from the training set, and reserved for later testing. The accuracy of SVM training is then measured by the accuracy of classification of the omitted vector. This is repeated for all inputs to provide a general measure of SVM learning.

Figure 5 displays the landscape of protention and retention as revealed through the SVM train and test paradigm, showing percent accuracy of SVM time projection from each of 375 test points. The time unit is 1.5 seconds, the time taken for each image in the fMRI series. The figure thus summarizes the temporal landscape looking 15 seconds forward and back from each time point. The left half of the graph represents retentional information recovered from brain images at different lags from the current time, showing that accuracy is greater in recovering the immediate past, compared to the immediate future.[3]

---

[3] SVMs are binary classifiers, so the experimental data set was decomposed into three binary sets, one for each of the three experimental conditions, Fixate, Drive, and Observe. So, for example, Fixate conditions would be encoded as +1, and all others −1, etc. Furthermore, separate SVMs were trained and tested for each time offset, moving the +1/−1 boundary from 15 seconds behind the actual transition to 15 second ahead. The resulting scores are aggregates from these various sub-conditions.

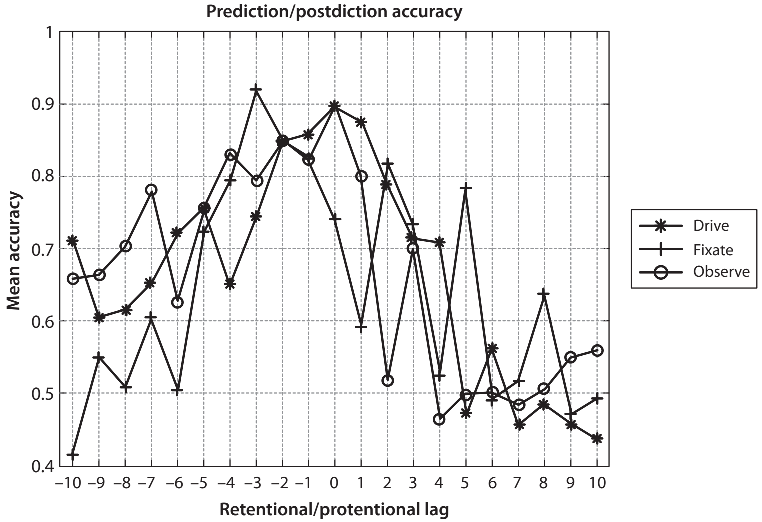

Figure 5. Embedded temporal information in each image in the drive experiment, mean of 375 times points in ten subjects. Accuracy is measured with a "leave one out" cross-validation strategy, measured on each image when it has been omitted from the training set. Because support vector machines are constrained to sort targets into binary categories, each experimental condition was tested separately. The graph implies that from each image, in general, projections are more than 50% accurate about six seconds into the future (4 time points on the graphic), and retentional information is accurate nine or more seconds into the past. (Chance performance is 50%.) At each time point, the brain appears to represent a temporal window of about fifteen seconds

5. Countdown to implementation

The evidence presented so far suggests that brains encode retentional and protentional information for several seconds and possibly longer, and that traces of this encoding are captured in fMRI image series. The impulse to localize this information is strong. But rather than follow this path, a final empirical foray explores a more general question about implementation: Is temporal information localized at all? At one extreme, one region of the brain may be the time-giver, a dedicated clock that can somehow be read by the SVM-fMRI combination. At the other extreme, time could be embedded in configurations of states, a distributed pattern of activation whose changes represent the evolving temporal landscape of this experiment. Phenomenology suggests that temporality is in every experience, suggesting a more distributed implementation. This is the suggestion to be tested here.

The test strategy is as follows: We successively remove each of the twenty independent components, and test the remaining nineteen using the methods of the previous section (Opper & Winther 2000). Performance degrades following some of these virtual lesions. The component whose absence most impairs performance is thus the component of greatest import in temporal representation. If performance dropped to chance levels, then we would have identified a component necessary to temporal encoding. But if temporal representation persists, even if somewhat impaired, we can repeat the process, now with the reduced set of nineteen components. That is, we can successively remove components one-by-one and test the remaining eighteen to discover which component contributes the most to pattern learning performance. This will be the second-most important component overall, and it too is set aside as the analysis continues, now lesioning each of the remaining eighteen components. And so forth, leading eventually to Figure 6, plotting the mean performance as the number of components is

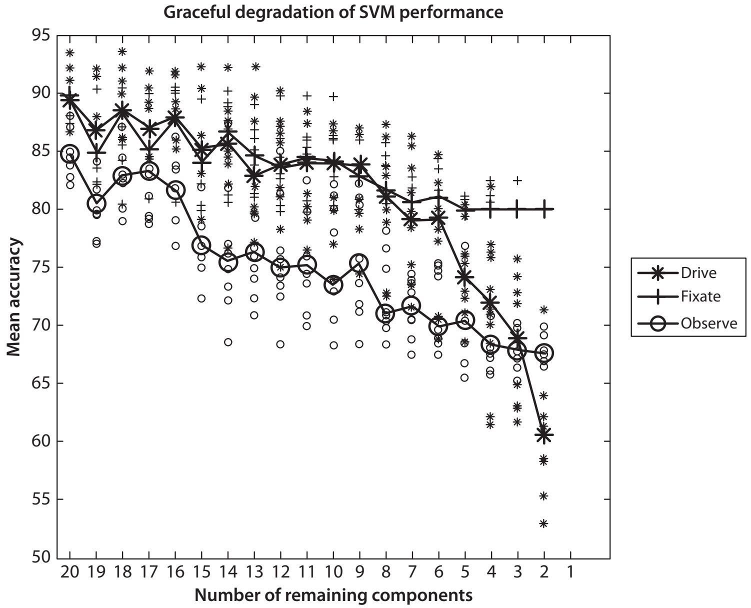

Figure 6. The effects of successive removal of the most effective components. For each subject, accuracy is tested with one component removed to determine the most important component (i.e. the components whose removal has the greatest impact on accuracy). That component is removed, and the process repeated on the reduced component set. Performance falls steadily toward chance levels in most subjects. Large symbols represent the median accuracy of the SVMs for 10 subjects. Small markers represent accuracy individually

decremented, removing the most important component with each decrement. The figure shows a graceful decline as approximately ten components are removed. Beyond ten, overall performance falls off more rapidly. Analysis using this "leave one out" strategy thus implies that temporal information is broadly distributed across regions of the brain. The encoding of protention and retention within the 30 second window considered here seems to depend on a varying pattern of activity rather than a localized clock.

6. Crossing the explanatory gap

The science of consciousness begins with correspondences between items in awareness and items in the brain, just as deciphering a text in an unknown language might begin by identifying synonyms between the known and the unknown lexicons. A large part of the study of a new language is just this, and likewise a great deal of consciousness research is a search for "neural correlates of consciousness." As the correlates accumulate the ties of brain to phenomenology diversify and skepticism about materialism becomes increasingly untenable. But a pile of correlates do not a theory of consciousness make, just as a lexicon is only one aspect of a language. A *theory* of consciousness is an edifice that shows how all the pieces fit together, comparable to an explicit grammar for a language. If the particulars of the neural correlates of consciousness are the first-order observations in the science of consciousness, then we must also look toward higher-order relationships as well.

Dynamical systems offer a useful framework for theorizing. State spaces afford a way to conceptualize higher order relationships that balances multivariate complexity against the limits of human imagination. Here the dynamical systems analogy structured an exploration of the higher-order neurophenomenology of time. In this one case study (with a mere ten subjects), we observed brainspace relationships that respected first-order distinctions (between the conditions of the experiment) while at the same time reflecting temporal distinctions between repetitions of the experimental conditions and within same-condition blocks. At this point, however, the analogy with translation falters. It's not enough to show that the dynamical states of the brain reflect the trajectories of conscious awareness. Consciousness is instantaneously temporal: each state of awareness includes an awareness of whence and whither, retention and protention. This seems to be more than just its dynamical causal tendencies, but rather an active presentation of temporal extension. Higher-order constraints govern the system and also appear in some form within states of consciousness.

To locate a synchronous representation of diachronic time required a more elaborate probe of the same brainstate data. Here the approach was indirect, using machine learning to seek the patterns across the brain that might embody the forward- and backward- gaze of temporal awareness. It was argued that if a support vector machine could successfully extract the experimental conditions both before and after the current time point, then the temporal information must already be in the functional brain image at that time point. Success in this learning paradigm leads naturally to an interest in its "neural correlate." In this case, we want to know which areas of the brain support the representation of time. The current chapter stops short of that answer, but does present evidence that broadly distributed patterns of activity are involved in the representation of temporal information.

Suppose, then, that further confirmations follow, based in different experiments and new methods. To what extent do reports like these inform a genuine theory of consciousness? Two general answers suggest themselves. First, it is essential that theories of consciousness address higher-order features of experience. Time is an example of a higher-order property, a structural aspect of all states of awareness. Temporality infuses everything else, and so every first-order neural correlate is also the vehicle of the higher-order structural awareness of time. Merleau-Ponty declared that "we are time," affirming the centrality of temporality to consciousness (Merleau-Ponty 1962). Temporal relationships provide a basis for many seemingly non-temporal properties. Space, for example, could be construed as a set of possible physical trajectories for objects of perception, such trajectories being represented in retention and protention. Objects are equally temporal, as Husserl himself described at length in 1907 (Husserl & Rojcewicz 1997).

More important, self-consciousness could be reanalyzed as a temporal trajectory as well. Self-consciousness and its problems are frequently taken to be the problem of consciousness overall. In my opinion, this is a mistake. Although all states of awareness are inflected by a point of view, they do not require a further reflective awareness of this point of view. In the phenomenological tradition, we encounter a distinction between consciousness "in itself" and consciousness "for itself" (As Sartre puts it: *en soi* and *pour soi*. (Sartre 1958)) Sartre, in accord with Husserl and Heidegger, affirms that much of conscious life is "in itself," that is, unaccompanied with self-awareness (for itself). Promoting self-awareness into an essential condition for experience leads to either an infinite regress or the implication that states of self-awareness (higher order thoughts) are not themselves conscious. On a more practical level Fekete and Edelman (2011) have argued that a synchronous awareness of the current state of the brain by that self-same brain is computationally infeasible. Temporality allows us to reconfigure self-awareness as a reflective access to retentional information about one's own prior states (both percepts and concepts), a modest and tractable self-awareness that better fits with the real phenomenology of self.

Thus, many aspects of conscious life can be reconstrued as an awareness of trajectory over time. A theory of consciousness that includes temporality is a worthy advance from the accumulation of first-order NCCs. But this brings us to the brink of another canyon of mystery: How does it work? The analyses in this chapter have been silent on the actual mechanisms of temporal awareness. Recurrent networks ("reentrant processing") appears repeatedly in proposed mechanisms of consciousness (Edelman 1993; Di Lollo, Enns et al. 2000; Clavagnier, Falchier et al. 2004; Lloyd 2004; Augustenborg 2010; Miller 2011). Such networks have the capacity for temporal representation. Indeed, they require it if they govern behaviors that extend over time. Unfortunately, functional MRI may be too coarse an instrument to probe these dynamics.

Which is more complex, the brain or consciousness? Both are vast, and it is no surprise that their union in the science of consciousness will be a long story indeed. But every story has its beginning. Among the proliferating beginnings by scientists and philosophers, we may find one or more with promise. Once upon a time....

Acknowledgements

Thanks to: Vince Calhoun and Godfrey Pearlson, who provided the study data sets, and to the Trinity Faculty Research Committee for ongoing research support. Independent Component Analysis was implemented with the ICA Toolbox (Calhoun, Adali et al. 2002). Support Vector Machine analysis was implemented with the OSU-SVM Toolbox (Ma, Zhao et al. 2002).

References

Augustenborg, C.C. (2010). The endogenous feedback network: A new approach to the comprehensive study of consciousness. *Conscious Cogn* **19**(2), 547–579.

Calhoun, V.D., Adali, T. et al. (2002). Independent component analysis of fMRI data in the complex domain. *Magn Reson Med* **48**(1), 180–192.

Calhoun, V.D., Carvalho, K. et al. (2005). Using virtual reality to study alcohol intoxication effects on the neural correlates of simulated driving. *Appl Psychophysiol Biofeedback* **30**(3), 285–306.

Calhoun, V.D., Pekar, J.J. et al. (2004). Alcohol intoxication effects on simulated driving: Exploring alcohol-dose effects on brain activation using functional MRI. *Neuropsychopharmacology* **29**(11), 2097–2017.

Carvalho, K.N., Pearlson, G.D. et al. (2006). Simulated driving and brain imaging: Combining behavior, brain activity, and virtual reality. *CNS Spectr* **11**(1), 52–62.

Clavagnier, S., Falchier, A. et al. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn Affect Behav Neurosci* **4**(2): 117–126.

Cox, D.D. & Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19**(2 Pt 1), 261–270.

Di Lollo, V., Enns, J.T. et al. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *J Exp Psychol Gen* **129**(4), 481–507.

Edelman, G.M. (1993). Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron* **10**(2), 115–125.

Edelman, S. (2008). *Computing the mind: How the mind really works*. New York: Oxford University Press.

Fekete, T. & S. Edelman (2011). Towards a computational theory of experience. *Consciousness and Cognition 20*, 807–827.

Husserl, E. (1966/1928). *Zur Phänomenologie des inneren Zeitbewusstseins (Phenomenology of Inner Time Consciousness)*. The Hague: Martinus Nijhoff.

Husserl, E. & Rojcewicz, R. (1997). *Thing and space: Lectures of 1907*. Dordrecht; Boston, Kluwer Academic Publishers.

LaConte, S., Strother, S. et al. (2005). Support vector machines for temporal classification of block design fMRI data. *Neuroimage* **26**(2), 317–329.

Lloyd, D.E. (2004). *Radiant cool: A novel theory of consciousness*. Cambridge, Mass.: MIT Press.

Logothetis, N.K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci* **357**(1424): 1003–1037.

Ma, J., Zhao, Y. et al. (2002). OSU-SVM Toolbox for Matlab.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge and Paul.

Miller, G. (2011). Neuroscience. Feedback from frontal cortex may be a signature of consciousness. *Science* **332**(6031), 779.

Opper, M. & Winther, O. (2000). Gaussian processes and SVM: Mean field results and leave-one-out. *Advances in large margin classifiers*. Cambridge, MA: MIT Press, 43–65.

Sartre, J.-P. (1958). *Being and nothingness*. London, Philosophical Library.

Zahavi, D. (1999). *Self-awareness and alterity*. Evanston, IL, Northwestern University Press.