Being in Time (extended abstract)

Shimon Edelman[1]
Tomer Fekete[2]

Any theory that purports to explain how phenomenal experience arises in the system in question (such as the brain) must do so in terms that are meaningful within the system itself. Simply put, because my experience of the world cannot be up to an outside observer, it cannot be explained by positing entities, such as whole-brain state variables, if those can only be known from the outside. This is the sense in which theories of experience must be *intrinsic* (Fekete, 2009; Fekete and Edelman, 2011).

Theories of phenomenal experience that are computational (Fekete and Edelman, 2011) are subject to another obvious constraint. The computations to which experience is reduced in such theories must be *tractable*, given the physical limitations of the implementing system (the brain). For instance, if a transition between two experiential states is to be explained by a computation performed on a certain variable, the brain must be capable of completing that computation within an appropriate time frame.

A popular theoretical construct that runs afoul of the above considerations is that of an instantaneous state of the brain (or of a part of the brain). A mind implemented in a brain is a distributed entity, because of the generally non-negligible delays in the propagation of signals within the brain.[3] If it takes a finite time for neurons that form a network to communicate, in what sense is an instantaneous state of the entire network "there" for the network itself, rather than for an omniscient zero-lag outside observer? Casting theories in terms of instantaneous brain states is also problematic from the functional standpoint because such states, taken in isolation, fail to constrain the system's dynamics, and therefore lack intrinsic causal powers.[4]

Only with time, as the system's far-flung components interact, does its dynamics become apparent — most importantly, to itself, that is, intrinsically. Not surprisingly, the passage of time is also critical for phenomenal experience. Because all the components that can in principle contribute to the system's dynamics participate in shaping its trajectories through the state space and in imposing structure on this space (ruling some classes of trajectories in and others out), phenomenal experience is holistic: it emerges from the dynamics of the entire brain (Fekete and Edelman, 2011). But as we just noted, because signaling within any network of neurons cannot be instantaneous, the holism of experience implies that it must be inherently temporally extended.

It is tempting to assume that this conclusion implies merely that there must be a delay between exposure to a stimulus and its experience, as expressed by the view that "the phenomenal experience emerges when all relevant neurons in a network are informed about their own population state" (Malach, 2007). This view is, however, troublesome in light of the computational difficulties associated with the problem of attaining agreement in asynchronous distributed systems (Pease, Shostak, and Lamport, 1980; Lamport, Shostak, and Pease, 1982). In realistic distributed systems, of which the brain is but one example, failures such as faulty elements and unreliable communication links conspire to make agreement (e.g., about the value of a global variable) hard to achieve.

[1]To whom correspondence should be addressed. Department of Psychology, Cornell University, Ithaca, NY 14853-7601, USA. mailto:edelman@cornell.edu; http://kybele.psych.cornell.edu/~edelman

[2]Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794-5281, USA.

[3]In fact, insofar as minds are embodied and situated, the dynamical processes that constitute them "spill over" from the brain into the body and the rest of the environment (Thompson and Varela, 2001; Spivey, 2006; Edelman, 2008).

[4]Fekete and Edelman (2011) call this predicament of standard neuroscientific accounts of representation the "silent units" problem.

A further complication is introduced by asynchrony: intuitively, communication protocols cannot distinguish failed elements from delayed messages (Kapron, Kempe, King, Saia, and Sanwalan, 2010). Indeed, with just a single faulty processor, no deterministic asynchronous agreement is possible (Fischer, Lynch, and Paterson, 1983). Even with randomization, the best known algorithm, which succeeds with probability 1, has an expected running time that is exponential in the number of elements (Ben-Or, 1983). The computational tractability analysis thus indicates that the wait until all the neurons in a network have reached an agreement may be long indeed.

These results need not, however, rule out the possibility of a viable computational theory of phenomenal experience. Instead, they highlight the conceptual shortcoming of the default scenario within which brain dynamics has been studied, in which a stimulus is delivered to the subject, who after a certain delay experiences it, while producing or withholding an overt response. This Pavlovian staple of 20th century psychology rests on the tacit assumption that over and above priming, the brain's activity prior to the onset of the stimulus is of little consequence to its "processing." Tellingly, the cover of the 1961 Science Editions paperback printing of Donald Hebb's *The Organization of Behavior* (1949) bears the subtitle "Stimulus and response — and what occurs in the brain in the interval between them."

It is now becoming increasingly clear that this assumption is wrong (Spivey, 2006, p.47).[5] Given that external events modulate ongoing experience rather than generating it *ab vacuo*, it makes little sense to phrase questions about brain dynamics in terms of the "delay" between the stimulus and its experience, which, moreover, is more profitably viewed not as settling into an attractor and waiting for the next stimulus, but rather as "chaotic itinerancy" through a series of saddle points in the state space (van Leeuwen, 2007). On the level of dynamics, therefore, one should inquire about the time scales on which experience unfolds (Rudrauf, Lutz, Cosmelli, Lachaux, and le van Quyen, 2003, p.38), so as to determine the characteristic time constants that describe the changes in experience brought about by stimulus onset (e.g., Ito, Nikolaev, and van Leeuwen, 2007). Crucially, such analysis must be complemented by a study of the structure of the system's state space (Fekete, Pitowsky, Grinvald, and Omer, 2009; Fekete, 2009), which needs to be related to the structure of the experience that arises from it (Fekete and Edelman, 2011).

While considering the brain's dynamics, it is easy to slip into language that glosses over the fundamental constraint on any theory of experience, namely, that it should be intrinsic. To enforce this constraint, we need to explain in what sense and by what means the system's dynamics is known to the system itself, and to do so without running into the tractability problem associated with asynchronous distributed agreement about global variables. A possible conceptual solution here is to consider temporally extended *diagnosis* of the system's state-space structure by a subset of its elements.[6]

The mathematical basis for this idea is a theorem proved by Takens (1981), according to which a series of measurements of a single variable within a dynamical system becomes with time diagnostic with respect to qualitative structure and quantitative parameters of the system at large (Quyen, 2003).[7] We note that methods derived from the Takens Theorem receive much attention in empirical studies of complex dynamical systems (including ecosystems, which, like the brain on the relevant time scale, are also spatially distributed; Wilson and Rand, 1997), and that such methods become more effective if multiple "observing" variables are used instead of one (Deyle and Sugihara, 2011).

These results imply that a clique of neurons can diagnose the dynamics of the rest of the brain, given

[5]Cf. Freeman (2007): "In an awake subject engaged with the environment those driven cortical neurons, even before stimulation, are already self-organized and active through intracortical and extracortical feedback loops."

[6]Cf. Quine (1985): "Mental states, construed as states of nerves, are like diseases. A disease may be diagnosed in the light of observable signs though the guilty germ be still unknown to science."

[7]For a cautionary note on the conditions for the applicability of Takens' Theorem, see (Smith, 1997).

enough time. Moreover, if the dynamics in question is not too high-dimensional (a plausible assumption, for a number of reasons; Skarda and Freeman, 1987; Freeman, 2007), the number of samples required for a good approximation is manageable (Deyle and Sugihara, 2011). In a sense, then, the proposed approach makes Malach's (2007) idea of self-observation work by viewing neurons not as processors that communicate by message-passing and that must reach a global agreement, but as ongoing dynamical "observers" that are part of the observed system.

To summarize, we argue that the sketch of a computational theory of phenomenal experience offered in (Fekete and Edelman, 2011) passes the crucial test of being statable exclusively in intrinsic terms, while conforming *prima facie* to the implementational constraints stemming from the mind's embodiment as a distributed dynamical system. Because experience is massively endogenous and continuous, it must be seen not as convergence to an attractor, but rather as the unfolding of a metastable trajectory through a properly structured space of possible trajectories, as defined by the brain's dynamics. The self-observation idea and self-diagnosis theorems suggest the sense in which parts of the brain may know, on an ongoing basis, what it is up to, and point to the means whereby they can do it.

References

Ben-Or, M. (1983). Another advantage of free choice: Completely asynchronous agreement prototols. In *Proceedings of the Second Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 2730.

Deyle, E. R. and G. Sugihara (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS ONE 6*, e18295.

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

Fekete, F., T. Pitowsky, A. Grinvald, and D. B. Omer (2009). Arousal increases the representational capacity of cortical tissue. *Journal of Computational Neuroscience 27*, 211–227.

Fekete, T. (2009). Representational systems. *Minds and Machines 20*, 69–101.

Fekete, T. and S. Edelman (2011). Towards a computational theory of experience. *Consciousness and Cognition*. To appear.

Fischer, M. J., N. A. Lynch, and M. Paterson (1983). Impossibility of distributed consensus with one faulty process. In *Proceedings of the Second ACM SIGACT-SIGMOD Symposium on Principles of Database System (PODS)*, pp. 17.

Freeman, W. J. (2007). Indirect biological measures of consciousness from field studies of brains as dynamical systems. *Neural Networks 20*, 10211031.

Hebb, D. O. (1949). *The organization of behavior*. Wiley.

Ito, J., A. R. Nikolaev, and C. van Leeuwen (2007). Dynamics of spontaneous transitions between global brain states. *Human Brain Mapping 28*, 904–913.

Kapron, B. M., D. Kempe, V. King, J. Saia, and V. Sanwalan (2010). Fast asynchronous Byzantine agreement and leader election with full information. *ACM Transactions on Algorithms 6*.

Lamport, L., R. Shostak, and M. Pease (1982). The Byzantine generals problem. *ACM Trans. on Prog. Lang. and Syst. 4*, 382–401.

Malach, R. (2007). The measurement problem in consciousness research. *Behavioral and Brain Sciences 30*, 516–517.

Pease, M., R. Shostak, and L. Lamport (1980). Reaching agreement in the presence of faults. *J. ACM 27*, 228–234.

Quine, W. V. (1985). States of mind. *The Journal of Philosophy 82*, 5–8.

Quyen, M. L. V. (2003). Disentangling the dynamic core: a research program for a neurodynamics at the large-scale. *Biol. Res. [online] 36*(1), 67–88.

Rudrauf, D., A. Lutz, D. Cosmelli, J.-F. Lachaux, and M. le van Quyen (2003). From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. *Biological Research 36*, 27–65.

Skarda, C. and W. J. Freeman (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences 10*, 161–195.

Smith, L. (1997). The maintenance of uncertainty. In *Proceedings of the International School of Physics "Enrico Fermi"*, Volume Course CXXXIII, Berlin, pp. 177–246. Springer.

Spivey, M. J. (2006). *The continuity of mind*. New York: Oxford University Press.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand and L.-S. Young (Eds.), *Dynamical systems and turbulence*, Lecture Notes in Mathematics, pp. 366–381. Berlin: Springer.

Thompson, E. and F. Varela (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences 5*, 418–425.

van Leeuwen, C. (2007). What needs to emerge to make you conscious? *Journal of Consciousness Studies 14*, 115–136.

Wilson, H. B. and D. A. Rand (1997). Reconstructing the dynamics of unobserved variables in spatially extended systems. *Proceedings of the Royal Society of London B 264*, 625–630.