Dynamical emergence of phenomenal consciousness: an outline of a theory

Roy Moyal and Shimon Edelman

Cornell University, Ithaca NY 14853, USA
rm875@cornell.edu
se37@cornell.edu

Abstract. We outline a computational theory of phenomenal conscious experience. Our Dynamical Emergence Theory (DET) aims to explain the structure, the quantity, and the quality of phenomenal experience in terms of trajectories through the space of the system's emergent metastable macrostates and their intrinsic (observer-independent) topology and geometry. Section 1 discusses some of the constraints that must be satisfied by any computational theory of consciousness. Section 2 is an overview of the concepts of coarse-graining and emergence, which are central to the understanding of representation and computation in dynamical systems. DET is then stated in section 3. Section 4 concludes the paper with some predictions for experimental studies and suggestions for future empirical and theoretical investigations.

Keywords: dynamics, emergence, macrostates

1 Explaining phenomenal experience: axioms and constraints

In this brief paper, we offer a condensed overview of Dynamical Emergence Theory (DET), which equates some aspects of the phenomenal experience [17] of a conscious system with the geometric and topological structure of its dynamics (a full account appears in [14]). Our notion of phenomenal awareness corresponds to "primary-process consciousness" [29, p.31], which does not necessarily require the presence of a self [23] or of any higher-order awareness of awareness.

We hold that, insofar as the mind is fundamentally computational, so is consciousness (for some arguments supporting this claim, see [10, 11, 17]). Our construal of computation is rather broad and includes not only symbol manipulation by discrete automata but also continuous dynamics, as stated in [11] and further explicated and motivated in [17]. Any computation is an organizational invariant [3] in that it depends on the relationships that govern the patterns of transitions among the system's states rather than on the physical substrate underlying those states. In other words, computation is multiply realizable, insofar as the identity of an instance of computation resides not in the physical identities of the elements of the underlying system (such as states and transitions in automata, or neurons), but rather in their organization (which must be *intrinsic* to the system; see below). Further, because the molecular composition of an organism changes continually, its identity (and its causal contributions on higher levels of organization, such as that of social groups) resides not in the identities of its constituent molecules, but rather in the pattern of their interactions. If conscious phenomenal experience is indeed a kind of computation, this organizational principle, which affords a separation of levels, must apply to it too.

An axiomatic basis for consciousness theories has been introduced and motivated by Fekete and Edelman ([17, 12]; for a somewhat different set of axioms, see [27]). While all the requirements listed there apply equally, here we single out two of them (referred to below as Inherence and Structure) and add a third one (Effectiveness). Together, these three requirements enable a computational approach to phenomenality by placing strictures on physical substrates and computational realizers of experience:

Inherence. Because the phenomenal experience of a system is necessarily intrinsic to it, rather than a matter of outside interpretation or attribution, so must be any characteristics that define the physical *substrate of experience* (PSE), as well as those that define its *computational realizer* (CRE).

Structure. Consciousness can only be explained in terms of organizationally (that is, relationally) defined states and transitions, whose intrinsic structure (CRE) must match the psychological structure of experience. Most importantly, this intrinsic structure must reflect discernment among qualia.

Effectiveness. The structure of CRE must be in some well-defined sense intrinsically causally effective. Intuitively, we take this to mean that the states and transitions comprising CRE must be predictively effective, as discussed in section 2 below.

The physical substrate of experience (PSE) is *collective dynamics*. A set of elements with no collective dynamics—one in which not all elements causally interact—is not a "system" in the relevant (intrinsic) sense. If a system does possess collective dynamics on two or more distinct levels (in the same sense that subatomic and molecular dynamics are distinct), then each such level is a candidate PSE.

The computational realizer of experience (CRE) is an intrinsically structured emergent pattern of states and transitions over PSE. If the collective dynamics of PSE is unstructured—as, for instance, in the case of the molecules of gas in an enclosure, whose joint trajectory space has no interesting or meaningful (in the sense of [1]) intrinsic structure—then the system is incapable of experience [17]. If, however, the pattern that emerges from the collective dynamics of a system is properly intrinsically structured (most importantly, if it implements intrinsic discernment; [17, p.807]) and is causally effective—as in the case of a coarse-graining [33, 34] of the PSE dynamics, as defined and discussed next—then this pattern, the CRE, satisfies the Structure and Effectiveness requirements.

2 Coarse-graining and emergent macrostates

Crutchfield [5, p.12] describes emergence intuitively as self-organization [34] of novel structure “over time”—that is, dynamically. CRE-level patterns can emerge from a physical substrate through coarse-graining [33, 20], which involves aggregating PSE states into equivalence classes or “macrostates” on the basis of certain statistics of their properties. For the resulting macrostates to meet the Inherence requirement, the choice of properties and of the statistical criteria must be intrinsically self-consistent. Further, for the macrostates to meet the Effectiveness requirement, they must be intrinsically meaningful (that is, their effects should not be a matter of outside interpretation). The application of these requirements leads precisely to contextual emergence: a process in which “the neurodynamics is used to construct statistical neural states which are in one-to-one correspondence with properly defined mental states. Their dynamics is then topologically equivalent with the neurodynamics” [19, p.176].

We adopt as a working hypothesis a set of systematic and formal [19] contextual constraints, which are based on the work of Shalizi ([32, 33]). First, this approach requires that the macrostates arise out of a generating partition of the original domain—one in which the boundaries between macrostates are preserved over time under the system’s dynamics, which ensures that the macrostates are intrinsically self-consistent. Second, it requires that macrostates be Markovian, and thus “states which predict their own future” in a mathematically explicit and “provably optimal” fashion [33, p.1].

These requirements together ensure that the macrostates are stable and intrinsically meaningful. From outside the system, one can resort to the macrostate discovery procedure of [33], which is intended to serve as a tool for the scientist seeking a useful high-level characterization of the system, and which necessarily begins with an externally made choice of observations. The initial approximation is then successively refined until an optimal set of macrostates is arrived at. Specifically, [33, p.9] “define a relation of ‘emergence’ between two sets of causal variables if (1) one is a coarse-graining of the other and (2) the coarse-grained variables can be predicted more efficiently.” Importantly, the resulting recasting of the system’s dynamics in terms of the macrostates is not merely an observer-relative description of the system.

To complete the characterization of emergent macrostates, we must consider their causal role. Our preferred approach to this issue is the idea of *proportionate causation*, as introduced in [40] and discussed at length by Harbecke and Atmanspacher [19], who make a careful distinction between counterfactual sufficiency and necessity and use it to support the idea that distinct and parallel causal contributions of different levels are possible and should be seen as complementing rather than excluding each other. A key principle in their framework is that the process of emergence is essentially dynamical: “For mental states [macrostates] to be causally efficacious, they must be dynamically stable” [19] (as per the notion of generating partition). An upshot of this principle is that it takes time for an emergent set of macrostates to be causally efficacious by being stable—just as posited by DET and its conceptual predecessor, Geometric Theory [17].

3 Quantifying experience

With the above conceptions of PSE and CRE in mind, we define *phenomenal experience* as a system's trajectory through its space of coarse-grain macrostates (CRE) that emerge from a physical substrate endowed with structured (in the sense described below) intrinsic collective dynamics (PSE).

This definition closely follows that of [17], except that here we allow for the relevant (CRE) dynamics—states and trajectories—to emerge from (rather than be necessarily identical to) the dynamics of the physical substrate (PSE). The CRE structure that emerges via coarse-graining inherits from the underlying dynamics some of the structure of its trajectory space; the coarse-grained structure affords fewer distinctions among trajectories (which on this level are sequences of macrostates), but all the distinctions that it does support are still intrinsic in the requisite sense, simply because the macrostates are a partition of the original set of states.

Our basic identification of experience with the dynamics of the system's trajectories leads to two quantitative measures of consciousness. One measure, a scalar, should quantify the *amount of experience* (AoE) that the system is having, by tracking the gradations in representational capacity (e.g., from coma to full alertness; cf. [15,16]). It vanishes for systems that are structurally incapable of meeting the criteria for PSE or CRE, as well as for systems that fail to do so contingently (as in the case of dreamless sleep or coma). The other measure should quantify the *nature of experience* (NoE) as it unfolds in time, and as such should be structured to a degree and in a manner that match the phenomenal structure of the ongoing experience.

We propose to identify AoE with the topological complexity of the individual state-space trajectories of the system. Intuitively, a topologically complex trajectory, as it unfolds in time, would correspond to a class of rich ongoing experiences. Note that the actual precise shape of a trajectory that belongs to such a class, which is what NoE is intended to capture, is constrained, but not uniquely determined, by its topological complexity. Thus, the amount of experience (AoE) corresponds to the trajectory's topological complexity, while the nature of experience (NoE), with all its idiosyncratic and likely ineffable nuances, corresponds to the trajectory's geometry.

Importantly, both AoE and NoE are empirical measures, not logical consequences of our definitions of PSE or CRE. Their values must be estimated using a sliding window and as such they depend on the size of the window and other measurement parameters. They are expected to fluctuate as the clique of the system's elements that affect the current stretch of the trajectory changes dynamically. Despite these dependencies and fluctuations, AoE and NoE should be useful because they offer insight into the intrinsic topology and geometry of the system. This is true even when the estimation is based on measurements that are rather crude in comparison with the system's dynamics, as is the case with EEG data and brain activity. In any case, given the empirical nature of AoE and NoE, they should be seen as relative, not absolute, measures. Rather than attempting to interpret the values they yield for specific states, we propose to use them to draw comparisons among several such states, each corresponding to some well-defined and preferably controlled condition (as suggested next).

4 Future work and conclusions

The question to which DET provides a tentative answer ("what does phenomenal awareness consist in?") is rarely explicitly engaged with by theories of consciousness. A detailed discussion of the relationships between DET and other theories that do—notably, Integrated Information Theory [39, 27]—can be found elsewhere [14]. In the little space that remains here, we briefly mention some directions for empirical work motivated by DET.

First, we propose to estimate, for a variety of EEG (and perhaps fast fMRI [7, 18]) data, both the representational capacity (RC) as defined in [16] and the DET AoE and NoE measures. The measures should be compared across changes in stimulation: rest, "simple" stimuli such as undifferentiated fields of uniform color, and composite stimuli such as shapes or scenes of increasing complexity. RC is expected to reflect arousal [15] and the difference between rest and a simple stimulus, and so is AoE. In comparison, differences in NoE should covary with changes across stimuli more closely than differences in RC or AoE (note that as NoE is not a scalar, its interpretation requires a metric to be defined over the structures resulting from its estimation).

Second, we propose to look for differences in AoE and NoE between a reference state and a reportable perception state, identified using a staircase procedure. Across trials, changes in AoE and NoE between states of unawareness and awareness (obtained for the same physical stimulus, e.g., by making use of the hysteresis in increasing vs. decreasing contrast) are expected. Studying such changes harks back to the classical concept of the neural correlates of consciousness [4]. Importantly, however, DET predicts that the best NCC (in the explanatory or predictive sense, similar to that of IIT [27]) will be found at the level of emergent macrostates, rather than at the "micro" level of, e.g., the activities of individual neurons.

Third, we predict that the AoE corresponding to a particular simple stimulus should be similar across such stimuli, but not necessarily across individuals; the NoE corresponding to a particular stimulus need not be similar either across such stimuli or across individuals. For complex stimuli that form a controlled pattern in the design space, the configuration formed by NoE structures in the NoE similarity space should reflect the design pattern (this corresponds to the notion of second-order isomorphism [35, 36] between representation spaces and the world [6, 13, 9, 28]).

Fourth, we propose to rank multiple versions of candidate macrostate dynamics, derived empirically by a variety of algorithms and from a variety of sources (such as EEG, fMRI, and perhaps invasive electrophysiological measures), by their predictiveness. We expect that the criterion of maximal predictiveness will reveal similar patterns of macrostates and transitions (dynamics), regardless of the source of the signal, given that the states of awareness and complex stimuli are controlled as suggested above.

To recapitulate, DET offers a computational framework for explaining basic phenomenal experience, of which all sentient beings (natural and perhaps eventually artificial) are capable. Minds are, as Minsky [26] quipped, what brains do; following Sperry [37] and others, DET posits that some such brain doings—specifically, those that are characterized by properly structured intrinsic dynamics—amount to feelings. These include, first and foremost, basic consciousness, or the feeling of being aware and awake, as well as the various qualia or phenomenal discernments, which the aware (but not necessarily self-aware) system can resolve. By giving these fundamental concepts an explicit computational interpretation, DET complements existing theories, such as the Global Workspace Theory [2, 8], which focuses on information processing at the expense of phenomenology, and the Information Integration Theory [39, 27], which assumes the kind of computational substrate that DET attempts to explain. In identifying phenomenality with certain dynamical properties of system trajectories, DET follows the approach of [17], which it modifies and extends. The kind of computational account of phenomenality offered by DET is an essential component of any comprehensive theory of consciousness, which for humans would also include the phenomenal self [23, 24] and its brain basis [21, 22].

References

1. Atmanspacher, H.: On macrostates in complex multi-scale systems. Entropy 18, 426 (2016).
2. Baars, B.J.: Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Progress in Brain Research 150, 45–53 (2005).
3. Chalmers, D.J.: A computational foundation for the study of cognition (1994).
4. Crick, F., Koch, C.: Towards a neurobiological theory of consciousness. Seminars in the Neurosciences 2, 263–275 (1990).
5. Crutchfield, J.P.: The calculi of emergence: computation, dynamics, and induction. Physica D 75, 11–54 (1994).
6. Cutzu, F., Edelman, S.: Faithful representation of similarities among three-dimensional shapes in human vision. Proceedings of the National Academy of Science 93, 12046–12050 (1996).
7. Davis, T., Poldrack, R.A.: Measuring neural representations with fMRI: practices and pitfalls. Annals of the New York Academy of Sciences 1296, 108–134 (2013).
8. Dehaene, S., King, L.C.J.R., Marti, S.: Toward a computational theory of conscious processing. Current Opinion in Neurobiology 25, 76–84 (2014).
9. Edelman, S.: Representation is representation of similarity. Behavioral and Brain Sciences 21, 449–498 (1998).
10. Edelman, S.: Computing the mind: how the mind really works. Oxford University Press, New York, NY (2008).
11. Edelman, S.: On the nature of minds, or: Truth and consequences. Journal of Experimental and Theoretical AI 20, 181–196 (2008).
12. Edelman, S., Fekete, T.: Being in time. In: Edelman, S., Fekete, T., Zach, N. (eds.) Being in Time: Dynamical Models of Phenomenal Experience, pp. 81–94. John Benjamins (2012).
13. Edelman, S., Grill-Spector, K., Kushnir, T., Malach, R.: Towards direct visualization of the internal shape representation space by fMRI. Psychobiology 26, 309–321 (1998).
14. Edelman, S., Moyal, R., Fekete, T.: Dynamical Emergence Theory (DET): a computational account of phenomenal consciousness. – –, – (2019), submitted.

15. Fekete, F., Pitowsky, T., Grinvald, A., Omer, D.B.: Arousal increases the representational capacity of cortical tissue. Journal of Computational Neuroscience 27, 211–227 (2009).
16. Fekete, T.: Representational systems. Minds and Machines 20, 69–101 (2010).
17. Fekete, T., Edelman, S.: Towards a computational theory of experience. Consciousness and Cognition 20, 807–827 (2011).
18. Grill-Spector, K., Malach, R.: fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologica 107, 293–321 (2001).
19. Harbecke, J., Atmanspacher, H.: Horizontal and vertical determination of mental and neural states. Journal of Theoretical and Philosophical Psychology 32, 161–179 (2012).
20. Hoel, E.P., Albantakis, L., Marshall, W., Tononi, G.: Can the macro beat the micro? Integrated information across spatiotemporal scales. Neuroscience of Consciousness 1, 1–13 (2016).
21. Merker, B.: From probabilities to percepts: A subcortical 'global best estimate buffer' as locus of phenomenal experience. In Edelman, S., Fekete, T., Zach, N. (eds.) Being in Time: Dynamical Models of Phenomenal Experience, pp. 37–80. John Benjamins (2012).
22. Merker, B.: The efference cascade, consciousness, and its self: naturalizing the first-person pivot of action control. Frontiers in Psychology 4(501), 1–20 (2013).
23. Metzinger, T.: Being No One: The Self-Model Theory of Subjectivity. MIT Press, Cambridge, MA (2003).
24. Metzinger, T.: The subjectivity of subjective experience: A representationalist analysis of the first-person perspective. Networks 3-4, 33–64 (2004).
25. Metzinger, T.: Splendor and misery of self-models: Conceptual and empirical issues regarding consciousness and self-consciousness. ALIUS Bulletin 1(2), 53–73 (2018), interviewed by J. Limanowski and R. Milliere.
26. Minsky, M.: The Society of Mind. Simon and Schuster, New York (1985).
27. Oizumi, M., Albantakis, L., Tononi, G.: From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. PLoS Computational Biology 10(5), e1003588 (2014).
28. Op de Beeck, H., Wagemans, J., Vogels, R.: Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nature Neuroscience 4, 1244–1252 (2001).
29. Panksepp, J.: Affective consciousness: Core emotional feelings in animals and humans. Consciousness and Cognition 14, 30–80 (2005).
30. Rosenthal, D.: A theory of consciousness. Research Group on Mind and Brain, ZiF Report No. 40, University of Bielefeld (1990).
31. Ross, D., Spurrett, D.: What to say to a skeptical metaphysician: A defense manual for cognitive and behavioral scientists. Behavioral and Brain Sciences 27, 603–647 (2004).
32. Shalizi, C.R.: Causal architecture, complexity and self-organization in time. Ph.D. thesis, University of Wisconsin, Madison, WI (2001).
33. Shalizi, C.R., Moore, C.: What is a macrostate? (2003), arXiv:cond-mat/0303625
34. Shalizi, C.R., Shalizi, K.L., Haslinger, R.: Quantifying self-organization with optimal predictors. Physical Review Letters 93, 118701–1 – 118701–4 (2004).
35. Shepard, R.N.: Cognitive psychology: A review of the book by U. Neisser. Amer. J. Psychol. 81, 285–289 (1968).
36. Shepard, R.N., Chipman, S.: Second-order isomorphism of internal representations: Shapes of states. Cognitive Psychology 1, 1–17 (1970).
37. Sperry, R.W.: A modified concept of consciousness. Psychological Review 76, 532–536 (1969).
38. Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.S. (eds.) Dynamical systems and turbulence, pp. 366–381. Lecture Notes in Mathematics, Springer, Berlin (1981).
39. Tononi, G.: Consciousness as integrated information: a provisional manifesto. Biol. Bull. 215, 216–242 (2008).
40. Yablo, S.: Mental causation. The Philosophical Review 101, 245–280 (1992).