COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13272

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13272

This paper part of the “Progress & Puzzles of Cognitive Science” letter series.

Consciousness, Exascale Computational Power, Probabilistic Outcomes, and Energetic Efficiency

Elizabeth A. Stoll

Western Institute for Advanced Study, Denver, CO

Received 18 October 2022; received in revised form 24 February 2023; accepted 4 March 2023

Abstract

A central problem in the cognitive sciences is identifying the link between consciousness and neural computation. The key features of consciousness—including the emergence of representative information content and the initiation of volitional action—are correlated with neural activity in the cerebral cortex, but not computational processes in spinal reflex circuits or classical computing architecture. To take a new approach toward considering the problem of consciousness, it may be worth re-examining some outstanding puzzles in neuroscience, focusing on differences between the cerebral cortex and spinal reflex circuits. First, the mammalian cerebral cortex exhibits exascale computational power, a feature that is not strictly correlated with the number of binary computational units; second, individual computational units engage in noisy coding, allowing random electrical events to gate signaling outcomes; third, this noisy coding results in the synchronous firing of statistically random populations of cells across the neural network, at a range of nested frequencies; fourth, the system grows into a more ordered state over time, as it encodes the predictive value gained through observation; and finally, the cerebral cortex is extraordinarily energy efficient, with very little free energy lost to entropy during the work of information processing. Here, I argue that each of these five key features suggest the mammalian brain engages in probabilistic computation. Indeed, by modeling the physical mechanisms of probabilistic computation, we may find a better way to explain the unique emergent features arising from cortical neural networks.

Correspondence should be sent to Elizabeth A. Stoll, Theoretical Research Program, Western Institute for Advanced Study, 1312 Seventeenth St Ste 745, Denver, CO 80202–1508, USA. E-mail: stoll@westerninstitute.org

Keywords: Neural computation; Probabilistic coding; Noisy coding; Thermodynamic computation; Energy-efficient computation; Perception; Agency; Nondeterministic computation

---

A central problem in the cognitive sciences is identifying the link between consciousness and neural computation. The key features of consciousness—including the emergence of representative information content and the initiation of volitional action—are correlated with neural activity in the cerebral cortex, but not computational processes in spinal reflex circuits or classical computing architecture. To take a new approach toward considering the problem of consciousness, it may be worth re-examining some outstanding puzzles in neuroscience, focusing on differences between the cerebral cortex and spinal reflex circuits. First, the mammalian cerebral cortex exhibits exascale computational power, a feature that is not strictly correlated with the number of binary computational units; second, individual computational units engage in noisy coding, allowing random electrical events to gate signaling outcomes; third, this noisy coding results in the synchronous firing of statistically random populations of cells across the neural network, at a range of nested frequencies; fourth, the system grows into a more ordered state over time, as it encodes the predictive value gained through observation; and finally, the cerebral cortex is extraordinarily energy efficient, with very little free energy lost to entropy during the work of information processing. Here, I argue that each of these five key features suggest the mammalian brain engages in probabilistic computation. Indeed, by modeling the physical mechanisms of probabilistic computation, we may find a better way to explain the unique emergent features arising from cortical neural networks.

A central problem in the field of cognitive science is the exact nature of the relationship between consciousness and neural computation. Current methods have not been able to crack this problem, so a new approach is needed.

Two major features of consciousness are the existence of qualitative perceptual experience and the initiation of voluntary behavior. The content of perceptual experience is paired with the encoding of sensory stimuli in modality-specific regions of the cerebral cortex (Christensen, Ramsoy, Lund, Madsen, & Rowe, 2006; Kerlin, Shahin, & Miller, 2010; Malnic, Hirono, Sato, & Buck, 1999), and the initiation of voluntary action is also tied to neural activity in the cerebral cortex (Amador & Fried, 2004; Churchland et al., 2012; Resulaj, Kiani, Wolpert, & Shadlen, 2009).

Meanwhile, neither spinal reflex circuits nor classical computing architectures generate representative information content or spontaneous volitional behavior. There must be something unique about cortical neuron computation that produces these two key features of consciousness. By examining five core puzzles of modern neuroscience, which highlight the distinctive characteristics of cortical neural circuits, we may identify a new approach to modeling consciousness as a computational process.

1. The computational power of the cerebral cortex is not tied to the number of binary computational units

The annals of neurology demonstrate that large numbers of neurons can be lost without any obvious deficit in perceptual content or voluntary behavior (Feuillet, Dufour, & Pelletier, 2007). While commenting on an illustrative case in a prominent science magazine, the cognitive scientist Axel Cleeremans rightly noted that “any theory of consciousness has to be able to explain why a person like that, who’s missing 90% of his neurons, still exhibits normal behavior” (Goldhill, 2016). The healthy human cerebral cortex contains 86 billion neurons, each of which are firing an action potential or not at any given moment. This biological system has been compared to a microchip with 86 billion transistors, each in an on-state or off-state at any given moment. Such microchips exist, but they do not generate qualitative perceptual experience or spontaneous goal-directed behavior. Clearly, the exascale computational power of the brain cannot be explained by modeling Shannon entropy, or summing the binary states of all computational units. The computational power of the brain must instead be measured in such a way that its information, or sum of possible system macrostates, is orders of magnitude higher than the number of binary computational units. A probability distribution of component microstates, or the von Neumann entropy of the system, may provide a better measure.

2. Cortical neurons engage in noisy coding, allowing random electrical noise to gate signaling outcomes

The firing patterns of invertebrate sensory neurons (Bialek & Rieke, 1992) and spinal motor neurons (Powers & Binder, 1995) are easily predicted by summing upstream inputs. By contrast, neurons in the mammalian cerebral cortex allow stochastic ion leak (Dorval & White, 2005) and spontaneous subthreshold fluctuations in membrane potential (Stern, Kincaid, & Wilson, 1997) to contribute to the likelihood of firing. Cortical neurons even exist in a coordinated “up-state,” right near the action potential threshold, allowing random electrical noise to gate signaling outcomes (Haider, Duque, Hasenstaub, & McCormick, 2006). The Hodgkin–Huxley equations provide an excellent approximation for predicting individual cortical neuron firing patterns under steady-state conditions. But the underlying relationship between membrane voltage, ion conductances, and channel activation—given by these partial differential equations—must ultimately be described by either modeling all eigenvectors in the system or by modeling a Hopf bifurcation to find the critical point where the cortical neuron flips from an off-state to an on-state (Austin, 2008; Rinzel & Miller, 1980; Rowat, 2007). These are inherently probabilistic processes.

3. Cortical neurons exhibit synchronous firing patterns at a range of nested frequencies

Synchronous neural activity is observed across the cerebral cortex during wakeful awareness and is correlated with reported perceptual experience (Buzsaki & Draguhn, 2004; Engel & Singer, 2001). Synchronous firing is also observed across the motor cortex during tasks (Jackson, Gee, Baker, & Lemon, 2003) and during sleep (Steriade, Timofeev, & Grenier, 2001). Yet, it has been difficult to reproduce these sparsely distributed patterns of synchronous neural activity by modeling recruitment with classical mechanisms (Stacey, Krieger, & Litt, 2011). These firing patterns across the cerebral cortex have previously been modeled as a random statistical ensemble (Beck et al., 2008), as the result of a random-connection model (Maoz, Tkacik, Esteki, Kiani, & Schneidman, 2020), and using fanofactor analysis of spike variance over time (Fayaz, Fakharian, & Ghazizadeh, 2022). These population-level firing patterns also indicate that cortical neuron signaling outcomes are inherently probabilistic, and must be modeled accordingly.

4. The cerebral cortex grows into a more ordered state over time, locally violating the second law of thermodynamics

The system remodels itself into a more compatible state with its surrounding environment over time (Pulvermuller et al., 1996; Turrigiano, Leslie, Desai, Rutherford, & Nelson, 1998; Zarnadze et al., 2016). This is an *unlikely* event, unless there is some physical mechanism by which a nonequilibrium thermodynamic system traps heat to drive computational work, reducing a broad probability distribution (high entropy) into a more ordered system state (low entropy). Identifying an optimal system state to effectively encode stimuli in the surrounding environment reduces uncertainty for the organism. Yet, while the biological mechanisms of plasticity are well-established, the physical laws underlying this systematic decrease in entropy are not.

5. The cerebral cortex is highly energy efficient, with very little net free energy lost to entropy

The brain is extraordinarily energy-efficient, with caloric intake almost exactly matching the amount of caloric energy expended on work (Engl & Attwell, 2015; Harris & Attwell, 2012; Howarth, Gleeson, & Attwell, 2012; Zhu et al., 2012; Zhu, Wang, Pan, & Zhu, 2019). Again, this is an *unlikely* occurrence, unless there is some mechanism by which a nonequilibrium thermodynamic system traps heat to drive computational work with minimal inefficiency. The ability to select or favor an optimal system state that encodes the surrounding environment not only reduces a system-wide probability distribution but also reduces the thermodynamic entropy of the system, resulting in more free energy being available to do work.

6. Conclusion

All five of these puzzles point to one rational solution: The brain engages in probabilistic computation. Indeed, this point has been well-appreciated in neuroscience for 30 years. Yet, for 30 years, neuroscientists have modeled cortical neurons as *noisy but classical systems*, instead of *inherently probabilistic systems*. Meanwhile, a mechanistic explanation for the unique emergent properties of cortical computation has remained elusive.

The concept that probabilistic coding in cortical neural networks might be intrinsically connected to the emergent psychological features of cortical neural networks is not well-established in the field of cognitive sciences. However, recent work by multiple groups has converged on the importance of this perspective, demonstrating that probabilistic approaches enhance deep learning (Gutman & Hyvärinen, 2013), lead to the generation of predictive models (Jara-Ettinger, Schulz, & Tenenbaum, 2020), yield context-dependent psychological states (Bruza et al., 2023; Chang, Biehl, Yu, & Kanai, 2020), and contribute to a sense of agency when navigating uncertain situations (Majchrowicz, Kulakova, DiCosta, & Haggard, 2020). Of course, the question of whether humans can actually exert agency, and whether such a capability might be compatible with causal determinism, is still hotly debated (Dennett 2014; Khalighinejada, Schurger, Desantisa, Zmigrod, & Haggard, 2018; O'Connor & Franklin, 2020). And some theories of consciousness, including Integrated Information Theory, do not require probabilistic computation for a system to have conscious experience—positing instead that *any* information-encoding system may have some level of consciousness (Tononi, Boly, Massimini, & Koch, 2016).

Yet, it may be useful to consider the brain as a truly probabilistic computational system, rather than a classical computational system. Cortical neural networks encode the surrounding environment by selecting an optimal system state, in context, from a broad distribution of possible system states. In comparison with spinal reflex circuits, this computational process allows electrical noise to gate signaling outcomes, and results in synchronous yet statistically random signaling outcomes across the neural network. This neural activity in turn drives contextually relevant behavior, paired with qualitative experience. The entire system is surprisingly robust to losing computational units, is almost perfectly energy-efficient, and physically remodels itself into a more compatible state with its surrounding environment over time. These are not typical features of a classical system, and may indicate a fundamentally different kind of computation. Identifying the mechanisms by which cortical neurons engage in probabilistic computation, with extraordinary thermodynamic efficiency, could provide a more cohesive theoretical framework for neuroscience. This effort may also benefit other areas of cognitive science. Indeed, the emergent features of cortical neuron computation—including perceptual content and nondeterministic behavioral outcomes—may turn out to be deeply tied to the unique physiological properties of cortical neurons.

References

Amador, N., & Fried, I. (2004). Single-neuron activity in the human supplementary motor area underlying preparation for action. *Journal of Neurosurgery*, *100*(2), 250–259.

Austin, T. D. (2008). The emergence of the deterministic Hodgkin–Huxley equations as a limit from the underlying stochastic ion channel mechanism. *Annals of Applied Probability*, *18*, 1279–1325.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E., & Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152.

Bialek, W., & Rieke, F. (1992). Reliability and information transmission in spiking neurons. *Trends in Neurosciences*, *15*(11), 428–434.

Bruza, P. D., Fell, L., Hoyte, P., Dehdashti, S., Obeid, A., Gibson, A., & Moreira, C. (2023). Contextuality and context-sensitivity in probabilistic models of cognition. *Cognitive Psychology*, *140*, 101529.

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929.

Chang, A. Y. C., Biehl, M., Yu, Y., & Kanai, R. (2020). Information closure theory of consciousness. *Frontiers in Psychology*, *11*, 1504.

Christensen, M. S., Ramsoy, T. Z., Lund, T. E., Madsen, K. H., & Rowe, J. B. (2006). An fMRI study of the neural correlates of graded visual perception. *Neuroimage*, *31*(4), 1711–1725.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), 51–56.

Dennett, D. (2014). Reflections on “Free Will”. Retrieved from Naturalism.org/resources/book-reviews/reflections-on-free-will

Dorval, A. D., & White, J. A. (2005). Channel noise is essential for perithreshold oscillations in entorhinal stellate neurons. *Journal of Neuroscience*, *25*(43), 10025–10028.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*(1), 16–25.

Engl, E., & Attwell, D. (2015). Non-signalling energy use in the brain. *Journal of Physiology*, *593*(16), 3417–3429.

Fayaz, S., Fakharian, M. A., & Ghazizadeh, A. (2022). Stimulus presentation can enhance spiking irregularity across subcortical and cortical regions. *PLoS Computational Biology*, *18*(7), e1010256.

Feuillet, L., Dufour, H., & Pelletier, J. (2007). Brain of a white-collar worker. *Lancet*, *370*(9583), 262.

Goldhill, O. (2016). A civil servant missing most of his brain challenges our most basic theories of consciousness. *Quartz Magazine*, 722614.

Gutman, M. U., & Hyvärinen, A. (2013). A three-layer model of natural image statistics. *Journal of Physiology*, *107*(5), 369–398.

Haider, B., Duque, A., Hasenstaub, A. R., & McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, *26*(17), 4535–4545.

Harris, J. J., & Attwell, D. (2012). The energetics of CNS white matter. *Journal of Neuroscience*, *32*, 356–371.

Howarth, C., Gleeson, P., & Attwell, D. (2012). Updated energy budgets for neural computation in the neocortex and cerebellum. *Journal of Cerebral Blood Flow & Metabolism*, *32*(7), 1222–1232.

Jackson, A., Gee, V. J., Baker, S. N., & Lemon, R. N. (2003). Synchrony between neurons with similar muscle fields in monkey motor cortex. *Neuron*, *38*(1), 115–125.

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The Naïve Utility Calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334.

Khalighinejada, N., Schurger, A., Desantisa, A., Zmigrod, L., & Haggard, P. (2018). Precursor processes of human self-initiated action. *NeuroImage*, *165*, 35–47.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *Journal of Neuroscience*, *30*(2), 620–628.

Majchrowicz, B., Kulakova, E., DiCosta, S., & Haggard, P. (2020). Learning from informative losses boosts the sense of agency. *Quarterly Journal of Experimental Psychology*, *73*(12), 2272–2289.

Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell*, *5*(5), 713–723.

Maoz, O., Tkacik, G., Esteki, M. S., Kiani, R., & Schneidman, E. (2020). Learning probabilistic neural representations with randomly connected circuits. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(40), 25066–25073.

O’Connor, T., & Franklin, C. (2020). Free will. *The Stanford Encyclopedia of Philosophy*.

Powers, R. K., & Binder, M. D. (1995). Effective synaptic current and motoneuron firing rate modulation. *Journal of Neurophysiology*, *74*(2), 793–801.

Pulvermuller, F., Eulitz, C., Pantev, C., Mohr, B., Feige, B., Lutzenberger, W., Elbert, T., & Birbaumer, N. (1996). High-frequency cortical responses reflect lexical processing: An MEG study. *Electroencephalography and Clinical Neurophysiology*, *98*(1), 76–85.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266.

Rinzel, J., & Miller, R. N. (1980). Numerical calculation of stable and unstable periodic solutions to the Hodgkin–Huxley equations. *Mathematical Biosciences*, *49*, 27–59.

Rowat, P. (2007). Interspike interval statistics in the stochastic Hodgkin–Huxley model: Coexistence of gamma frequency bursts and highly irregular firing. *Neural Computation*, *19*(5), 1215–1250.

Stacey, W. C., Krieger, A., & Litt, B. (2011). Network recruitment to coherent oscillations in a hippocampal computer model. *Journal of Neurophysiology*, *105*(4), 1464–1481.

Steriade, M., Timofeev, I., & Grenier, F. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *Journal of Neurophysiology*, *85*(5), 1969–1985.

Stern, E. A., Kincaid, A. E., & Wilson, C. J. (1997). Spontaneous subthreshold membrane potential fluctuations and action potential variability of rat corticostriatal and striatal neurons in vivo. *Journal of Neurophysiology*, *77*(4), 1697–1715.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*, 450–461.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, *391*(6670), 892–896.

Zarnadze, S., Bauerle, P., Santos-Torres, J., Bohm, C., Schmitz, D., Geiger, J. R., Dugladze, T., & Gloveli, T. (2016). Cell-specific synaptic plasticity induced by network oscillations. *eLife*, *5*, 14912.

Zhu, F., Wang, R., Pan, X., & Zhu, Z. (2019). Energy expenditure computation of a single bursting neuron. *Cognitive Neurodynamics*, *13*, 75–87.

Zhu, X. H., Qiao, H., Du, F., Xiong, Q., Liu, X., Zhang, X., Ugurbil, K., & Chen, W. (2012). Quantitative imaging of energy expenditure in human brain. *Neuroimage*, *60*(4), 2107–2117.