Expanding perspectives on cognition in humans, animals, and machines

Alex Gomez-Marin[1,2,*] and Zachary F Mainen[1,*]

[1] Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon, Portugal.

[2] Current affiliation: Instituto de Neurociencias, Consejo Superior de Investigaciones Científicas & Universidad Miguel Hernández, Sant Joan d'Alacant, Spain.

[*] Corresponding authors: A.G.M. (agomezmarin@gmail.com) and Z.F.M. (zmainen@neuro.fchampalimaud.org).

Abstract

Over the past decade neuroscience has been attacking the problem of cognition with increasing vigor. Yet, what exactly *is* cognition, beyond a general signifier of anything seemingly complex the brain does? Here, we briefly review attempts to define, describe, explain, build, enhance and experience cognition. We highlight perspectives including psychology, molecular biology, computation, dynamical systems, machine learning, behavior and phenomenology. This survey of the landscape reveals not a clear target for explanation but a pluralistic and evolving scene with diverse opportunities for grounding future research. We argue that rather than getting to the bottom of it, over the next century, by deconstructing and redefining cognition, neuroscience will and should expand rather than merely reduce our concept of the mind.

1. Introduction

Neuroscience aims to explain how the workings of the mind and behavior arise from the biology of interacting cells and molecules. This is arguably the biggest scientific challenge of our era. One of the main reasons that this problem is so hard is obvious: it is the vast complexity of the brain. But a somewhat less obvious part of the problem is agreeing on the question: what exactly are we trying to explain? The term "cognition" is used in neuroscience as shorthand to refer to the most complex functions the brain enables [1]. Cognition has traditionally been defined in terms of a suite of "cognitive functions"—attention, working memory, and so on. But are these really the key building blocks for complex behavior? If not, the definition of cognition should be realigned on a better-grounded empirical or theoretical approach. But where is that to be found? These questions are critical because without a firm grasp of what exactly is cognition, apparent progress in understanding its neural basis may evaporate as it is realized that the wrong object was under scrutiny.

1

In this review we survey contemporary approaches to brain complexity that might serve as a basis for redefining cognition. We begin by very briefly reviewing over 40 years of progress in studying the single neuron correlates of cognition. We emphasize the recent theoretical proposal that the complexity of these cognitive tasks is extremely low, implying that we might indeed be missing something in the definition. From this we launch into a broader discussion of possible theoretical bases for defining cognition. We then turn to consider the more empirical bases found in machine learning, behavioral studies, and neuro-phenomenology.

2. Neural basis of cognition in monkeys and rodents

The idea of cognition goes back at least to Aristotle (*gignoskein*, to know, the process of acquiring knowledge) but didn't really get going in a scientific form until the 19th century. The field of psychology has sought for many years to understand the basis of intelligence—cognition being the name for the higher capacities of the mind. By the time of Ebbinghaus and James in the late 19th century, many forms of cognition were beginning to be subjected to scientific study by psychologists.

Starting in the 1960's neuroscientists began to translate work in cognitive psychology from humans into monkeys. The goal of this work was to reveal direct physical correlates of cognitive functions such as working memory [2]. To do this, they systematically recorded the electrical activity of single neurons in behaving animals. This required them to teach monkeys (macaques) to perform cognitive tasks. This line of work demonstrated that monkeys not only exhibit behaviors considered to constitute the most basic building blocks of cognition—learning & memory, decision-making—but can perform more complex tasks including selective attention [3], flexible categorization [4], integration of evidence [5], probabilistic reasoning [6], metacognitive judgements [7], numerical reasoning [4], context-dependent decisions [8] and more.

Starting in the 1980's, researchers began to show that even rodents can also perform a wide variety of standard cognitive tasks including selective attention [9], flexible sensorimotor mapping [10] and working memory. In the last decade or so, this line of work in freely-behaving animals was refined to even more exacting standards, allowing quantitative measurements of single neuron activity in relationship to functions including confidence judgments [11], psychophysical thresholds [12], integration of evidence [13], self-initiated actions [14], cue combination [15], and flexible sensory-motor associations [16]. A burgeoning line of work has taken this approach to even more constrained, head-fixed, mice performing tasks such as memory guided decisions [17], sometimes in a virtual reality [18].

3. Cognition in just a few dimensions

This work has set the stage for huge progress in "opening the black box", as it allows rodents—which are smaller, cheaper, far easier to experiment with, less

ethically demanding—to stand in for monkeys or even humans. Particularly because this allows the use of molecular and genetic manipulations, which have become more and more powerful [19,20], many researchers using macaques have begun to switch to rats or mice or smaller, faster-breeding marmosets, in which genetic manipulations may be performed with greater ease.

Yet the desire to account for behavior on the basis of single spikes in single neurons raises a rather perplexing issue. It seems preposterous to hope to understand the concerted actions of a brain of 100 million neurons by studying a handful of neurons at a time. Yet, somehow this does not seem to be the case. Single neurons or populations <100 have typically revealed meaningful correlates of target cognitive phenomena. When analyzed using dimensionality reduction techniques, even the most complex tasks mentioned have yielded a dimensionality that is far lower than even the relatively small number of neurons being analyzed [8,21].

What is going on? Ganguli and colleagues have proposed a theory of "neural task complexity" [22]. They argue that as long as task parameters are limited and neuronal activity patterns vary smoothly across task parameters, that even for what we think of as complex cognitive tasks, the neural task complexity is very small. This work suggests that rather than recording more neurons, in order to make progress we need richer, more complex tasks [22]. This relates to the important issue of how much of what we measure in behaving animals is due to behavior itself or simply reflecting the environment in which the animal is placed in the laboratory. It also raises the more radical possibility that neuroscience may already be outgrowing the traditionally defined set of cognitive functions.

4. A basis in psychology or in the brain itself?

Nearly 80 years ago, Tolman proclaimed that "everything important in psychology (...) can be investigated in essence through the continued experimental and theoretical analysis of the determiners of rat behavior at a choice-point in a maze" [23]. Our current cognitive taxonomy appears to include a much richer set of behaviors, but much of our basic cognitive toolkit was already old hat to Tolman, having been developed decades ago by psychologists such as William James [24]. Granting James' genius, how much confidence can be taken that the 19th century agenda of cognitive psychology covers the breadth and depth of mental function?

The idea of dissecting cognition into isolatable components, combined with the possibility of studying the neural basis of such a taxonomy of sub-functions (attention, working memory, meta-cognition, etc.) appears to be well-suited to the paradigm set by molecular biology and genetics for the study of biological mechanism. The prominence of neural reductionism, "the tendency to seek all causal relations of behavior in brain processes" [25], has caused some to wonder whether psychological explanation is in danger of extinction [26]. This is a bias that might be exacerbated in the current climate of increased competition for

publication and funding. What does it mean to explain cognition[27]? If explaining entails explaining it away (if making it clearer results in dissolving it into low level mechanisms) then the drive toward reductionistic accounts may ironically turn into a roadblock in our understanding. Instead, here we attempt to redefine cognition from its multiple approaches. Truly novel work does go on at the psychological level, even if it has a harder time finding an audience without a neural substrate. For instance, while Bayesian-inspired psychological research has made inroads into mechanistic cognitive neuroscience, cognitive psychologists continue to work beyond what is cutting edge in neuroscience on problems such as rapid structural learning and concept learning [28].

It has been argued that psychology suffers from human-invented categories that do not necessarily conform to nature—"we take a man-created word or concept (...) and search for brain mechanisms that may be responsible for the generation of this conceived behavior" [29]. An alternative is an inside-out or bottom-up approach building from brain mechanisms toward cognitive function. While this strategy offers freedom from potentially awkward hypotheses about the phenomena to be explained, for a system as complex as a brain, a pure bottom up approach appears to be a non-starter. The power of discovery based on sophisticated techniques relying on big data can fail to produce significant results but, in a sense, it can never be wrong. Therefore falsifiable theories and concrete testable hypothesis — not just impressive tour-de-force— are needed. Even with an organism for which the complete wiring diagram is available, *C. Elegans*, behavioral function cannot be deduced without additional hypotheses. This problem is also illustrated by the heroic but ultimately humbling attempt to model a single cortical column from the bottom-up [30].

5. Theories and frameworks for cognition

If one is to approach and define cognition at least partly from the top-down, what are the available theoretical frameworks from which this might be done? Here we find several alternatives, not necessarily mutually exclusive. Such theoretical approaches are rooted in mathematics and deep concepts—probabilistic inference, learning, entropy, complexity—that are being mined for their potential as an ultimate level of explanation for brain function.

Dynamical systems. A long-standing theoretical approach to cognition is based on nonlinear dynamical systems theory, which has been proposed to be a non-computational way to conceive of cognition [31]. Two claimed advantages of conceiving of a cognitive agent as a dynamical system, rather than a digital computer, are the possibility to interact with the world without "representing" it and the possibility to avoid discretizing time [32]. Dynamical systems approaches were the basis for understanding the basis of neuronal function at the cellular level [33]. They are also the basis for the more "biophysical" network models, which have demonstrated that the dynamics of very simple neuronal circuits can be turned to perform elementary cognitive functions such as working memory and decision-making [34,35]. Moreover, with appropriate (albeit sometimes implausibly nonlocal) learning rules these networks can be endowed with

surprisingly rich dynamics, e.g. generating arbitrary dynamical mappings [36,37], providing mechanistic hypothesis for the basis of context-dependent decision-making. Dynamical network models based on engineering rather than learning solutions have begun to be assembled to perform complete tasks [38] but cannot say play a video game. Approaches based on optimization (learning) of relatively simple objective functions are remarkably successful in capturing the dynamics of complex actions (e.g. standing up) using nonlinear dynamical models of the motor plant and physical environment [39]. Work from this perspective argues that cognition and action are deeply intertwined, an "embodied" perspective reinforced by the loopiness of brain anatomy. Motor-control inspired dynamic systems approaches may provide fertile ground for future neuroscience of cognition.

Reinforcement learning. Over the last two decades, reinforcement learning (RL) theory has been one of the most impactful theoretical frameworks in neuroscience [40]. With roots in artificial intelligence, to which we return below, the success of RL in neuroscience has been forged by theorists who have taken a very much top-down or normative approach to ask what *should* the brain do. Importantly, RL theorists have also cared deeply about the underlying brain circuitry, particularly at the systems level (area by area), and the behavior, particularly that carried out by the "classical" animal learning theorists (essentially the line of Skinner). Although the name emphasizes one flavor of learning, RL naturally incorporates decision-making, and strong relationships with microeconomic theory exist [41]. RL deals with optimization based on an externally-supplied scalar objective function. Yet it should not be forgotten that learning is a much richer and murkier domain. Beyond RL are forms of learning that work without supervision of any kind, that make use of observation, and that require only a single exposure.

Bayesian decision theory. Decision theory is a branch of mathematical psychology that deals with choosing based on noisy samples. It was inspired by the problem of code-breaking and fueled for many years by the ability to explain human reaction time and performance data across a range of domains [42]. It thus has both a normative basis and empirical support. The diffusion-to-bound model has been linked to single neuron recordings, implemented using the circuit-level dynamic models and derived as a subcase of Bayesian theory. Decision-making is clearly a limited aspect of cognition, but integration is a key process not only in accumulation of evidence, but in working memory, cue combination, and Bayesian inference, etc. Evidence for non-leaky, noiseless integration even in rats [43] and, the ability to account for other aspect of behavioral data, for example, action timing [14], contribute to the attractiveness of these models as a basis for describing cognition.

Quantum cognition is an offshoot of decision theory and mathematical psychology that borrows the quantum probability formalism based on Dirac and Von Neumann axioms. Quantum decision models can explain several anomalous effects in sequential decision-making [44]. A quantum diffusion-to-bound model was introduced [45], which is a promising start toward other cognitive mechanisms. In the background, as either a beacon or a specter, is whether it is

ultimately tied to quantum mechanics or quantum computing, as infamously proposed by Penrose [46].

Information theory. An information processing metaphor, in which the brain is essentially an input/output system has dominated cognitive science since its origins. Information theory quantitatively formalizes the communication of signals using the fundamental concept of entropy. In neuroscience it has been used successfully to understand how neural signals propagate through sensory systems [47] and provides a useful basis for conceptualizing the efficiency of neural representations or codes. Moreover, information theory has been used to define what constitutes a "conscious" process [48,49] and recently, it has been shown that simple forms of "cognitive" behavior, such a solving a puzzle or cooperating socially, can emerge from principles of entropy maximization, where future entropy acts as a force in the present[50]. Under the rubric "free energy", and married with Bayesian and RL theory, such approaches form the core of a proposed a universal brain model [51].

Theory of computation. Together with information, the notion of computation has long been a dominant abstraction for conceiving the brain's most complex functions. Perhaps most fundamentally, Kolmogorov complexity provides a definition of randomness based on the shortest computer program that can be run on a Turing machine [52]. Algorithmic complexity may be useful in application to cognition by offering a grounded definition of what is hard or complex. Despite the dominance of the computational metaphor, there has been surprisingly little contact between cognitive neuroscience and most aspects of computer science. Interestingly, it has been suggested that practical advances in the field of computer engineering (e.g. operating systems) are not themselves well-understood by computer scientists [53], which may partly explain why current approaches to thinking about cognition remain relatively uninfluenced by advances in computer engineering.

6. Machine cognition

The theoretical approaches just outlined are rooted in mathematics and deep concepts—probabilistic inference, learning, entropy, complexity—that are being mined for their potential as an ultimate level of explanation for brain function. These theories have also driven, in parallel, the quest to artificially engineer intelligent machines. In particular, the field of neural networks or machine learning has been closely related to the field of computational neuroscience. What might be the relevance of machine learning, beyond a basis in some of the same normative theories, to defining cognition?

Artificial intelligence (AI) is a practical perspective by which to define what brain functions are truly "hard", and therefore "cognitive", under our current theoretical assumptions. It has long been understood that for "symbolic" AI, problems like chess are easy while problems like vision are impossible, not to mention the fact that robots continue to play miserable soccer. Neural networks based on "deep learning" appear to have cracked what had been considered

difficult problems such as image recognition. As these approaches are pushed, their limits may come into clearer focus before changing again as new algorithms emerge. Already, hybrid approaches combing, for example, neural networks with Turning machines [54], are proliferating.

Neural networks share some core features with brains, including distributed processing over simple units and modifiable connections. But they omit the vast majority of what we know about brains, from spikes to cortical layers. And what makes deep learning successful has not been the incorporation of more exacting neural principles but rather clever algorithms such as the back-propagation learning rule that has little to do with brain architecture [55]. It remains uncertain to what degree machine learning approaches will benefit from incorporating more of the brain's architecture. Consider how divergent are the architectures of the cephalopod and the mammalian brain. Pessimistically (or not), brains and machines might evolve two different solutions to intelligence built on two different sets of tricks, with neither illuminating much the other.

7. Cognition informed by behavior

If we get behavior wrong, we will not get cognition right. As we have argued recently [56] behavior is the foundation of neuroscience, not just a handy tool to investigate the function of the nervous system. When seen as a phenomenon of its own right —not as a mere neural side-effect— the study of behavior becomes invaluable to reveal the nature of cognition.

Bottom up approaches aim to discover, from sophisticated analyses, structure in the copious and high-resolution data currently available. In computational ethology [57], state-of-the-art approach is the discovery of stereotyped behaviors directly from the images themselves [58], [59]. Such remarkable data analysis progresses will ultimately need to be incorporated in the context of falsifiable theories. Behavior itself is also a worthy subject of "top-down" theoretical consideration that may inform and perhaps even upend our notion of cognition. Among other potential avenues, one that seems particularly necessary today is to reconsider the idea of negative feedback control, the basis for the "cybernetic revolution" of the 1940's. Maxwell's 1867 seminal paper on "governors" described autonomous systems that can maintain certain variables to a reference value despite external disturbances (such as the tray of a waiter as he walks through the restaurant). In fact, negative feedback control is closely related to the concept of homeostasis.

While cybernetic concepts thoroughly transformed engineering practice, they had relatively little impact on the study of behavior. Their potential relevance was however eloquently championed by Powers, who proclaimed that behavior is the control of perception [60]. Then, to understand behavior is to discover the variables that the organism controls, not how it responds to certain stimuli. The implication of this view is that linear cause-and-effect explanations will fail because they assume that real behavior is open loop, when it is not [61]. Behavioral approaches that focus on manipulating the input (stimuli, rewards) and

measuring the output (responses, actions) tend to confound controlled with controlling variables. The difference is crucial: the former are controlled by the behaving system whereas the latter are controlled by the human studying the behaving system[62]. One might then be studying the properties of the environmental feedback mistaking them for the behavioral function itself. Note, for instance, the distinction between the behavior of a fan and a thermostat: the first controls its output, the second its input. The thermostat certainly "emits" heat, but that is not really what it does. What it does (or attempts to do) is to specify the temperature to be sensed.

Therefore, rather than concentrating on what you see an animal doing, what is relevant is what the animal is trying to perceive. This double inversion (from my point of view to the animal's and from action to perception) has three critical behavioral implications for the study of cognition: (i) motor output is a side effect of perceptual control and so mere quantitative data collection data will not suffice, (ii) averaging across individuals may smear out control variables, and (iii) restrained experimental setups may not let animals control the relevant inputs; freedom in the requisite dimensions is required. In short, control (circular causality) and subjectivity (animal centrism) are essential ingredients in behavioral and cognitive neuroscience.

8. Cognition beyond the brain itself

Finally, we note two important and relatively new approaches that assert that the data relevant to cognition do not lie solely within the realm of neuroscience facts as traditionally construed. First, "neurophenomenology", introduced by Francisco Varela, aims to include introspection as a source of scientific evidence[63]. Neurophenomenology is based on first-person (subjective) moment-to-moment accounts coupled with classical quantification procedures such as neuroimaging. Together, these are proposed to provide mutual constraints to tighten the link between neurodynamics and conscious experience and, in particular, to offer new insights in the study of variability and spontaneity in cognition[64]. Second, "embodied cognition", also championed by Varela, centers on the recognition that cognition is embedded in a context beyond the brain itself, including the body itself as well broader environmental, biological and cultural systems [65]. Both neurophenomenology and embodied cognition illustrate the nascent fruits of substantive dialogue between philosophy and neuroscience that we hope will continue to instill new perspectives in the years to come.

9. Conclusions

The neuroscience of cognition is in its infancy. Greatly empowered by new tools, it is still largely hamstrung by the perspectives of the past. In formulating the problems of the future, the field should benefit by paying careful mind not only to neural data but also to the nature of the phenomena to be explained. If the questions we ask are restricted to the same terms they have traditionally been

formulated in, potential understanding will be greatly restricted. The various disciplines and frameworks we have summarized here offer rich and widely differing perspectives on how to conceive of cognition. Their mutual incoherence attests to a conceptual vitality that is not merely appropriate but critical to a young field tackling an immense problem. Going forward, it is important for the field to continue to evolve, drawing inspiration from any and all potentially relevant quarters and resisting the tendency to become stuck in a singular, archaic, view of the mind. When we define cognition, we are helping, after all, to define what it is to be human.
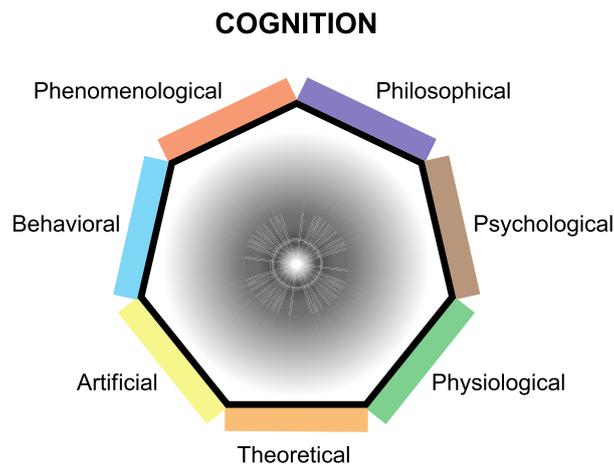
COGNITION

Figure. A pluralistic landscape of abstractions to approach the phenomenon of cognition

References

1. Cromwell, H. C. & Panksepp, J. Rethinking the cognitive revolution from a neural perspective: How overuse/misuse of the term 'cognition'and the neglect of affective controls in behavioral neuroscience could be delaying progress in understanding the BrainMind. *Neurosci. Biobehav. Rev.* **35,** 2026–2035 (2011).

2. Fuster, J. M. & Alexander, G. E. Neuron activity related to short-term memory. *Science* **173,** 652–654 (1971).

3. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18,** 193–222 (1995).

4. Nieder, A., Freedman, D. J. & Miller, E. K. Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297,** 1708–1711 (2002).

5. Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86,** 1916–1936 (2001).

6. Kiani, R. & Shadlen, M. N. Representation of confidence associated with a decision by neurons in the parietal cortex. *science* **324,** 759–764 (2009).

7. Hampton, R. R. Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comp. Cogn. Behav. Rev.* **4,** 17 (2009).

8. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78–84 (2013).

9. Carli, M., Robbins, T. W., Evenden, J. L. & Everitt, B. J. Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. *Behav. Brain Res.* **9,** 361–380 (1983).

10. Roberts, A. C., Robbins, T. W. & Everitt, B. J. The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Q. J. Exp. Psychol.* **40,** 321–341 (1988).

11. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455,** 227–231 (2008).

12. Miura, K., Mainen, Z. F. & Uchida, N. Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron* **74,** 1087–1098 (2012).

13. Hanks, T. D. *et al.* Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* **520,** 220–223 (2015).

14. Murakami, M., Vicente, M. I., Costa, G. M. & Mainen, Z. F. Neural antecedents of self-initiated actions in secondary motor cortex. *Nat. Neurosci.* (2014).

15. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* (2014).

16. Duan, C. A., Erlich, J. C. & Brody, C. D. Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. *Neuron* **86,** 1491–1503 (2015).

17. Guo, Z. V. *et al.* Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81,** 179–194 (2014).

18. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484,** 62–68 (2012).

19. Bernstein, J. G. & Boyden, E. S. Optogenetic tools for analyzing the neural circuits of behavior. *Trends Cogn. Sci.* **15,** 592–600 (2011).

20. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nat. Neurosci.* **18,** 1213–1225 (2015).

21. Kobak, D. *et al.* Demixed principal component analysis of population activity in higher cortical areas reveals independent representation of task parameters. *ArXiv Prepr. ArXiv14106031* (2014).

11

22. Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32,** 148–155 (2015).

23. Tolman, E. C. The determiners of behavior at a choice point. *Psychol. Rev.* **45,** 1 (1938).

24. James, W. *1902/1992) The Varieties of Religious Experience: A Study in Human Nature.* (Birmingham, AL: The Classics of Psychiatry and Behavioral Science Library, 1901).

25. Lashley, K. S. Basic neural mechanisms in behavior. *Psychol. Rev.* **37,** 1 (1930).

26. Wright, C. D. Is Psychological Explanation Becoming Extinct? *Matter Mind* 249 (2009).

27. Craik, K. J. W. *The nature of explanation.* (CUP Archive, 1967).

28. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350,** 1332–1338 (2015).

29. Buzsáki, G. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus* **15,** 827–840 (2005).

30. Markram, H. *et al.* Reconstruction and simulation of neocortical microcircuitry. *Cell* **163,** 456–492 (2015).

31. Van Gelder, T. What might cognition be, if not computation? *J. Philos.* 345–381 (1995).

32. Van Gelder, T. The dynamical hypothesis in cognitive science. *Behav. Brain Sci.* **21,** 615–628 (1998).

33. Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117,** 500–544 (1952).

34. Machens, C. K. & Brody, C. D. Design of continuous attractor networks with monotonic tuning using a symmetry principle. *Neural Comput.* **20,** 452–485 (2008).

35. Schulze, A. *et al.* Dynamical feature extraction at the sensory periphery guides chemotaxis. *eLife* **4,** e06694 (2015).

36. Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63,** 544–557 (2009).

37. Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* **14,** 2531–2560 (2002).

38. Eliasmith, C. *et al.* A large-scale model of the functioning brain. *science* **338,** 1202–1205 (2012).

39. Todorov, E. Optimality principles in sensorimotor control. *Nat. Neurosci.* **7,** 907–915 (2004).

40. Dayan, P. & Niv, Y. Reinforcement learning: the good, the bad and the ugly. *Curr. Opin. Neurobiol.* **18,** 185–196 (2008).

41. Glimcher, P. W. & Fehr, E. *Neuroeconomics: Decision making and the brain*. (Academic Press, 2013).

42. Ratcliff, R. Methods for dealing with reaction time outliers. *Psychol. Bull.* **114,** 510 (1993).

43. Brunton, B. W., Botvinick, M. M. & Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340,** 95–98 (2013).

44. Pothos, E. M. & Busemeyer, J. R. Can quantum probability provide a new direction for cognitive modeling? *Behav. Brain Sci.* **36,** 255–274 (2013).

45. Kvam, P. D., Pleskac, T. J., Yu, S. & Busemeyer, J. R. Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proc. Natl. Acad. Sci.* **112,** 10645–10650 (2015).

46. Penrose, R. & Hameroff, S. Consciousness in the universe: Neuroscience, quantum space-time geometry and Orch OR theory. *J. Cosmol.* **14,** 1–17 (2011).

47. Rieke, F. *Spikes: exploring the neural code.* (MIT press, 1999).

48. Tononi, G. An information integration theory of consciousness. *BMC Neurosci.* **5,** 42 (2004).

49. Tononi, G. & Koch, C. Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370,** 20140167 (2015).

50. Wissner-Gross, A. D. & Freer, C. E. Causal entropic forces. *Phys. Rev. Lett.* **110,** 168702 (2013).

51. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11,** 127–138 (2010).

52. Turing, A. M. On computable numbers, with an application to the Entscheidungsproblem. *J Math* **58,** 5 (1936).

53. Doyle, J. C. & Csete, M. Architecture, constraints, and behavior. *Proc. Natl. Acad. Sci.* **108,** 15624–15630 (2011).

54. Graves, A., Wayne, G. & Danihelka, I. Neural Turing Machines. *ArXiv Prepr. ArXiv14105401* (2014).

55. Gomes, L. Machine-learning maestro michael jordan on the delusions of big data and other huge engineering efforts. *IEEE Spectr. Oct* **20,** (2014).

56. Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M. & Mainen, Z. F. Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nat. Neurosci.* **17,** 1455–1462 (2014).

57. Anderson, D. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84,** 18–31 (2014).

58. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11,** 20140672 (2014).

59. Wiltschko, A. B. *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* **88,** 1121–1135 (2015).

60. Powers, W. T. *Behavior: The control of perception*. (Aldine Chicago, 1973).

61. Yin, H. H. in *Computational and robotic models of the hierarchical organization of behavior* 319–347 (Springer, 2013).

62. Marken, R. S. Controlled variables: Psychology as the center fielder views it. *Am. J. Psychol.* **114,** 259–282 (2001).

63. Varela, F. J. Neurophenomenology: A methodological remedy for the hard problem. *J. Conscious. Stud.* **3,** 330–349 (1996).

64. Fazelpour, S. & Thompson, E. The Kantian brain: brain dynamics from a neurophenomenological perspective. *Curr. Opin. Neurobiol.* **31,** 223–229 (2015).

65. Varela, F. J., Thompson, E. & Rosch, E. *The embodied mind*. (CogNet, 1992).