

Review



**Cite this article:** Zurn P, Bassett DS. 2020 Network architectures supporting learnability. *Phil. Trans. R. Soc. B* **375**: 20190323. <http://dx.doi.org/10.1098/rstb.2019.0323>

Accepted: 31 December 2019

One contribution of 11 to a theme issue 'Unifying the essential concepts of biological networks: biological insights and philosophical foundations'.

**Subject Areas:**

cognition, neuroscience, theoretical biology, computational biology, bioinformatics

**Keywords:**

network neuroscience, knowledge networks, constraints, curiosity, collective dynamics

**Author for correspondence:**

Danielle S. Bassett  
e-mail: [dsb@seas.upenn.edu](mailto:dsb@seas.upenn.edu)

# Network architectures supporting learnability

Perry Zurn<sup>1</sup> and Danielle S. Bassett<sup>2,3,4,5,6,7</sup>

<sup>1</sup>Department of Philosophy, American University, Washington, DC 20016, USA

<sup>2</sup>Department of Bioengineering, School of Engineering and Applied Science, <sup>3</sup>Department of Physics and Astronomy, College of Arts and Sciences, <sup>4</sup>Department of Electrical and Systems Engineering, School of Engineering and Applied Science, <sup>5</sup>Department of Neurology, Perelman School of Medicine, and <sup>6</sup>Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>7</sup>Santa Fe Institute, Santa Fe, NM 87501, USA

PZ, 0000-0003-1964-2946; DSB, 0000-0002-6183-4493

Human learners acquire complex interconnected networks of relational knowledge. The capacity for such learning naturally depends on two factors: the architecture (or informational structure) of the knowledge network itself and the architecture of the computational unit—the brain—that encodes and processes the information. That is, learning is reliant on integrated network architectures at two levels: the epistemic and the computational, or the conceptual and the neural. Motivated by a wish to understand conventional human knowledge, here, we discuss emerging work assessing network constraints on the learnability of relational knowledge, and theories from statistical physics that instantiate the principles of thermodynamics and information theory to offer an explanatory model for such constraints. We then highlight similarities between those constraints on the learnability of relational networks, at one level, and the physical constraints on the development of interconnected patterns in neural systems, at another level, both leading to hierarchically modular networks. To support our discussion of these similarities, we employ an operational distinction between the modeler (e.g. the human brain), the model (e.g. a single human's knowledge) and the modelled (e.g. the information present in our experiences). We then turn to a philosophical discussion of whether and how we can extend our observations to a claim regarding explanation and mechanism for knowledge acquisition. What relation between hierarchical networks, at the conceptual and neural levels, best facilitate learning? Are the architectures of optimally learnable networks a topological reflection of the architectures of comparably developed neural networks? Finally, we contribute to a unified approach to hierarchies and levels in biological networks by proposing several epistemological norms for analysing the computational brain and social epistemes, and for developing pedagogical principles conducive to curious thought.

This article is part of the theme issue 'Unifying the essential concepts of biological networks: biological insights and philosophical foundations'.

## 1. Introduction

The human mind is equipped with rich materials and diverse strategies with which to interpret flows of information into structured units and relations [1,2]. In many cases, such interpretive inferences are drawn from temporally extended streams of stimuli where bits of information are presented to our perceptive apparatus in a sequential manner [3–8]. We read a book or listen to a lecture composed of word sequences. We listen to a song or instrumental piece composed of sound sequences. We engage in discussions with sequential arcs. We perceive a visual scene composed of light and colour sequences. We walk through the day and experience sequences of heat, air currents and

human touch. From those one-dimensional streams of information we infer the complex structure of the world, and our potential knowledge thereof [9,10].

In fact, reality is knowable as a set of informational units and relations among them. It is these units and their relations that scientists devote their lives to understanding. Henri Poincaré noted in his book *Science and Hypothesis* (1902) that, ‘The aim of science is not things themselves, as the dogmatists in their simplicity imagine, but the relations among things; outside these relations there is no reality knowable.’ [11, p. xxiv] While Poincaré’s post-Kantian claim may seem eminently reasonable and straightforward, its implications are perhaps more intriguing than he knew. If science is concerned not with reality itself, but with the relations among things, then scientific knowledge relies on perpetually fine-tuning the network architectures of information. That is, what can be known is a network of relations, and knowledge itself is a network of that information. John Dewey in his book *Democracy and Education* (1916) suggested as much when he wrote that, ‘[K]nowledge is a perception of those connections of an object which determine its applicability in a given situation. [...] An ideally perfect knowledge would represent such a network of interconnections that any past experience would offer a point of advantage from which to get at the problem presented in a new experience.’ [12, p. 116] Scientific knowledge, then, is an increasingly effective network of ideas that model interconnections in the world.

And what is the apparatus by which we process and construct the relations between things and reason through those relations (i.e. engage in relation and relational learning) [13–15]? Our primary computational infrastructure is the brain [16]. Much of what we now know about the brain relates to the putative functions of single areas, and has been derived from lesion and imaging studies in both human and non-human animals [17,18]. Yet, recent work has noted a marked increase in explanatory power from circuit-level descriptions that map the location, transmission, and manipulation of information throughout spatially distributed networks of neural units [19–21]. Over the past decade or more, the study of the network architecture of neural circuits has been formalized in the field of *network neuroscience* [22], which draws on graph theory, statistical mechanics and network science to create and study network models of neural systems [23–25]. In some ways, this appreciation of the brain as a networked system is new, particularly in its formal mathematical nature; yet in other ways, this appreciation is simply a remembrance of what we have speculated about for almost two centuries since Schwann’s proposal in 1839 [26], and known decisively for more than one century as the Neuron Doctrine [27]. In 1906, Cajal and Golgi were awarded the Nobel Prize for Physiology or Medicine for their demonstrative experiments confirming that nerve cells are the discrete units that make up brain tissue, and that they comprise a connected network system by discrete sites of contact.

We pause at this juncture in the ever-vigorous progress of scientific and philosophical investigation into the nature of mind and reality. We focus our attention on the networked nature of knowledge as well as the networked nature of the computational unit—the brain—that allows the human mind to process and construct that knowledge. We review recent evidence for network constraints on learnability of information and network constraints on the architecture of neural circuits that support that learning. We discuss similarities in those constraints and attempt to reason about why there might exist a

marked correspondence in hierarchically modular organization in both types of networks. Our discussions of the architecture of knowledge, the architecture of the brain, and their relations allow us to then explicitly reason about the relationship between the modelled, the modeller, and the model. That reasoning leads us to ask how recent empirical evidence could inform deeper explanations and mechanisms for knowledge acquisition. We engage in an interdisciplinary discussion of possible epistemological norms for studying brain network architecture and its role in the acquisition of knowledge networks. Finally, we close with a few thoughts on how these discussions could inform pedagogical principles conducive to curious thought engendering knowledge acquisition. Because we come from quite different areas of inquiry (philosophy, physics and neuroscience), and because we hope that this piece will be accessible across fields, we aim for a simple and clear presentation of the ideas, and eschew jargon wherever possible. We have been free with our citations to ensure that practitioners in a given field are directed to relevant work in their disciplinary domain. Notably, what we provide is a review of extant literatures, selected for their relevance and insight into the network architectures supporting learnability, upon which future experimental and theoretical analyses might be built.

## 2. Network constraints on the learnability of relational knowledge

To make our discussion here a bit more concrete, let us consider a professor sitting down at their (likely disheveled) desk to develop tomorrow’s lecture or discussion plan. For simplicity, let us ignore the precise topic of the lecture or discussion (and the related problem of how students learn representations useful for modelling the world), and instead focus solely on the structure of the content. Perhaps the content can be quite easily subdivided into 15 narrowly defined concepts, which are related to one another in a non-trivial topology. Concepts 1 through 5 may be strongly related, forming a module; concepts 6 through 10 may be strongly related, also forming a module; and concepts 11 through 15 may be strongly related, forming a third module. But the three modules are not completely independent of one another; instead, module 1 is conceptually linked to module 2, and module 2 is conceptually linked to module 3, which in turn harks back to module 1. How does the professor choose to take this potentially high-dimensional network architecture between concepts, and translate it to the students when time is one-dimensional and uni-directional, and thus only one word can be spoken at a time, and presumably only one concept presented or discussed at a time? The same challenge is faced by any writer or speaker: how must one take a bit of knowledge, with some inherent network architecture of relations between informational units, and translate that knowledge into a continuous stream of words? Will the reader or listener or discussant be able to infer the pattern of relations between units? If so, how? Is there an optimal mapping of the network into a stream that supports rapid inference on the part of the receiver or interlocutor?

### (a) Statistical learning and the relevance of transition probabilities

Broadly, the problem of inferring patterns of pairwise dependencies from incoming streams of data is in fact much more

general than simply listening to a lecture or engaging in a discussion. Indeed, the capacity to make such inferences allows us to learn language [28], segment visual events [3], parse tonal groupings [4], parse spatial scenes [5,29], infer social networks [6,30] and perceive distinct concepts [7,8,31,32]. The underlying general learning mechanism is known as *statistical learning*, which can be defined as the ability for humans and other animals to extract statistical regularities from the world around them to learn about the environment [33]. For example, a baby can listen to a stream of syllables and detect the probabilities with which syllables follow one another. Sets of syllables that follow one another with high probability are perceived as units (such as words); when one syllable rarely follows a second syllable, the transition is perceived as a boundary between units (a break between words). Although first identified in human infant language acquisition, statistical learning is now thought to be a generalized learning mechanism that is relevant across information modalities and operationalized in multiple species [34].

### (b) Moving beyond local transition probabilities

While it was clear from its inception that statistical learning offered a compelling description for sensitivity to pairwise dependencies between informational units, it was not immediately clear whether that description could be extended to explain sensitivity to a complex network structure underlying sequential input from our world [35]. The foundational work in statistical learning manipulated the transition probability between two adjacent stimuli in a temporal stream. Yet, evidence quickly accumulated that supported the notion that humans were also capable of learning from the probabilities between non-adjacent stimuli [36,37], quaintly referred to as ‘learning at a distance’ harking back to the quantum mechanical notion of ‘action at a distance’ [38]. For example, we come to know not only that ‘Peter’ and ‘Rabbit’ are distinct words, but also that we are more likely to see or hear those two words in the same story than to hear ‘Thayne’ and ‘Rabbit’ in the same story. Human sensitivity to structure beyond adjacent transition probabilities was further underscored by pioneering work from Schapiro and colleagues, who drew a sequence of visual stimuli from a random walk on a network while keeping all transition probabilities fixed at a constant value [39]. The network contained three main modules and the investigators observed that humans were able to demarcate module boundaries from the temporal stream, supported by neural activity in the hippocampus [40].

### (c) Explicitly probing learnability of network architectures

Following these important studies that provided initial suggestions that humans were sensitive to a network architecture guiding the statistics of their experiences, the field faced two main challenges. First, an experimental paradigm was needed that could provide an assessment of exactly how much each relation (or edge) in the network was learned. Like Shapiro *et al.* [39], Karuza and colleagues used a task in which a stream of visual stimuli was constructed by traversing a given network using a particular type of walk and in which humans were given a cover task of detecting whether stimuli were upright or rotated [41]. From the cover task, the investigators were able to extract a reaction time for each transition

between two stimuli; from the type of walk (random, Eulerian and Hamiltonian), the investigators were able to determine that the manner in which the network was traversed impacted human expectations. Second, the field needed a clear demonstration that human expectations could be manipulated differently by different network architectures. Kahn and colleagues studied human expectations derived from a stream of stimuli drawn from a random walk on three different network architectures (modular, lattice and random), and showed that humans reacted with differential swiftness to sequences constructed from each network type [42]. The work also mapped the original context of visual stimuli [6,39,41] to motor commands, thereby demonstrating that network learning was robust across modalities.

### (d) Hallmarks of network learning in humans

Throughout the existing literature, the human capacity to acquire expectations about a network architecture underlying a temporal stream of information is particularly marked by the so-called cross-cluster surprisal [41]: humans react more slowly to transitions between modules in a network than to transitions within modules [6,30,41–43]. This finding suggests that humans are able to infer the presence of higher dimensional topological clusters within one-dimensional streams of information. As a human behaviour, this effect on reaction time is particularly striking in light of the fact that the transition probabilities of all edges are identical, indicating that humans must be sensitive to a meso-scale or global organization, unfolding over long time scales within the information stream. Perhaps even more strikingly, humans react more swiftly when the stream of information is drawn from a modular network than when it is drawn from a lattice network [42,43]. In turn, this behaviour suggests that humans find the modular architecture relatively easy to learn, although it is as yet unclear whether that ease is explained by innate knowledge of certain graphical motifs [1], a flexible learning algorithm [44–46], or constraints on the computational complexity of associated cognitive processes [43,47]. The human response to network-based temporal streams of information is remarkable when we consider the mental computations that subserve it. Neither the cross-cluster surprisal effect nor the modular-lattice effect would be observed from simulated agents with optimal rationality, who instead would accurately learn the transition probabilities that are held constant across all edges and all network architectures in these experiments.

### (e) Building mental models of our world

To explain these curious, non-artificial (some would even say non-optimal) features of human behaviour, we turn to the question of exactly how humans build models of their world. While this question has been asked in different ways for millennia [48], and from within the discipline of psychology for decades [1,2], here we focus on the specific question of how humans perceive relational knowledge, building models of network architectures explaining transition probabilities of sequentially experienced stimuli. We consider the relatively reasonable hypothesis that humans seek to minimize both computational resources and errors, which can be formalized by the free-energy principle [43]. We then draw on a subfield of theoretical physics known as statistical mechanics to stipulate a maximum entropy (minimal complexity) solution, thereby blending principles of thermodynamics and information theory. The formal

mathematical model explains human behaviour by predicting that humans perform a sort of fuzzy temporal integration, which serves to strengthen their expectations of edges in local clusters. Using this model, we can account for both the cross-cluster surprisal effect and the modular-lattice effect in current human experiments, and we can further predict human responses to arbitrary network architectures [43]. In exercising the model on simulated data, we expect that humans will be able to learn information most swiftly and accurately on hierarchically modular networks [49], a prediction that can be directly tested in future experiments in both real and artificial learning systems [50]. But first, we turn to the biological apparatus that allows network learning to occur.

### 3. Network constraints on interconnection patterns in neural systems

As our professor sits down to their dishevelled desk to prepare tomorrow's lecture or discussion plan, they may or may not consider the learning organ in the minds of their students. That organ—the human brain—is a richly structured apparatus that has been built to model the world. The acts of building have occurred slowly over evolutionary time scales, and are also modulated within an organism's lifetime by developmental programmes as well as the prevailing forces of the local environment. As with any remarkably useful tool, there exists a systematic map between the physical architecture of the brain and the functions made possible by that architecture. This is not to say that a single function can only be supported optimally by a single structure [51,52], but instead to say that there exist constraints on the class of structures that can or cannot support a given function [53,54]. For example, the structure of synapses between neurons in the nematode *Caenorhabditis elegans* allows for motoric capabilities and mechanized action [55], while the structure of primary afferent connections in the *Drosophila* olfactory system explains odour lateralization behaviour [56]. Similarly, in the human, the connection pattern of white matter tracts linking large-scale brain areas allows for information flow between visual and motor cortices supporting motor skill acquisition [57]. While collating these discrete observations can be useful, it would arguably be more satisfying to identify broad principles that can serve to parameterize the relation between structure and function in neural systems. Here, we briefly review the literature on the network structure of neural systems, and the clear constraints upon it.

#### (a) Energy expenditure and metabolism

The brain evolves, develops and functions under constraints on energy expenditure [58–60]. Early work noted that the shape of neuronal arbours appears to be explained by a minimization of wiring, which in turn minimizes the energy required for synaptic communication in local neural circuits [61,62]. At resolutions larger than the subcellular scale, the principle of wiring minimization also explains why the layout of ganglia in the nematode nervous system requires the least total connection length out of 40 000 000 alternative layouts [63]. Wiring minimization may be balanced by constraints that are topological in nature; for example, early evidence in the rhesus macaque demonstrated that neural networks are more similar to network layouts that minimize the length of processing paths, rather than the length of wires [64,65]. In a sparse network, processing

paths allow two units that are not directly connected to nevertheless communicate across a string of direct paths between serially ordered intermediate units. Minimization of physical lengths or of processing paths leads to a network topology marked by (i) strong local clustering, supporting local processing and (ii) short average path lengths from any point in the network to any other point, supporting global processing [66]. The combination of local clustering and short path lengths is consistent with existing models of small-worldness [67], which in turn are associated with efficient communication in many informational systems spanning technology [68], physics [69], linguistics [70] and biology [23,71].

#### (b) Information processing and computation

It seems sensible to state that optimal information processing requires both local and global components, but it is unclear whether those two constraints are sufficient to produce ideal neural systems [72,73]. Let us consider information transmission as distinct from processing, and note that reasonable architectures to support transmission are bipartite structures [74–76], in which a set of network nodes are strongly and preferentially connected to another set of network nodes, but nodes within a set are not connected to each other [77]. Such bipartite connectivity is observed in neural networks across *C. elegans*, *Drosophila*, the rhesus macaque, the mouse and the human [78], and offers utility in predicting how the activity of neural systems responds to perturbations [79]. Next, consider the potential necessity for information broadcast and receipt; these processes are best supported by core-periphery architectures [80], in which a densely intra-connected set of nodes (the core) extends connections to a sparsely intra-connected set of nodes (the periphery). Core-periphery organization is noted in the structural networks of neural systems across several species [78] as well as in functional brain networks in humans [81–84], allowing for broadcast and receipt functions [85], error prediction [86] and adaptation during learning [87]. Together, small-world organization, bipartitivity and core-periphery structure allow for a diverse array of informational processes that could support the function of neural systems as modellers of our world.

#### (c) Evolution, development, adaptation and learning

A key feature of neural systems that is *not* directly explained by the constraints and structural motifs described thus far is their capacity to evolve, develop and adapt. Evolutionary theory suggests that such adaptability is made possible by structural modularity [88,89], which arises naturally in systems that must satisfy different goals in a changing landscape [90,91]. Moreover, work in both evolutionary biology and evolutionary computer science [92] suggests that hierarchical modularity—the recursive composition of submodules—arises naturally in these same systems when they evolve under constraints for wiring minimization [93]. Hierarchical modularity has been described as the generic architecture of complexity [54], and is observed beyond the neurosciences, in metabolic, ecological and gene regulatory networks, and in human-made systems, such as large organizations and the Internet [93]. The current structure of the human brain is a reflection of evolutionary pressures to optimize neural function and constraints from what other systems and capacities had already developed at each stage of evolution; recent studies suggest that these pressures and constraints naturally guide the system towards

hierarchical modularity [93–98]. In the human brain, hierarchical modularity has been noted in the structural networks linking large-scale areas [65] and in functional networks linking these same areas by shared information [99].

From a psychological perspective, hierarchical modularity is a natural substrate for the separation of cognitive processes [100] and a conduit for the specialization of function in distinct volumes of neural tissue [101]. Yet, it is important to note that not all of the specifics of the early ideas of cognitive or mental modularity withstood the test of time or deeper scientific investigation [100–103]. Those early ideas have been altered and fine-tuned in the light of new empirical data and the capacity to test such theories across large cohorts, for example, in the more than 1000 humans who participated in the Human Connectome Project [104,105]. A recent study used an author-topic model of cognitive functions across 9208 experiments of 77 cognitive tasks to demonstrate a strong spatial correspondence between cognitive functions and brain network modules, suggesting that each module performs a discrete cognitive function [106]. A subsequent study further suggested that specific brain regions tune the connectivity of their neighbouring regions to be more modular while allowing for the integration of task appropriate information across communities, in a manner that facilitates cognitive performance [107]. Such studies lend support to the notion that a map between cognitive modularity and brain modularity does in fact exist, but its specific form may be different from that postulated several decades ago. The existence of such a map also suggests that adaptable brain modules may support adaptable cognition. Indeed, the predicted support of modularity for adaptability is particularly evident in recent work demonstrating that the modules within functional brain networks flexibly reconfigured over time in support of human learning [108–111], planning and reasoning [112], and cognitive flexibility [113]. From a theoretical perspective, the relation between network modularity and adaptable function can be understood in a more mechanistic manner by considering the fact that network architecture directly constrains the trajectories that a system can take through the adaptation landscape [114].

#### 4. Similarities in constraints, leading to hierarchically modular networks

If the professor we have been following understood the architecture of the brain, would that understanding change how they chose the content and structure of their lecture or discussion plan? Most experts could describe, if asked, the direct relations between any pair of the 15 concepts they chose to cover in the class period. In other words, the expert could see the topic as a fully connected graph if they wished; they have all of the requisite knowledge. Yet, an expert can also crystallize that fully connected graph into a sparse network or spanning tree when they wish to use it, or to communicate it; a fully connected network is unlikely to be particularly useful or particularly easy to communicate or apprehend. Which set of important links between ideas should be chosen? Which are sufficient to find a path that connects any pair of ideas in the domain? Should the network architecture of knowledge to be transmitted and the network architecture of the brain inform one another, and if so how and why?

The question brings to mind a passage from Aristotle's *Metaphysics*, where he considers precisely what happens to

the mind when it contemplates. He writes, 'Mind thinks itself because it shares the nature of the object of thought; for it becomes an object of thought in coming into contact with and thinking its objects, so that mind and object of thought are the same.' [115, p. 1072] While the notion that mind and object of thought are the same might initially appear fanciful and rather arcane, there are many metaphors and research programmes that reflect the human intuition that there exists some structural similarity in how we think about knowledge architecture and brain architecture. Moreover, emerging evidence offers preliminary support for one candidate operationalization of precisely the notion that mind and object of thought are intimately connected [39,40,116–118]. When a mind is shown relational knowledge with a specified network architecture, brain activity reflects that architecture in a particular manner. Specifically, the pattern of activity in response to a given item (network node) is similar to the pattern of activity in response to another item (network node) to a degree dictated by the topological distance between the items in the network [117,118]. One could think of this form of representation as one in which the brain represents the inter-item distance as a particular type of relation encoded as a node itself in a labelled graph. In fact, humans appear to organize conceptual knowledge in the brain in a manner that is similar to how they organize spatial knowledge [116], coding topological paths akin to physical distances [40,117]. Suppose that this process of producing patterns of activity whose relations match the relations of the items they represent (or the parts of the world they model) occurs consistently over a human's lifetime, and in fact also over the course of evolution; then what architecture might most effectively underlie the active units to optimize this process?

#### (a) Concordance between modeller and modelled

The terms reflect, model, process, represent and encode are distinct and can each separately help us to understand whether and when the modeller and the modelled are somehow concordant. Does a good apparatus display a form that reflects the form of the material on which it works? Not always; the apparatus for Millikan and Fletcher's 1909 oil-drop experiment has a form far from that of the electron that it is meant to measure [119]. Does a good modeller display a form that reflects the form of the subject to be modelled? Also not always; a 3D printer has a form unlike all of the models that it can build, barring one (itself). Does a representer display a form that reflects the form of the represented? Sometimes; the form of a stationary artist in Times Square or an actor in the West End is the same as the form they represent (although for an alternative view see [120]). Does a processor display a form that reflects the form of the processed? Perhaps; very large scale integrated circuits in computer chips display hierarchically modular structure [65,121], consistent with the structure of information that the chips will represent, manipulate, and store [122]. Does an encoder display a form that reflects the form of the encoded? Efforts in the field of artificial neural networks are continuing to develop architectures and models that can encode the features (both categorical and non-categorical) of an image across hierarchical layers of the neural network. In some cases, the encoding in the artificial system maps to the structure in the real image in an interpretable way [123]: for example, a high-resolution feature is encoded in early layers and a low-resolution feature is encoded in later layers [124–126].

Broadly, across modelling, representing, processing and encoding, the relation between the \*er and the \*ed can differ. Thus, whether it is happenstance or meaningful that brains, learnable networks and knowledge structures have modular architecture depends upon the nature of the relation between brain and knowledge. Much of the current thought in neuroscience and psychology builds upon the notion that the brain's principal purpose is to model [127–134]. Thus, a discussion of the architecture of the brain and the architecture of knowledge would be impoverished without a discussion of the relation (the act of modelling) that can formally link the two. And notably, that relation remains to be clarified; decades of prior work demonstrate that it is non-trivial to successfully represent relational structure in neural systems [14,135–143]. Such representations may depend upon the nature of the relations or the content being related, and may manifest distinctly in the scale accessible to fMRI compared to the scale accessible to cellular imaging. Finally, such representations may also differ across regions [118], being precise reflections of the graph or more akin to a predicate logic.

### (b) Correspondence by relation versus by shared constraint

Does correspondence in architecture tell us something important about the nature of modelling in the brain, thus offering hints regarding explanations and mechanisms for knowledge acquisition? There may exist multiple reasonable answers to this question, and those answers might depend on the specific brain area(s) whose architecture we are considering. Is the given brain area (or the entire brain) a modeller, representer, processor, encoder, or all of the above? First, note that to the degree that the brain *represents* knowledge, correspondence between the network structure of neural representations and the network structure of object relations is perhaps expected based on recent empirical studies [39,40,116–118] (although note that further studies are needed that explicitly compare the neural representations developed in response to different network structures). Second, to the degree that the brain *processes* information, correspondence between the network structure of informational connections and the network structure of the information is also perhaps expected [65,121]. These two correspondences come about owing to the nature of the functions *represent* ('depict', 'constitute', or 'amount to' [144]) and *process* (perform a series of mechanical or chemical operations on something in order to change or preserve it [144]). In both cases, a function can lead to a correspondence in the architecture of the modelled and the modeller. But is the reverse inference accurate? If a correspondence exists in architecture between the modelled and the modeller, can we conclude that the correspondence is owing to a functional relation? Not necessarily. Perhaps the simplest counter example is that a modeller can come into existence under similar constraints to the modelled; in this case, the correspondence in architecture is owing to shared constraints rather than a functional relation.

### (c) Concordant versus discordant constraints

Do there exist shared and divergent constraints on network architecture in the brain and in knowledge (or in the reality to which knowledge maps)? Both the world around us and the world within us must obey the laws of physics, and

therefore exist under marked constraints on energy and tendencies towards entropy. The pressures specifically for wiring minimization—both to conserve energy and to remain adaptable in a changing environment—are pervasive across both natural and human-made systems from genetic regulatory networks to the Internet [93]. Both the brain and the world around it must maintain robustness over evolutionary time scales, a constraint that could explain their shared modular structure [145], and the redundancy evident in distinct elements serving similar functions within the network [146]. Yet the world and the human brain may not be constrained by all of the same factors; while the human species (and therefore the human mind) must reproduce, must the world reproduce? Or must knowledge reproduce? Moreover, knowledge of the world is not exactly the same as the world itself, and therefore the constraints that impinge on the nature of the world might not always perfectly map onto the constraints that impinge on the nature of knowledge. Any discordances in constraints between knowledge networks and brain networks could explain differences in their architecture or function. But perhaps more importantly, divergence between cognitive constructs and neural instantiations could also allow the two systems to function independently; perfect isomorphisms in the topology of two interconnected networks induce system fragility and vulnerability to control.

## 5. Epistemological norms for analysing neural and social epistemes

Are the constraints impinging upon a brain network or a knowledge network relevant beyond a single individual? Certainly, there exists a distinction between individual knowledge and collective knowledge, and a distinction between brain networks and social networks. While we grant the distinction between these entities and the often different analytics required to study them, their interdigitation is crucial to the advancement of relevant scientific and philosophical inquiry. Here, we extend the discussion of individual knowledge patterns and practices to relational and collective knowledge, in keeping with the contemporary philosophical turn towards social epistemology [147,148] and network epistemology [149–151]. These fields take, as their point of departure, the recognition that an individual knower cannot ultimately be isolated from the social environments in which that knower is said to know. Moreover, we extend the discussion of brain networks to social networks, in keeping with the contemporary neuroscientific turn towards social neuroscience [152] and population neuroscience [153]. These fields recognize that individual brain networks shape social networks [154], that social ties in turn shape the brain [154], and that collective knowledge can alter individual cognition, from attentional capacities and memory processes to social perceptions and decisions [155]. We begin our interscale discussion with epistemology.

### (a) An expanding epistemology

Traditional epistemology, as crystallized by reigning accounts in twentieth-century analytic philosophy, makes some assumptions that, while useful under certain conditions, are no longer considered adequate to our epistemic realities. These assumptions include that knowledge is (1) the purview of an individual human (2) whose beliefs, intentions and

propositional attitudes are a critical component of that knowledge. Today, however, it is increasingly important—not to mention useful—to recognize not only the presence of non-human and/or machinic knowers [156,157] and the reality of group or collective knowing [158,159], but what might be called *extended knowing* [160] that traverses knowers of different species, system dynamics and social structures. Such a recognition necessitates, on the one hand, redefining knowledge not as an individual human's justified true belief [161] but in a more generalizable sense as an evidentially supported explanatory model of some elements of a system [149]. Given the ways in which these models are shared, as well as co-constructed, it is equally important to grapple with the biases implicit in knowledge models in both organic [162] and computational systems [163,164]. These various tasks of revisioning epistemology are largely undertaken by the recent subfields of network epistemology and social epistemology. Building on social epistemology's insights into the constitutive effects of social relations, investments and institutions on knowledge itself [147,165–167], network epistemology uses formal network theory to elucidate those constitutive effects [148–150,168–170]. Together, network and social epistemology provide a systems-level approach to the processes of knowledge production, as well as the structural limitations of those processes.

### (b) From representation to network architecture

It is largely recognized, across epistemological literature and the history of science, that knowledge neither resembles nor represents, in the technical sense of these terms, things as they are, but rather interprets and constructs things as they are experienced [171–175]. Network epistemology reframes knowledge as a practice of system modelling or network building. As such, it applies a new frame to classic epistemological issues, including the nature of content and testimony [149], consensus [168], communication structures [150], factionalization [151], belief diffusion [148] and curiosity [9,10]. Understanding knowledge as an increasingly effective network of ideas that models interconnections in the world does not preclude standards of efficacy in knowledge network construction or of elegance in the knowledge network architecture, nor does it preclude standards of correctness in knowledge network acquisition or of effectiveness in knowledge network communication. Network epistemology simply extends the epistemic systems under consideration and the questions that can be asked of them. When defining epistemological norms within this framework, for example, it is important not only to attend to structural characterizations, but also functional and causal characterizations. That is, we must ask, 'What is the architecture of the network?' but also 'What is the function of the network' and 'What are its causes?' Such causes are not always perfectly explained by the system's function, but they can instead be explained by other forces from the system's environment. In computational, collaborative systems, such as brains, computers and human or non-human collectives, questions of function can be explainable as much by optimization requirements as by suboptimal protocols [150,151].

### (c) Model–modeller–modelled

Let us consider an operational distinction between modeller, model and modelled in the context of our topic of interest.

Consider the 'modeller' to be that which models (e.g. the brain), the 'model' to be that which the modeller makes (e.g. the representation of knowledge that the brain produces) and the 'modelled' to be that which is modelled by the modeller (e.g. the information or knowledge present in the human experience). If the brain is taken to model the world, it is incumbent to identify the model–modeller–modelled relationship by which it does so. On the one hand, the form of the model, modeller and modelled may be the same; an example is the actor in the West End. The modelled is Hamlet, the modeller is an actor, and the model is the acted Hamlet. This type of modelling brings to mind the following passage from Rosenblueth & Wiener in their 1945 article in the journal *Philosophy of Science*: 'That is, in a specific example, the best material model for a cat is another, or preferably the same, cat.' [176, p. 320] On the other hand, the form of the model and modelled may be the same, but the form of the modeller may be different; an example is the 3D printer. The modelled is a tree and the model is a tree, but the printer is in no way a tree. What is the best way to categorize the brain, as it builds models of the world by learning knowledge networks?

Any discordant constraints between the two systems might lead us to posit that the model–modeller–modelled relationship that we are facing is of the second sort, where modeller is different from model and modelled. But let us consider for a moment whether we see any evidence for the first sort, where model, modeller and modelled are in some meaningful sense the same. Consider that an optimal learning system (the brain) has a modular architecture that allows it to adapt and change, which is the fundamental essence of learning. And what is the system learning? For a moment, let knowledge refer to the knowledge network present in a single mind; it is a subgraph of the Knowledge network extended across that individual's society, which is in turn a subgraph of the Knowledge network present in the combined humanity of today and yesteryear. Collective knowledge can be viewed as a complex system that also must be able to adapt and change; when we find a new piece of information, it must be possible to add it to the knowledge network without rebuilding the system from scratch. Otherwise, knowledge would not serve its purpose, which is to illuminate the 'veil interposed between reality and the eye of the [mind]' [177], allowing humanity to interact with the world while not perceiving it fully. To the degree that collective knowledge is an adaptable complex system, it must display modular architecture for precisely the same reason that the brain displays modular architecture. Thus, we have evidence for the first type of model–modeller–modelled relationship.

### (d) Systemic, network and modular bias

From the vantage point of social network epistemology, what is known and reflected in a single brain or a network of brains, in a single computational device or a network of computational devices, will never be simply the result of immediate interaction with perceptions, bits, or data points, but always also with structural limitations and sedimented frames [171,178,179]. Knowledge that is created, shared and distributed across network systems will always reflect the history, goals and limitations of those systems. This modular bias is multi-dimensional and multi-vector. In additionally manifesting evolutionary demands across time, modular bias will manifest current and local demands on organic and inorganic systems, as well as competing goals and epistemic factions.

It will also evidence perspectival limitations, including but not limited to inherited and/or algorithmic bias [180], stereotypes [181], structured ignorance and other forms of epistemic injustice [182]. To understand and address issues of modular bias in knowledge network systems and synaptic communication requires increasingly robust work in the politics of human and artificial intelligence, particularly focused on social equity and educational justice.

## 6. Pedagogical principles conducive to curious thought

Developing a deeper understanding of the network architectures of knowledge and knowledge-processing systems such as the brain is of interest in its own right. More than a satisfying intellectual exercise, however, the acquisition of such understanding has the potential to inform and transform our learning environments. As an extension of the robust educational literature exploring the relationship between knowledge networks and social networks [183–187], we posit the relationship between knowledge networks and neural networks as a new pathway for individualizing, optimizing and diversifying pedagogical techniques. Equipped with this knowledge, for example, would our professor sitting at their disheveled desk prepare a different sort of lecture, discussion, or neither? How might we use the existing laboratory experiments in network learning [6,30,39–43,49] to guide best practices in how to present or process information in a way that empowers student learning? As a start, we predict that a modular network architecture underlying information transmission will result in better learning than random or lattice-like architecture, based on the swifter human reaction times observed in visual perceptual learning and visual-motor learning tasks. This prediction could be tested in classroom experiments where a lecture is organized around a set of modularly related concepts versus around a set of linearly related concepts. But beyond the networks studied thus far (only 3 out of the possible  $805\,491\ k=4$  regular architectures of the  $10^{14}$  15-node 30-edge graphs), it is important to distill the optimally learnable graph [49] and to ask whether it has a topology that is common in language or in nature. Is the architecture of the optimally learnable graph also the architecture of a well-written paper or a well-written textbook that effectively communicates networked knowledge to the reader [188]?

### (a) Individualization of knowledge presentation

In exploring the network architectures supporting learnability, we would be remiss if we did not ask, ‘Supporting learnability for whom, how, and in what contexts?’ In the search to calibrate learnability to different systems, with their unique learning capabilities and developmental trajectories, new queries are incumbent. For example, do different humans prefer to learn information on different graph architectures [43] specifically because of their cognitive apparatus, which in turn is constrained by their underlying neural substrate [57,189]? If so, would presenting information to humans in their preferred architecture enhance learning? In experimental neuroscience, it is increasingly clear that marked individual differences exist in many types of learning and associated cognitive processes, such as fear learning [190], social learning [191],

sensorimotor learning [192], language acquisition and processing [193], media multitasking [194] and executive functions [195]. Moreover, it is clear that humans differ in their general statistical learning capacities [196], as well as in their specific network learning capacities [43]. Such different learning capacities, strategies and preferences motivate a careful study of the network architectures of knowledge that are most easily acquired by a given person. Beyond neuro-typical humans, it is possible that those with disabilities, disorders, or other neuro-atypicalities could further benefit from individualization of knowledge presentation. Such a benefit is underscored by the fact that statistical learning as a general mechanism serves as a window into developmental disabilities such as autism spectrum disorder, specific language impairments, Williams’s syndrome and developmental dyslexia [197]. Even more broadly, it is notable that differences and dysfunctions of basic learning mechanisms accompany a wide range of mental disorders including substance abuse, depression and schizophrenia [198]. Future work could seek to explain neuro-typical and neuro-atypical individual differences in network learning by assessing the trajectories of adaptation that are possible from the underlying neural network architecture [114].

### (b) Exemplifying information-seeking

While we frequently learn from information that is presented to us by an external agent whose goal is for us to acquire knowledge, we often learn best when this process stimulates or supervenes upon an internally driven search for information [199–201]. But is this search innate, something we know how to do without any training [202]? Or is it itself learned as we watch our caretakers, our friends and our mentors exemplify curious search [200]? As a set of investigative practices, curiosity is ultimately a tool. Just as animal and human primates deploy hammers as physical tools [203,204], so they use curious search as an intellectual tool. In the context of pedagogy, we need to investigate how curious search can be both facilitated in students and exemplified by instructors. Whether through lectures, group discussions, hands-on activities, or student research, curiosity can be motivated and modelled [9]. From a network learning perspective, instructors can facilitate student curiosity via a random walk search on the knowledge network (moving from disconnected idea to disconnected idea), or a local walk search on the knowledge network (moving from an idea to a tightly related idea). Instructors can also exemplify the richness of other walk topologies (reflecting other curious typologies), such as a Levy walk in which the probability distribution of step-lengths is heavy-tailed [205]. On a flat landscape, the Levy walk can create a small-world network architecture [205]; by contrast, on the existing knowledge network with a non-lattice topology, a Levy walk can create other more nuanced structures [206,207]. By testing the efficacy of different techniques for facilitating and exemplifying patterns of curious thought, we can begin to build a pedagogy that more robustly encourages curiosity, thereby increasing learnability and well-being [208].

### (c) Curious practice as knowledge network building

What is the logical consequence of idiosyncratic information seeking on knowledge networks unfolding over the time scales of months and years? Preferences for seeking



information along certain types of relations, or across specific semantic or conceptual distances, will naturally lead to idiosyncratic architectures of knowledge networks in individual human minds [9]. For example, humans who prefer to close triangles (if A is related to B, and B is related to C, then they want to understand how A is related to C) will naturally build a mesh-like knowledge network architecture. It is interesting to ask whether such individual preferences for styles of knowledge acquisition are evident today or across recent millennia. A recent historical study of the Greek, Latin, German, French and English words for *curiosity* from Plutarch to today demonstrated the existence of at least three key types of curious practice, each characterized by a distinct kinesthetic signature [209]. The *busybody* seeks disconnected bits of information similar to trivia, the *hunter* seeks a specific bit of information in a focused, linear search, and the *dancer* seeks information in local neighbourhoods of knowledge space intermixed with leaps (of analogical or other reasoning) to distant knowledge spaces. Each kinesthetic signature produces a distinct network architecture: respectively a network with many disconnected components, a network with chain-like architecture, and a network with local clustering and long-distance connections, leading to small-world modular architectures [9,10]. Evidence from young children learning the English language supports the notion that such learning is most consistent with the last phenotype, being pocked with gaps in knowledge that are later filled [210]. It would be interesting in future work to determine whether different styles of gappy learning relate to different styles of curiosity [211].

## References

- Kemp C, Tenenbaum JB. 2008 The discovery of structural form. *Proc. Natl Acad. Sci. USA* **105**, 10687–10692. (doi:10.1073/pnas.0802631105)
- Lake BM, Lawrence ND, Tenenbaum JB. 2018 The emergence of organizing structure in conceptual representation. *Cogn. Sci.* **42** (Suppl 3), 809–832. (doi:10.1111/cogs.12580)
- Stahl AE, Romberg AR, Roseberry S, Golinkoff RM, Hirsh-Pasek K. 2014 Infants segment continuous events using transitional probabilities. *Child Dev.* **85**, 1821–1826. (doi:10.1111/cdev.12217)
- Saffran JR. 2003 Musical learning and language development. *Ann. N Y Acad. Sci.* **999**, 397–401. (doi:10.1196/annals.1284.050)
- Karuza EA, Emberson LL, Roser ME, Cole D, Aslin RN, Fiser J. 2017 Neural signatures of spatial statistical learning: characterizing the extraction of structure from complex visual scenes. *J. Cogn. Neurosci.* **29**, 1963–1976. (doi:10.1162/jocn\_a\_01182)
- Tompson SH, Kahn AE, Falk EB, Vettel JM, Bassett DS. 2019 Individual differences in learning social and nonsocial network structures. *J. Exp. Psychol. Learn. Mem. Cogn.* **45**, 253–271. (doi:10.1037/xlm0000580)
- Redington M. 1998 Distributional information: a powerful cue for acquiring syntactic categories. *Cogn. Sci.* **22**, 425–469. (doi:10.1207/s15516709cog2204\_2)
- Hirsh-Pasek K, Kemler Nelson DG, Jusczyk PW, Cassidy KW, Druss B, Kennedy L. 1987 Clauses are perceptual units for young infants. *Cognition* **26**, 269–286. (doi:10.1016/S0010-0277(87)80002-1)
- Bassett DS. 2020 A network science of the practice of curiosity. In *Curiosity studies: toward a new ecology of knowledge*. Minneapolis, MN: University of Minnesota Press. (See <https://www.upress.umn.edu/book-division/books/curiosity-studies>)
- Zurn P, Bassett DS. 2018 On curiosity: a fundamental aspect of personality, a practice of network growth. *Pers. Neurosci.* **1**, e13. (doi:10.1017/pen.2018.3)
- Poincaré H. 1902 *Science and hypothesis*. New York, NY: Walter Scott Publishing.
- Dewey J 1916 *Democracy and education*. New York, NY: Macmillan.
- Christie S, Gentner D. 2010 Where hypotheses come from: learning new relations by structural alignment. *J. Cogn. Dev.* **11**, 356–373. (doi:10.1080/15248371003700015)
- Doumas LA, Hummel JE, Sandhofer CM. 2008 A theory of the discovery and predication of relational concepts. *Psychol. Rev.* **115**, 1–43. (doi:10.1037/0033-295X.115.1.1)
- Lu H, Chen D, Holyoak KJ. 2012 Bayesian analogy with relational transformations. *Psychol. Rev.* **119**, 617–648. (doi:10.1037/a0028719)
- Azarfar A, Calcini N, Huang C, Zeldenrust F, Celikel T. 2018 Neural coding: a single neuron's perspective. *Neurosci. Biobehav. Rev.* **94**, 238–247. (doi:10.1016/j.neubiorev.2018.09.007)
- Petersen SE, Fiez JA, Corbetta M. 1992 Neuroimaging. *Curr. Opin Neurobiol.* **2**, 217–222. (doi:10.1016/0959-4388(92)90016-E)
- Corbetta M, Siegel JS, Shulman GL. 2018 On the low dimensionality of behavioral deficits and alterations of brain network connectivity after focal injury. *Cortex* **107**, 229–237. (doi:10.1016/j.cortex.2017.12.017)
- Cummings JL. 1993 Frontal-subcortical circuits and human behavior. *Arch. Neurol.* **50**, 873–880. (doi:10.1001/archneur.1993.00540080076020)
- Felleman DJ, Van Essen DC. 1991 Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47. (doi:10.1093/cercor/1.1.1)
- Pulvermuller F, Garagnani M, Wenekers T. 2014 Thinking in circuits: toward neurobiological explanation in cognitive neuroscience. *Biol. Cybern.* **108**, 573–593. (doi:10.1007/s00422-014-0603-9)
- Bassett DS, Sporns O. 2017 Network neuroscience. *Nat. Neurosci.* **20**, 353–364. (doi:10.1038/nn.4502)
- Bassett DS, Bullmore E. 2006 Small-world brain networks. *Neuroscientist* **12**, 512–523. (doi:10.1177/1073858406293182)

24. Bullmore E, Sporns O. 2009 Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198. (doi:10.1038/nrn2575)
25. Bassett DS, Zurn P, Gold JI. 2018 On the nature and use of models in network neuroscience. *Nat. Rev. Neurosci.* **19**, 1–13. (doi:10.1038/s41583-018-0038-8)
26. Schwann T. 1839 *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachstum der Thiere und Pflanzen*. Berlin, Germany: Sander.
27. Jones EG. 1999 Colgi, cajal and the neuron doctrine. *J. Hist. Neurosci.* **8**, 170–178. (doi:10.1076/jhin.8.2.170.1838)
28. Saffran JR, Kirkham NZ. 2018 Infant statistical learning. *Annu. Rev. Psychol.* **69**, 181–203. (doi:10.1146/annurev-psych-122216-011805)
29. Fiser J, Aslin RN. 2001 Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* **12**, 499–504. (doi:10.1111/1467-9280.00392)
30. Tompson SH, Kahn AE, Falk EB, Vettel JM, Bassett DS. 2020 Functional brain network architecture supporting the learning of social networks in humans. *NeuroImage* **210**, 116498. (doi:10.1016/j.neuroimage.2019.116498)
31. Endress AD, Scholl BJ, Mehler J. 2005 The role of salience in the extraction of algebraic rules. *J. Exp. Psychol. Gen.* **134**, 406–419. (doi:10.1037/0096-3445.134.3.406)
32. Frick RW, Lee YS. 1995 Implicit learning and concept learning. *Q. J. Exp. Psychol. A* **48**, 762–782. (doi:10.1080/14640749508401414)
33. Saffran JR. 2003 Statistical language learning: mechanisms and constraints. *Curr. Dir. Psychol. Sci.* **12**, 110–114. (doi:10.1111/1467-8721.01243)
34. Santolin C, Saffran JR. 2018 Constraints on statistical learning across species. *Trends Cogn. Sci.* **22**, 52–63. (doi:10.1016/j.tics.2017.10.003)
35. Karuza EA, Thompson-Schill SL, Bassett DS. 2016 Local patterns to global architectures: influences of network topology on human learning. *Trends Cogn. Sci.* **51364-6613**, 30 071–30 077.
36. Newport EL, Aslin RN. 2004 Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **48**, 127–162. (doi:10.1016/S0010-0285(03)00128-2)
37. Newport EL, Hauser MD, Spaepen G, Aslin RN. 2004 Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* **49**, 85–117. (doi:10.1016/j.cogpsych.2003.12.002)
38. Hesse MB. 1955 Action at a distance in classical physics. *Isis* **46**, 337–353. (doi:10.1086/348429)
39. Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. 2013 Neural representations of events arise from temporal community structure. *Nat. Neurosci.* **16**, 486–492. (doi:10.1038/nn.3331)
40. Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM. 2016 Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**, 3–8. (doi:10.1002/hipo.22523)
41. Karuza EA, Kahn AE, Thompson-Schill SL, Bassett DS. 2017 Process reveals structure: how a network is traversed mediates expectations about its architecture. *Sci. Rep.* **7**, 12733. (doi:10.1038/s41598-017-12876-5)
42. Kahn AE, Karuza EA, Vettel JM, Bassett DS. 2018 Network constraints on learnability of probabilistic motor sequences. *Nat. Hum. Behav.* **2**, 936–947. (doi:10.1038/s41562-018-0463-8)
43. Lynn CW, Kahn AE, Bassett DS. 2018 Structure from noise: mental errors yield abstract representations of events. *arXiv*, vol. 1805, p. 12491. (doi:10.32470/CCN.2018.1169-0)
44. Fodor JA 1975 *The language of thought*, vol. 5. Cambridge, MA: Harvard University Press.
45. Fodor JA 1981 *Representations—philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.
46. Doumas LAA, Martin AE. 2018 Learning structured representations from experience. *Psychol. Learn. Motiv.* **69**, 165–203. (doi:10.1016/bs.plm.2018.10.002)
47. Danks D 2014 *Unifying the mind: cognitive representation as graphical models*. Cambridge, MA: MIT Press.
48. Aristotle ca. 350 B.C. Prior analytics and posterior analytics. In *The New Aristotle Reader* 1988 (ed. JL Ackrill), pp. 24–60. Princeton, NJ: Princeton University Press.
49. Lynn CW, Papadopoulos L, Kahn AE, Bassett DS. 2019 Human information processing in complex networks. (See <https://arxiv.org/abs/1906.00926>)
50. Lu Z, Bassett DS. 2018 A parsimonious dynamical model for structural learning in the human brain. *arXiv*, 1807.05214. (<https://arxiv.org/abs/1807.05214>)
51. Prinz AA, Bucher D, Marder E. 2004 Similar network activity from disparate circuit parameters. *Nat. Neurosci.* **7**, 1345–1352. (doi:10.1038/nn1352)
52. Calabrese RL. 2018 Inconvenient truth to principle of neuroscience. *Trends Neurosci.* **41**, 488–491. (doi:10.1016/j.tins.2018.05.006)
53. Alon U. 2007 Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461. (doi:10.1038/nrg2102)
54. Simon H. 1962 The architecture of complexity. *Am. Phil. Soc.* **106**, 467–482.
55. Yan G, Vertes PE, Towilson EK, Chew YL, Walker DS, Schafer WR, Barabasi AL. 2017 Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* **550**, 519–523. (doi:10.1038/nature24056)
56. Tobin WF, Wilson RI, Lee WA. 2017 Wiring variations that enable and constrain neural computation in a sensory microcircuit. *Elife* **6**, e24838. (doi:10.7554/eLife.24838)
57. Kahn AE, Mattar MG, Vettel JM, Wymbs NF, Grafton ST, Bassett DS. 2017 Structural pathways supporting swift acquisition of new visuo-motor skills. *Cereb. Cortex* **27**, 173–184. (doi:10.1093/cercor/bhw335)
58. Niven JE. 2016 Neuronal energy consumption: biophysics, efficiency and evolution. *Curr. Opin Neurobiol.* **41**, 129–135. (doi:10.1016/j.conb.2016.09.004)
59. Niven JE, Laughlin SB. 2008 Energy limitation as a selective pressure on the evolution of sensory systems. *J. Exp. Biol.* **211**(Pt 11), 1792–1804. (doi:10.1242/jeb.017574)
60. Laughlin SB. 2001 Energy as a constraint on the coding and processing of sensory information. *Curr. Opin Neurobiol.* **11**, 475–480. (doi:10.1016/S0959-4388(00)00237-3)
61. Chermiak C. 1992 Local optimization of neuron arbors. *Biol. Cybern.* **66**, 503–510. (doi:10.1007/BF00204115)
62. Karbowski J. 2001 Optimal wiring principle and plateaus in the degree of separation for cortical neurons. *Phys. Rev. Lett.* **86**, 3674–3677. (doi:10.1103/PhysRevLett.86.3674)
63. Chermiak C. 1994 Component placement optimization in the brain. *J. Neurosci.* **14**, 2418–2427. (doi:10.1523/JNEUROSCI.14-04-02418.1994)
64. Kaiser M, Hilgetag CC. 2006 Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* **2**, e95. (doi:10.1371/journal.pcbi.0020095)
65. Bassett DS, Greenfield DL, Meyer-Lindenberg A, Weinberger DR, Moore SW, Bullmore ET. 2010 Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput. Biol.* **6**, e1000748. (doi:10.1371/journal.pcbi.1000748)
66. Betzel RF, Bassett DS. 2018 The specificity and robustness of long-distance connections in weighted, interareal connectomes. *Proc. Natl Acad. Sci. USA* **115**, E4880–E4889. (doi:10.1073/pnas.1720186115)
67. Watts DJ, Strogatz SH. 1998 Collective dynamics of ‘small-world-networks’. *Nature* **393**, 440–442. (doi:10.1038/30918)
68. Latora V, Marchiori M. 2001 Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87**, 198701. (doi:10.1103/PhysRevLett.87.198701)
69. Mulken O, Blumen A. 2006 Efficiency of quantum and classical transport on graphs. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **73**(Pt 2), 066117. (doi:10.1103/PhysRevE.73.066117)
70. Sheppard JP, Wang JP, Wong PC. 2012 Large-scale cortical network properties predict future sound-to-word learning success. *J. Cogn. Neurosci.* **24**, 1087–1103. (doi:10.1162/jocn\_a\_00210)
71. Bassett DS, Bullmore ET. 2017 Small-world brain networks revisited. *Neuroscientist* **23**, 499–516. (doi:10.1177/1073858416667720)
72. Bullmore E, Sporns O. 2012 The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349. (doi:10.1038/nrn3214)
73. Avena-Koenigsberger A, Misis B, Sporns O. 2017 Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* **19**, 17–33. (doi:10.1038/nrn.2017.149)

74. Hernandez DG, Risau-Gusman S. 2013 Epidemic thresholds for bipartite networks. *Phys. Rev. E* **88**, 052801. (doi:10.1103/PhysRevE.88.052801)
75. Lu L, Liu W. 2011 Information filtering via preferential diffusion. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **83**(Pt 2), 066119. (doi:10.1103/PhysRevE.83.066119)
76. Wang P, Hunter T, Bayen AM, Schechtner K, Gonzalez MC. 2012 Understanding road usage patterns in urban areas. *Sci. Rep.* **2**, 1001. (doi:10.1038/srep01001)
77. Bollobas B 1979 *Graph theory: an introductory course*. New York, NY: Springer.
78. Betzel RF, Medaglia JD, Bassett DS. 2018 Diversity of meso-scale architecture in human and non-human connectomes. *Nat. Commun.* **9**, 346. (doi:10.1038/s41467-017-02681-z)
79. Kim JZ, Soffer JM, Kahn AE, Vettel JM, Pasqualetti F, Bassett DS. 2018 Role of graph architecture in controlling dynamical networks with applications to neural systems. *Nat. Phys.* **14**, 91–98. (doi:10.1038/nphys4268)
80. Borgatti SP, Everett MG. 2000 Models of core/periphery structures. *Soc. Netw.* **21**, 375–395. (doi:10.1016/S0378-8733(99)00019-2)
81. Gollo LL, Zalesky A, Hutchison RM, van den Heuvel M, Breakspear M. 2015 Dwelling quietly in the rich club: brain network determinants of slow cortical fluctuations. *Phil. Trans. R. Soc. B* **370**, 1668. (doi:10.1098/rstb.2014.0165)
82. Gollo LL, Roberts JA, Cocchi L. 2017 Mapping how local perturbations influence systems-level brain dynamics. *Neuroimage* **160**, 97–112. (doi:10.1016/j.neuroimage.2017.01.057)
83. Gu S, Yang M, Medaglia JD, Gur RC, Gur RE, Satterthwaite TD, Bassett DS. 2017 Functional hypergraph uncovers novel covariant structures over neurodevelopment. *Hum. Brain Mapp.* **38**, 3823–3835. (doi:10.1002/hbm.23631)
84. Battiston F, Guillon J, Chavez M, Latora V, De Vico Fallani F. 2018 Multiplex core–periphery organization of the human connectome. *J. R. Soc. Interface* **15**, 146. (doi:10.1098/rsif.2018.0514)
85. Mantzaris AV, Bassett DS, Wymbs NF, Estrada E, Porter MA, Mucha PJ, Grafton ST, Higham DJ. 2013 Dynamic network centrality summarizes learning in the human brain. *J. Complex Netw.* **1**, 83–92. (doi:10.1093/comnet/cnt001)
86. Ekman M, Derrfuss J, Tittgemeyer M, Fiebach CJ. 2012 Predicting errors from reconfiguration patterns in human brain networks. *Proc. Natl Acad. Sci. USA* **109**, 16714–16719. (doi:10.1073/pnas.1207523109)
87. Bassett DS, Wymbs NF, Rombach MP, Porter MA, Mucha PJ, Grafton ST. 2013 Task-based core–periphery organization of human brain dynamics. *PLoS Comput. Biol.* **9**, e1003171. (doi:10.1371/journal.pcbi.1003171)
88. Kirschner M, Gerhart J. 1998 Evolvability. *Proc. Natl Acad. Sci. USA* **95**, 8420–8427. (doi:10.1073/pnas.95.15.8420)
89. Schlosser G, Wagner GP 2004 *Modularity in development and evolution*. Chicago, IL: Chicago University Press.
90. Kashtan N, Alon U. 2005 Spontaneous evolution of modularity and network motifs. *Proc. Natl Acad. Sci. USA* **102**, 13 773–13 778. (doi:10.1073/pnas.0503610102)
91. Chen Y, Wang S, Hilgetag CC, Zhou C. 2013 Trade-off between multiple constraints enables simultaneous formation of modules and hubs in neural systems. *PLoS Comput. Biol.* **9**, e1002937. (doi:10.1371/journal.pcbi.1002937)
92. Wagner GP, Altenberg L. 1996 Complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976. (doi:10.1111/j.1558-5646.1996.tb02339.x)
93. Mengistu H, Huizinga J, Mouret JB, Clune J. 2016 The evolutionary origins of hierarchy. *PLoS Comput. Biol.* **12**, e1004829. (doi:10.1371/journal.pcbi.1004829)
94. Rubinov M, Sporns O, Thivierge J-P, Breakspear M. 2011 Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. *PLoS Comput. Biol.* **7**, e1002038. (doi:10.1371/journal.pcbi.1002038)
95. Yamamoto H, Moriya S, Ide K, Hayakawa T, Akima H, Sato S, Kubota S, Tani T, Niwano M, Teller S, Soriano J, Hirano-Iwata A. 2018 Impact of modular organization on dynamical richness in cortical networks. *Sci. Adv.* **4**, eaau4914. (doi:10.1126/sciadv.aau4914)
96. Rodriguez N, Izquierdo E, Ahn YY. 2019 Optimal modularity and memory capacity of neural reservoirs. *Netw. Neurosci.* **3**, 551–566. (doi:10.1162/netn\_a\_00082)
97. Caetano-Anolles G, Aziz MF, Mughal F, Grater F, Koc I, Caetano-Anolles K, Caetano-Anolles D. 2019 Emergence of hierarchical modularity in evolving networks uncovered by phylogenomic analysis. *Evol. Bioinform.* **15**, 1176934319872980. (doi:10.1177/1176934319872980)
98. Espinosa-Soto C. 2018 On the role of sparseness in the evolution of modularity in gene regulatory networks. *PLoS Comput. Biol.* **14**, e1006172. (doi:10.1371/journal.pcbi.1006172)
99. Meunier D, Lambiotte R, Bullmore ET. 2010 Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* **4**, 200. (doi:10.3389/fnins.2010.00200)
100. Fodor JA. 1983 *Modularity of mind*. Massachusetts, MA: MIT.
101. Sporns O, Betzel RF. 2016 Modular brain networks. *Annu. Rev. Psychol.* **67**, 613–640. (doi:10.1146/annurev-psych-122414-033634)
102. Carruthers P 2006 *The architecture of the mind: massive modularity and the flexibility of thought*. Oxford, UK: Oxford University Press.
103. Samuels R. 1998 Evolutionary psychology and the massive modularity hypothesis. *Br. J. Phil. Sci.* **49**, 575–602. (doi:10.1093/bjps/49.4.575)
104. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium W-MH. 2013 The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79. (doi:10.1016/j.neuroimage.2013.05.041)
105. Glasser MF, al. et. 2016 The human connectome project's neuroimaging approach. *Nat. Neurosci.* **19**, 1175–1187. (doi:10.1038/nn.4361)
106. Bertolero MA, Yeo BT, D'Esposito M. 2015 The modular and integrative functional architecture of the human brain. *Proc. Natl Acad. Sci. USA* **112**, E6798–E6807. (doi:10.1073/pnas.1510619112)
107. Bertolero MA, Yeo BTT, Bassett DS, D'Esposito M. 2018 A mechanistic model of connector hubs, modularity and cognition. *Nat. Hum. Behav.* **2**, 765–777. (doi:10.1038/s41562-018-0420-6)
108. Bassett DS, Wymbs NF, Porter MA, Mucha PJ, Carlson JM, Grafton ST. 2011 Dynamic reconfiguration of human brain networks during learning. *Proc. Natl Acad. Sci. USA* **108**, 7641–7646. (doi:10.1073/pnas.1018985108)
109. Bassett DS, Yang M, Wymbs NF, Grafton ST. 2015 Learning-induced autonomy of sensorimotor systems. *Nat. Neurosci.* **18**, 744–751. (doi:10.1038/nn.3993)
110. Mattar MG, Wymbs NF, Bock AS, Aguirre GK, Grafton ST, Bassett DS. 2018 Predicting future learning from baseline network architecture. *Neuroimage* **172**, 107–117. (doi:10.1016/j.neuroimage.2018.01.037)
111. Gerraty RT, Davidow JY, Foerster K, Galvan A, Bassett DS, Shohamy D. 2018 Dynamic flexibility in striatal-cortical circuits supports reinforcement learning. *J. Neurosci.* **38**, 2442–2453. (doi:10.1523/JNEUROSCI.2084-17.2018)
112. Pedersen M, Zalesky A, Omidvarnia A, Jackson GD. 2018 Multilayer network switching rate predicts brain performance. *Proc. Natl Acad. Sci. USA* **115**, 13 376–13 381. (doi:10.1073/pnas.1814785115)
113. Braun U et al. 2015 Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl Acad. Sci. USA* **112**, 11 678–11 683. (doi:10.1073/pnas.1422487112)
114. Helsen J, Frickel J, Jelier R, Verstreppe KJ. 2019 Network hubs affect evolvability. *PLoS Biol.* **17**, e3000111. (doi:10.1371/journal.pbio.3000111)
115. Aristotle ca. 350 B.C. *Metaphysics*. In *The New Aristotle Reader* 1988 (ed. JL Ackrill), Book XII, ch. 7, p. 1072. Princeton, NJ: Princeton University Press.
116. Constantinescu AO, O'Reilly JX, Behrens TEJ. 2016 Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468. (doi:10.1126/science.aaf0941)
117. Garvert MM, Dolan RJ, Behrens TE. 2017 A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife* **6**, e17086. (doi:10.7554/eLife.17086)
118. Henin S, Turk-Browne N, Friedman D, Liu A, Dugan P, Flinker A, Doyle W, Devinsky O, Melloni L. 2019 Statistical learning shapes neural sequence representations. *BioRxiv* **583856**. (doi:10.1101/583856)

119. Millikan RA. 1913 On the elementary electric charge and the avogadro constant. *Phys. Rev.* **2**, 109–143. (doi:10.1103/PhysRev.2.109)
120. Lamb C 1986 On the tragedies of Shakespeare, considered with reference to their fitness for stage. *Reflector* **1811**, 97–111.
121. Bakoglu HB 1990 *Circuits, interconnections, and packaging for VLSI*. Boston, MA: Addison Wesley.
122. Liu J, Wang J, Zheng Q, Zhang W, Jiang L. 2012 Knowledge-based systems. *Knowl. Based Syst.* **36**, 260–267. (doi:10.1016/j.knsys.2012.07.011)
123. Ertekin S, Rudin C. 2015 A Bayesian approach to learning scoring systems. *Big Data* **3**, 267–276. (doi:10.1089/big.2015.0033)
124. Khaligh-Razavi SM, Henriksen L, Kay K, Kriegeskorte N. 2017 Fixed versus mixed RSA: explaining visual representations by fixed and mixed feature sets from shallow and deep computational models. *J. Math. Psychol.* **76**(Pt B), 184–197. (doi:10.1016/j.jmp.2016.10.007)
125. Jozwik KM, Kriegeskorte N, Storrs KR, Mur M. 2017 Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Front. Psychol.* **8**, 1726. (doi:10.3389/fpsyg.2017.01726)
126. St-Yves G, Naselaris T. 2018 The feature-weighted receptive field: an interpretable encoding model for complex feature spaces. *Neuroimage* **180**(Pt A), 188–202. (doi:10.1016/j.neuroimage.2017.06.035)
127. Vilares I, Kording K. 2011 Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann. N Y Acad. Sci.* **1224**, 22–39. (doi:10.1111/j.1749-6632.2011.05965.x)
128. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. 2011 How to grow a mind: statistics, structure, and abstraction. *Science* **331**, 1279–1285. (doi:10.1126/science.1192788)
129. Parr T, Friston KJ. 2018 The anatomy of inference: generative models and brain structure. *Front. Comput. Neurosci.* **12**, 90. (doi:10.3389/fncom.2018.00090)
130. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. 2003 Temporal difference models and reward-related learning in the human brain. *Neuron* **38**, 329–337. (doi:10.1016/S0896-6273(03)00169-7)
131. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND. 2015 Model-based choices involve prospective neural activity. *Nat. Neurosci.* **18**, 767–772. (doi:10.1038/nn.3981)
132. Kriegeskorte N, Diedrichsen J. 2019 Peeling the onion of brain representations. *Annu. Rev. Neurosci.* **42**, 407–432. (doi:10.1146/annurev-neuro-080317-061906)
133. Kriegeskorte N, Diedrichsen J. 2016 Inferring brain-computational mechanisms with models of activity measurements. *Phil. Trans. R. Soc. B* **371**, 20160278. (doi:10.1098/rstb.2016.0278)
134. Mack ML, Love BC, Preston AR. 2018 Building concepts one episode at a time: the hippocampus and concept formation. *Neurosci. Lett.* **680**, 31–38. (doi:10.1016/j.neulet.2017.07.061)
135. Hummel JE, Biederman I. 1992 Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* **99**, 480–517. (doi:10.1037/0033-295X.99.3.480)
136. Doumas LAA, Hummel JE. 2005 Approaches to modeling human mental representations: what works, what doesn't, and why. In *The Cambridge handbook of thinking and reasoning* (eds KJ Holyoak, RG Morrison), pp. 73–91. Cambridge, UK: Cambridge University Press.
137. Doumas LAA, Hummel JE. 2012 Computational models of higher cognition. In *The Oxford handbook of thinking and reasoning* (eds KJ Holyoak, RG Morrison), pp. 52–66. Oxford, UK: Oxford University Press.
138. Halford GS, Wilson WH, Phillips S. 1998 Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behav. Brain Sci.* **21**, 803–864. (doi:10.1017/S0140525X98001769)
139. Halford GS, Wilson WH, Phillips S. 2010 Relational knowledge: the foundation of higher cognition. *Trends Cogn. Sci.* **14**, 497–505. (doi:10.1016/j.tics.2010.08.005)
140. Hummel JE, Holyoak KJ. 1997 Distributed representations of structure: a theory of analogical access and mapping. *Psychol. Rev.* **104**, 427–466. (doi:10.1037/0033-295X.104.3.427)
141. Hummel JE, Holyoak KJ. 2003 A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* **110**, 220–264. (doi:10.1037/0033-295X.110.2.220)
142. von der Malsburg C. 1981 The correlation theory of brain function. *MPI Biophysical Chemistry, Internal report*, pp. 81–82. Reprinted in *Models of Neural Networks II* 1994 (eds E Domany, JL van Hemmen, K Schulten), pp. 95–119. Berlin, Germany: Springer.
143. von der Malsburg C. 1999 The what and why of binding: the modeler's perspective. *Neuron* **24**, P95–P104. (doi:10.1016/S0896-6273(00)80825-9)
144. Simpson JA, Weiner ESC. 1989 *The Oxford English Dictionary*. Oxford, UK: Oxford University Press.
145. Felix MA, Wagner A. 1998 Robustness and evolution: concepts, insights, and challenges from a developmental model system. *Heredity* **100**, 132–140. (doi:10.1038/sj.hdy.6800915)
146. Cropper EC, Dacks AM, Weiss KR. 2016 Consequences of degeneracy in network function. *Curr. Opin Neurobiol.* **41**, 62–67. (doi:10.1016/j.conb.2016.07.008)
147. Fricker M, Graham PJ, Henderson D, L. Pedersen NJL 2019 *The Routledge handbook of social epistemology*. New York, NY: Routledge.
148. Goldman A 2019 *Social epistemology – stanford encyclopedia of philosophy*. Stanford, CA: Stanford University Press.
149. Humphreys P. 2009 Network epistemology. *Episteme* **6**, 221–229. (doi:10.3366/E1742360009000653)
150. Zollman KJS. 2013 Network epistemology: communication in epistemic communities. *Phil. Compass* **8**, 15–27. (doi:10.1111/j.1747-9991.2012.00534.x)
151. Weatherall J, O'Connor C. 2018 Endogenous epistemic factionalization: a network epistemology approach. Available at SSRN: <https://ssrn.com/abstract=3304109>. (doi:10.2139/ssrn.3304109)
152. Cacioppo JT, Decety J. 2011 Social neuroscience: challenges and opportunities in the study of complex behavior. *Ann. N Y Acad. Sci.* **1224**, 162–173. (doi:10.1111/j.1749-6632.2010.05858.x)
153. Falk EB, al. et. 2013 What is a representative brain? Neuroscience meets population science. *Proc. Natl Acad. Sci. USA* **110**, 17615–17622. (doi:10.1073/pnas.1310134110)
154. Falk EB, Bassett DS. 2017 Brain and social networks: fundamental building blocks of human experience. *Trends Cogn. Sci.* **21**, 674–690. (doi:10.1016/j.tics.2017.06.009)
155. Firth J, al. et. 2019 The 'online brain': how the Internet may be changing our cognition. *World Psychiatry* **18**, 119–129. (doi:10.1002/wps.20617)
156. Haraway D 2008 *When species meet*. Minneapolis, MN: University of Minnesota Press.
157. Haraway D. 1985 A manifesto for cyborgs. *Soc. Rev.* **15**, 65–107.
158. Gilbert M 2013 *Collective epistemology*. Oxford, UK: Oxford University Press.
159. Hans Bernhard Schmid DS, Weber M eds. 2011 *Collective epistemology*. Berlin, Germany: De Gruyter.
160. Karachalios K, Ito J. 2018 Human intelligence and autonomy in the era of 'extended intelligence'. *Council on extended Intelligence*. See [https://globalcxi.org/wp-content/uploads/CXI\\_Essay.pdf](https://globalcxi.org/wp-content/uploads/CXI_Essay.pdf).
161. Gettier E. 1963 Is justified true belief knowledge? *Analysis* **23.6**, 121–123. (doi:10.1093/analys/23.6.121)
162. Brownstein M, Saul J 2016 *Implicit bias and philosophy*. Oxford, UK: Oxford University Press.
163. Friedman B, Nissenbaum H. 1996 Bias in computer systems. *ACM Trans. Inf. Syst.* **14**, 330–347. (doi:10.1145/230538.230561)
164. Tschider CA. 2018 Regulating the internet of things: discrimination, privacy, and cybersecurity in the artificial intelligence age. *Denver Law Rev.* **96**, 87–143. (doi:10.2139/ssrn.3129557)
165. Goldman A 1999 *Knowledge in a social world*. Oxford, UK: Oxford University Press.
166. Goldman A, Whitcomb D 2011 *Social epistemology: essential readings*. Oxford, UK: Oxford University Press.
167. Lackey J 2014 *Essays in collective epistemology*. Oxford, UK: Oxford University Press.
168. Zollman KJS. 2012 Social network structure and the achievement of consensus. *Politics, Phil. Econ.* **11**, 26–44. (doi:10.1177/1470594X11416766)
169. Rosenstock S, Bruner J, O'Connor C. 2017 In epistemic networks, is less really more? *Phil. Sci.* **84**, 234–252. (doi:10.1086/690717)
170. O'Connor C, Weatherall JO. 2018 Scientific polarization. *Eur. J. Phil. Sci.* **8**, 855–875. (doi:10.1007/s13194-018-0213-9)
171. Kuhn T 1962 *The Structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
172. Foucault M 1970 *The order of things*. New York, NY: Pantheon.
173. Hacking I 2000 *The social construction of what?* Cambridge, MA: Harvard University Press.
174. Daston L, Galison P. 2007 *Objectivity*. Cambridge, MA: Zone Books.
175. Frigg R. 2010 Friction and scientific representation. In *Beyond Mimesis and Nominalism: Representation*

- in *Art and Science* (eds R Frigg, M Hunter), pp. 97–138. Springer Netherlands.
176. Rosenblueth A, Wiener N. 1945 The role of models in science. *Phil. Sci.* **12**, 316–321. (doi:10.1086/286874)
  177. Poincaré H 1946 *The value of science*. New York, NY: The Science Press.
  178. Fricker M 2007 *Epistemic injustice: power and the ethics of knowing*. Oxford, UK: Oxford University Press.
  179. Steele C 2011 *Whistling Vivaldi: how stereotypes affect us and what we can do*. New York, NY: W. W. Norton & Company.
  180. Danks D, London AJ. 2017 Algorithmic bias in autonomous systems. In *Proc. of the 26th Int. Joint Conf. on Artificial Intelligence*, pp. 4691–4697. (doi:10.24963/ijcai.2017/654)
  181. Caliskan BJJ, Narayanan A. 2017 Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186. (doi:10.1126/science.aal4230)
  182. Chin-Yee B, Upshur R. 2019 Three problems with big data and artificial intelligence in medicine. *Perspect. Biol. Med.* **62**, 237–256. (doi:10.1353/pbm.2019.0012)
  183. Illich I 1971 *Deschooling society*. London, UK: Marion Boyars Publishers.
  184. Chivers GE, Poell RF, Van der Krogt FJ, Wildemeersch D. 2000 Learning-network theory: organizing the dynamic relationships between learning and work. *Manage. Learn.* **31**, 25–49. (doi:10.1177/1350507600311004)
  185. d. Krogt FJ v. 2006 Learning network theory: the tension between learning systems and work systems in organizations. *Hum. Resour. Dev. Q.* **9**, 157–177. (doi:10.1002/hrdq.3920090207)
  186. Daly AJ 2010 *Social network theory and educational change*. Cambridge, MA: Harvard University Press.
  187. Carvalho L, Goodyear P 2014 *The architecture of productive learning networks*. New York, NY: Routledge.
  188. Chai LR, Bassett DS. 2018 Evolution of semantic networks in biomedical texts. *arXiv* **1810**, 10534. (doi:10.1093/comnet/cnz023)
  189. Roberts RE, Anderson EJ, Husain M. 2013 White matter microstructure and cognitive function. *Neuroscientist* **19**, 8–15. (doi:10.1177/1073858411421218)
  190. Gershman SJ, Hartley CA. 2015 Individual differences in learning predict the return of fear. *Learn. Behav.* **43**, 243–250. (doi:10.3758/s13420-015-0176-z)
  191. Mesoudi A, Chang L, Dall SRX, Thornton A. 2016 The evolution of individual and cultural variation in social learning. *Trends Ecol. Evol.* **31**, 215–225. (doi:10.1016/j.tree.2015.12.012)
  192. Seidler RD, Carson RG. 2017 Sensorimotor learning: neurocognitive mechanisms and individual differences. *J. Neuroeng. Rehabil.* **14**, 74. (doi:10.1186/s12984-017-0279-1)
  193. Kidd E, Donnelly S, Christiansen MH. 2018 Individual differences in language acquisition and processing. *Trends Cogn. Sci.* **22**, 154–169. (doi:10.1016/j.tics.2017.11.006)
  194. Uncapher MR, Lin L, Rosen LD, Kirkorian HL, Baron NS, Bailey K, Cantor J, Strayer DL, Parsons TD, Wagner AD. 2017 Media multitasking and cognitive, psychological, neural, and learning differences. *Pediatrics* **140**(Suppl 2), S62–S66. (doi:10.1542/peds.2016-1758D)
  195. Friedman NP, Miyake A. 2017 Unity and diversity of executive functions: individual differences as a window on cognitive structure. *Cortex* **86**, 186–204. (doi:10.1016/j.cortex.2016.04.023)
  196. Siegelman N, Bogaerts L, Christiansen MH, Frost R. 2017 Towards a theory of individual differences in statistical learning. *Phil. Trans. R Soc. B* **372**, 20160059. (doi:10.1098/rstb.2016.0059)
  197. Saffran JR. 2018 Statistical learning as a window into developmental disabilities. *J. Neurodev. Disord* **10**, 35. (doi:10.1186/s11689-018-9252-y)
  198. Heinz A, Schlagenhaut F, Beck A, Wackerhagen C. 2016 Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. *J. Neural. Transm. (Vienna)* **123**, 809–821. (doi:10.1007/s00702-016-1561-2)
  199. Gottlieb J, Oudeyer PY, Lopes M, Baranes A. 2013 Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* **17**, 585–593. (doi:10.1016/j.tics.2013.09.001)
  200. Engel S 2015 *The hungry mind: the origins of curiosity in childhood*. Cambridge, MA: Harvard University Press.
  201. Freire P 2001 *Pedagogy of freedom: ethics, democracy, and civil courage*. Lanham, MD: Rowman and Littlefield Publishers.
  202. Kidd C, Hayden BY. 2015 The psychology and neuroscience of curiosity. *Neuron* **88**, 449–460. (doi:10.1016/j.neuron.2015.09.010)
  203. Cattaneo L, Rizzolatti G. 2009 The mirror neuron system. *Arch. Neurol.* **66**, 557–560. (doi:10.1001/archneurol.2009.41)
  204. Carcea I, Froemke RC. 2019 Biological mechanisms for observational learning. *Curr. Opin Neurobiol.* **54**, 178–185. (doi:10.1016/j.conb.2018.11.008)
  205. Jespersen S, Blumen A. 2000 Small-world networks: links with long-tailed distributions. *Phys. Rev. E* **62**(Pt A), 6270–6274. (doi:10.1103/PhysRevE.62.6270)
  206. Brockmann D, Geisel T. 2003 Levy flights in inhomogeneous media. *Phys. Rev. Lett.* **90**, 170601. (doi:10.1103/PhysRevLett.90.170601)
  207. Guo Q, Cozzo E, Zheng Z, Moreno Y. 2016 Levy random walks on multiplex networks. *Sci. Rep.* **6**, 37641. (doi:10.1038/srep37641)
  208. Lydon-Staley D, Zurn P, Bassett DS. 2019 Within-person variability in curiosity during daily life and associations for well-being. *J. Personality.* (doi:10.1111/jopy.12515)
  209. Zurn P. 2019 Busybody, hunter, dancer: three historico-philosophical models of curiosity. In *Toward new philosophical explorations of the desire to know: just curious about curiosity* (ed. M Papastephanou), pp. 26–49. Cambridge, UK: Cambridge Scholars Press. Cambridge, UK: Cambridge Scholars Press, forthcoming.
  210. Sizemore AE, Karuza EA, Giusti C, Bassett DS. 2018 Knowledge gaps in the early growth of semantic feature networks. *Nat. Hum. Behav.* **2**, 682–692. (doi:10.1038/s41562-018-0422-4)
  211. Lydon-Staley DM, Zhou D, Zurn P, Bassett DS. 2019 Hunters, busybodies, and the knowledge network building associated with curiosity. *PsyArXiv*. (doi:10.31234/osf.io/undy4)