

CONTENTS

<i>Acknowledgments</i>	ix
<i>Abbreviations</i>	xi
<i>Preface</i>	xiii
Section I. From Brain Flows to Brain Functions	1
1 Seeking a Generative Understanding of the Human Brain and Its Flexible Cognitive Repertoire	3
2 Grounding Cognition in Brain Flows and the Dynamic Connectome	35
3 A Chunk of Matter that Pursues Goals	75
4 Novelty in a Compositional Universe	107
5 Flow-based Generation of Cognitive Levels of Organization	147
Section II. How Brain Flows Generate the Human Mind	179
6 Flow Fields and the Generation of Consciousness	181
7 Effective Goal Pursuit and the Generation of Intelligence	205
8 Causality and the Generation of Free Will	236
9 Dysfunctional Flows and the Generation of Mental Health	259

viii CONTENTS

10	Novel Flow Pattern Search and the Generation of Creativity	287
	Epilogue	319
	<i>Glossary</i>	325
	<i>Notes</i>	331
	<i>Bibliography</i>	345
	<i>Index</i>	363
	<i>Color plates follow page</i>	178

1

Seeking a Generative Understanding of the Human Brain and Its Flexible Cognitive Repertoire

1.1. A Cognitive Challenge and the Origin of Human Cognitive Flexibility

Let's unpack this: That evolutionary pressures drove selection for intelligence and creativity, due to problem solving (goal pursuit) improving survival of self and offspring. This enables immense flexibility in our thoughts and behaviors that is at the core of who we are as humans, and which is also one of the most profound remaining scientific mysteries.

The human brain contains 86 billion neurons linked by over a trillion connections and clustered into hundreds of brain regions, each of which is part of a functional network with semi-specialized functionality (Ji *et al.*, 2019)^E. Why would such an object evolve? A core thesis of this book is that the movement of activity over neural connections (activity flows) are central to brain functionality^S. So let's reframe the question: What led evolution—using neural connections as brush strokes—to produce the human brain? To begin fashioning an answer to this question we will start with a complex enough situation to illustrate the human brain's unique properties and perhaps justify its complexity. Note that even the seemingly simple aspects of the situation I will describe—such as me walking or driving—are actually extremely complex, as illustrated by the many decades of difficulty experienced by engineers trying to replicate these skills in machines. Beyond being a complex real world (i.e., outside the laboratory) experience ideal for illustrating cognitive complexity,

this experience is also notable in being one of the most joyful, whimsical, meaningful, and challenging experiences of my life.

In my daughter's final year of preschool there was an event called Dad Day. This two-hour experience was built up the entire year, leaving my daughter with high expectations for the event. All I knew was that it meant a lot to her, and that (mysteriously) I needed to "bring a hammer." I took the train to work as usual, planning to leave early to make sure any commute hiccups didn't make me late for this important event.

Unfortunately, when I tried to take my train home, all trains were stopped due to a brushfire near the tracks. How could I get home in time for the Dad Day event? With help from my wife and my daughter's teacher, I rapidly replanned for us to go to a later showing of the Dad Day event (meant for another class). I took a rideshare in place of my train, then drove home to pick up my daughter, giving us just enough time for the second showing of Dad Day.

To this point, I had selected a goal (to get to my daughter's event on time), prospectively taken action to help achieve that goal (planning and plan execution), and adjusted my plan in the service of the goal (flexibility). Moreover, I had coordinated a change in scenario (when the event would occur) through social interaction based on a temporal prediction and took action to catch a ride home. While somewhat stressful, solving this problem was also exhilarating.

The everyday nature of this story may make it seem simple, yet even the easy parts were remarkably complex and computationally challenging. This has been a recurring lesson in the field of artificial intelligence (AI) research—the attempt to computationally replicate (or exceed) human intelligence. Over the decades many AI researchers have made the mistake of thinking that tasks that are easy for us humans will be easy to implement in AI algorithms. Instead, something as simple as tying shoelaces requires state of the art AI algorithms and robotics in the first quarter of the 21st century (Luo and Demiris, 2024). We should keep this hidden complexity in mind during our quest to understand the brain processes behind Dad Day.

Once home, I set down my work bag and grabbed the one thing I was told we needed for this event: a hammer. I then drove my daughter to her preschool. After a brief "circle time" in which the teacher read a book to all of us dads and kids sitting on the floor, we moved on to the woodworking portion of the event. The dads were asked to build something with their child, then given random pieces of wood, nails, and glue (Figure 1.1A). I asked my daughter what she wanted to build, thinking she would say a box or letter T (both in my wheelhouse from when I had last done woodworking as a pre-teen).

She wanted a unicorn. I smiled at how fun this idea was, and also to hide my uncertainty about how to go about building a unicorn. Surrounded by ear-splitting hammering from the other father-child dyads, my daughter led us through the box of wood parts we were given, together identifying a horse-like configuration that she was happy with. After some hammering, we had something that (kind of) looked like a horse. We high-fived each other with a strong sense of accomplishment.

But without a horn, it was not a unicorn. We found a pointed piece of wood that could make for a horn, yet with no drill to insert the horn into the horse's head there was no way to attach it. We tried glue, but it would not dry by the time we had to leave, and the horn would not stay attached to the head until the glue dried. The teacher noticed us looking stuck and quickly gave us rubber bands and paperclips, hoping those would help.

My daughter and I stared blankly at the collection of items, unsure how to proceed. After some thinking and trial-and-error, we found a way for a rubber band to exert the correct force on the horn to hold it in place while the glue dried. See Figure 1.1B for that configuration and Figure 1.1C for the finished unicorn. My daughter loved it, and so I was happy with it too. We had accomplished our woodworking goal, and I had accomplished my larger goal of a successful Dad Day for my daughter.

The above story illustrates a sample of the variety of cognitive brain processes all of us humans use on a regular basis, if not typically in such rapid-fire succession. Among the cognitive brain processes I will focus on in this book, I had carried out processes consistent with intelligence and creativity. Importantly, I do not think I was especially intelligent or creative relative to others (anyone could have done the same), but the experience illustrates these abilities nonetheless.

One might ask why we have intelligence and creativity. These abilities were likely directly selected by evolution. While my goal wasn't a matter of survival, it demonstrated a strong drive and creative intelligence toward doing what was best for my offspring. Of course, the highest priority I have with regard to doing what is best for my daughter—even more important than building her a unicorn—is her survival. This suggests that if her survival were at stake, I would have been even more driven, creative, and intelligent in my problem solving. It stands to reason that the genes of individuals in our ancestral past who exhibited these survival-enhancing abilities toward themselves and their family would have tended to be selected for. Over millennia these traits were likely amplified many times over, spreading throughout the population.

But how could there be a biological basis for something as seemingly nebulous as abstract abilities like creativity and intelligence? The human brain has a wide variety of mechanisms, largely continuous in nature with other animals,

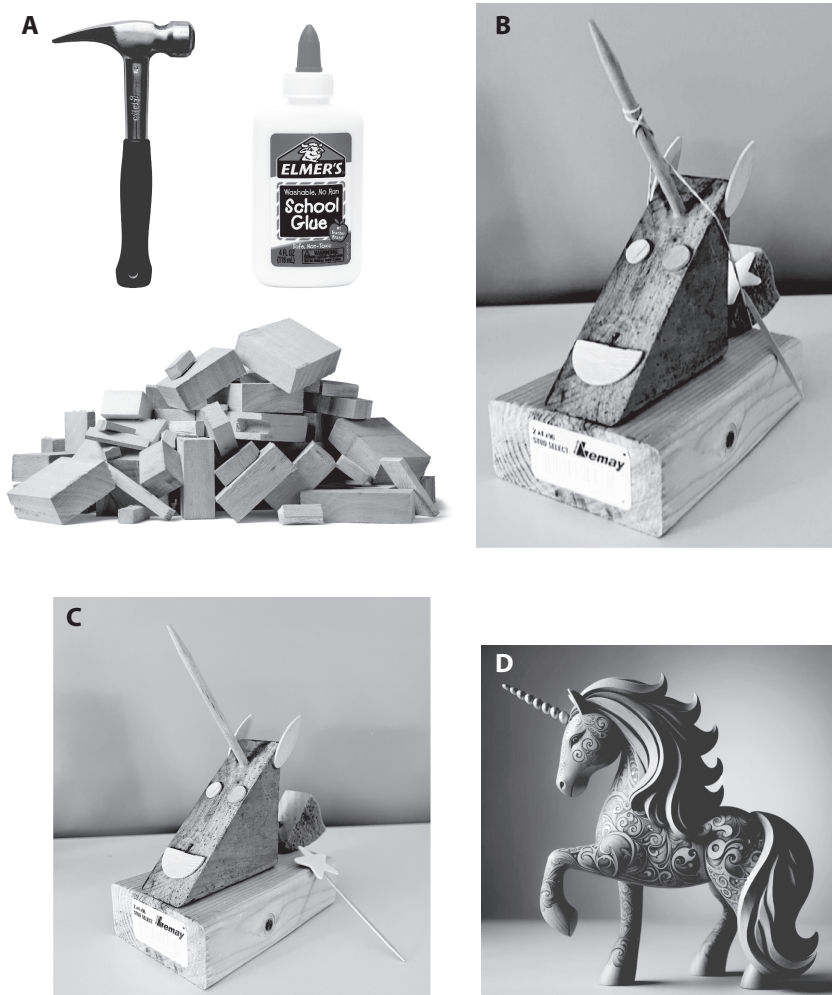


FIGURE 1.1. Components, goals, and unicorns.

A) The components of the unicorn building task. **B)** The solution to holding the unicorn horn while the glue dried (after the addition of paperclips [not used] and rubber bands). **C)** The finished unicorn. **D)** How my daughter said she sees/imagines the finished unicorn. Image generated by ChatGPT in January 2024.

that gift us these abilities. One key feature of both creativity and intelligence is the combinatorial nature of new ideas for solving our problems. As we'll see in chapter 7, defining intelligence is challenging. However, we can use the ability to accomplish a variety of possible novel goals as a working definition. It is no coincidence that the human brain's immensely complex network architecture

is—by its very nature of being a network—well positioned to rapidly explore a wide variety of combinations of its elements^P. This ability to represent ideas in multiple possible combinations depends on compositionality; the ability to represent those ideas in a reusable and recombining fashion. I will return to this notion of compositionality frequently.

My Dad Day adventure was filled with examples of compositional reasoning and creativity. Perhaps the most salient example was the goal of building a unicorn using a limited set of parts and tools. These parts and tools are the components in the compositional reasoning that needed to occur to achieve the unicorn-building goal. Somehow my brain (in coordination with my daughter's brain) needed to represent each component (piece of wood), then work backward from possible unicorn-like forms to identify a sequence of actions (e.g., hammering nails) to build one of those forms. This would not have been possible without my brain treating each piece of wood as a reusable and reconfigurable component—an instance of compositional reasoning.

To go deeper into understanding how my daughter and I were able to build a unicorn we need to go beyond these cognitive-level descriptions, focusing on neural mechanisms supporting the abilities we exhibited that day. A typical neuroscience description would involve a list of brain regions and what they do to support the cognitive functions of interest. While useful, this list of region-function mappings is only part of the story. A richer description would at least include how those regions get involved in a particular order, and even how they interact. Underlying these details—the theoretical target I would argue such efforts are searching for—is a fundamental neuroscientific construct I have termed *activity flow*.

The remainder of this chapter is a series of subsections—starting with an overview of activity flow—that mirror the remainder of the book. Each subsection is a short introduction to the topic of the corresponding book chapter, with the final subsection before the chapter summary introducing the entire second half of the book.

1.2. The Centrality of Activity Flows to Brain Function

Let's unpack this: That brain activity flow patterns are central to the generation of cognitive phenomena and underlie the bulk of useful computation in the brain. Every mental function is generated by a set of brain flows, such that we can understand each mental function by mapping its constituent brain flows and (for deeper understanding) what caused those flows. Following cognitive scientist David Marr (1982), this is the *implementation* level of understanding. As we'll see in later subsections, *algorithmic* (representational) and *computational*

(goal-oriented) levels of understanding build on this initial flow-based implementational level to generate minds from brains.

Activity flows are the movements of activity between neural populations (Cole *et al.*, 2016). These dynamic neural processes are nothing short of the animating force of the brain, the very essence of what makes the brain function^{1,5}. Indeed, activity flows appear to be a primary integrative mechanism in neural computation, constituting the main conduit through which diverse biophysical processes converge to generate mental functions. Mental functions are synonymous with cognition and encompass our experience, our abilities, our memories, and our goals. We will be focusing on how activity/brain flows generate the most fundamental aspects of the mind (such as learned associations and goal pursuit) in the first half of the book, focusing on more complex functions (such as consciousness, intelligence, and creativity) in the second half of the book. Beyond evidencing the simple fact that brain flows compose the human mind, I will share an account of *how* brain flows go about generating functionality despite the computational challenges we face as we pursue goals in a complex world.

The idea of brain flows has a long history. Indeed, philosophers speculated about brain flows at least as far back as René Descartes in the 17th century. However, these initial speculations were based on analogy to hydraulic systems. Much has been learned about brain flows since then, such as their electrical and chemical basis. Over the centuries, substantial evidence has accumulated that brain flows are the primary mechanism by which the brain functions.

Despite the centrality of brain flows to cognitive functionality, someone reading the recent neuroscience literature would be forgiven for thinking brain flows play only a minor role in neural processing. Rather than playing a central role in hypothesis development, data analyses, and data interpretation, activity flows are typically relegated to background knowledge only vaguely related to the main conclusions of most neuroscientific studies. Most contemporary neuroscience studies focus instead on either mapping functions to brain regions or mapping connectivity between brain regions². While helpful, such activity and connectivity estimates are the mere correlates of brain flows, and thus do not address the brain's core functionality. Brain activity flows are fundamental to neural processing and have the potential to unify efforts to map brain functions and connectivity into rich explanatory accounts of brain functions. This journey to the essence of brain function begins with the action potential.

From slugs to mice to monkeys to humans, and from anatomy³ to computational models, the action potential (or spike) has been shown to be central

to neural processing (Humphries, 2021)^E. This pivotal role starts in the periphery, with every sensory modality communicating with the brain via action potentials. These electrical signals flow rapidly along neurons' axons, triggering the release of chemicals (neurotransmitters) that in turn trigger action potentials in the next group of neurons in the chain (Figure 1.2A). This flow of action potential rate increases—activity flow—eventually enters the brain, where much less is known. For instance, not much is known about perception of simple patterns like circles, and even less for perception of complex mythical creatures like unicorns. However, we do know that, if a stimulus causes a motor response, coming out the “other side” are more action potentials from the brain that trigger muscle movements.

This lack of knowledge once action potential flows enter the brain is due in part to technical barriers to recording from multiple neurons at once. After all, there are many billions of cells within mammalian brains and, until the last two decades, neurons were recorded one-at-a-time. Even then, these methods have focused primarily on recording many neurons within single regions, rather than mapping out how activity flows between those neurons or between brain regions. This lack of a method to measure activity flows, and the resulting lack of knowledge, may be a major reason why activity flows have taken a back seat in neuroscience research.

Methodological barriers are even greater for the human brain, where it is normally ethically infeasible to cut through the scalp and skull to record neurons directly. Instead, unless someone is undergoing brain surgery for a medical condition, we must use noninvasive neuroimaging to see the goings on within the human brain. Notably, however, noninvasive neuroimaging methods have utility beyond their noninvasiveness (Tsao *et al.*, 2006; Bandettini, 2009; Van Essen and Glasser, 2018)^E. This is due to their much larger field of view—the entire brain rather than small sections of the brain. While the high-resolution recording from thousands of neurons in animal models is useful for understanding *local* neural populations, it turns out that low-resolution recording of clusters of all ~86 billion neurons can be especially useful for understanding the *distributed* neural codes and connectivity patterns linked together via activity flows. Thus, what started as a technical limitation in human neuroscience—being forced to observe only large neural populations—ended up enabling us to observe global activity flows within the brain for the first time.

Given the functional centrality of the flow of action potential rate increases outside the brain, one would assume it would be important within the brain as well. And indeed this has been a core feature of many neural theories (Figure 1.2B). The relegation of this dynamic flow process to peripheral background theory for most empirical neuroscience studies is like we've been doing

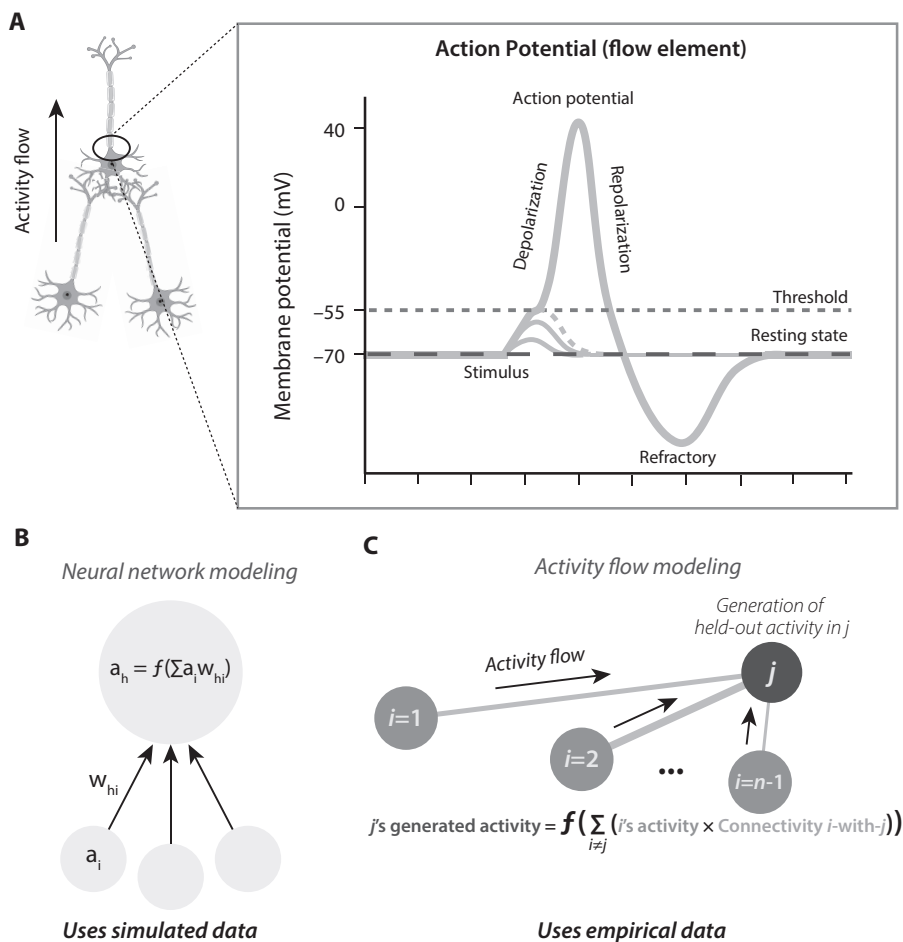


FIGURE 1.2. Activity flows are fundamental to how neural networks function and are modeled.

A) Minimal neural circuit illustrating signal integration, with activity flowing from two neurons to a third. Action potentials flow from the axon hillock (circled on one cell) to the end of the axon (upward in the diagram). Right: Illustration of an action potential, which integrates inputs and is initiated if the inputs reach a threshold. Activity flows—the movement of neural activity (aggregate action potentials) between neural populations—are abstracted from these details, but are ultimately implemented by them. Created with BioRender.com. **B)** Fundamental mathematical framework underlying all neural population simulations, including the artificial neural network methods used in modern artificial intelligence and activity flow modeling. **C)** The activity flow modeling neural data analysis approach. Each sphere represents a neural population, with the to-be-generated activity in neural population j . Adapted from Cole *et al.* (2016).

neuroscience research with our hands tied behind our backs. Yet this presents us with a substantial opportunity for improving our understanding of brain function by shifting to the brain flow perspective, and away from more standard activity-amplitude-centric and connectivity-centric views. Luckily for us, activity flows can be estimated as the combination of connectivity and task-related activity^S (Figure 1.2C), both of which are being mapped via thousands of neuroscience studies every year. We can therefore leverage this vast and ever-expanding knowledge in our quest to understand the essence of brain function.

Consider this salient example of mapping a brain function, which is a starting point for understanding more complex cases (e.g., representations of mythical creatures like unicorns): There appears to be a sort of module in the human visual system for processing faces^E. This brain region is termed the fusiform face area, and it is more active when a participant sees a face relative to seeing other visual categories (such as a house or tool) (O’Craven and Kanwisher, 2000; Tsao *et al.*, 2006; Haushofer and Kanwisher, 2007). While this region does more than just process faces (Haxby *et al.*, 2001), decades of research has established that this chunk of cortex buried within the temporal lobe really is more active in response to faces than any other category, even scrambled faces (i.e., stimuli with the same low-level features). Can we use the brain flow perspective to gain additional insight into this localized function?

Let’s begin by considering what an activity-flow-based account of the fusiform face area would look like, in principle. Importantly, an activity-flow-based account would not replace function maps or connectivity maps. Rather, activity flow modeling builds on this knowledge, mixing maps of task-evoked activity with connectivity maps to estimate activity flows. In the case of the fusiform face area, an activity flow model would utilize this region’s unique connectivity pattern to identify which other regions contribute to its functionality. It would then simulate the flow of activity from those regions to the fusiform face area as it generates the function of interest—here, elevated activity in response to faces (face selectivity). Given that the model generates the function of interest (while being highly empirically constrained) we can consider this to be a *generative* understanding of face selectivity. This generative understanding goes well beyond the function map and the connectivity map that we began with, demonstrating a highly plausible mechanism by which a function of interest emerges in this brain region⁴. Notably, this generative network neuroscience perspective is highly related to neuroscientific *encoding models*—wherein a basis set (e.g., a set of simple visual features) is used to account for a complex neural response (e.g., visual responses to a movie), then used to test for generative sufficiency in independent data (Naselaris *et al.*, 2011; Kay, 2018). Activity flow modeling can be thought of as connectivity-based encoding modeling (Ito, Hearne, Mill, *et al.*, 2020).

Thinking more generally, imagine a neuroscientist's dream scenario: mapping every function of every neuron (and neural population) perfectly. So, every neuron's response could be perfectly described even for highly complex stimuli (e.g., unicorns), cognitive goals/tasks (e.g., building unicorns), and motor responses. We would understand the brain much better than we do now, and we could do all sorts of amazing things, such as predicting neural responses to a variety of stimuli and situations. But would we now understand the brain in any satisfactory sort of way? I would argue no.

The issue is that we would not have gained a *generative* understanding of brain function. For that, one would need to understand what causal mechanisms led to each neural function of interest. As illustrated above, generative understanding of face selectivity in the fusiform face area would require identifying the processes that produced increased activity in that brain region in response to faces relative to other object categories (e.g., places, tools, and non-facial body parts). While brain function mapping is essential for identifying and beginning to characterize the to-be-explained phenomena (thus contributing to understanding), no amount of brain function mapping—identifying response properties of neurons—is going to yield generative understanding.

Instead, we can enhance understanding by considering the activity (action potential) flows that likely do for the central nervous system what they so obviously do for the peripheral nervous system—enable its most essential functionality. In contrast to only mapping brain functions or brain connectivity, mapping activity flows yields generative understanding of neurocognitive functions (Ito, Hearne, *et al.*, 2020). This is due to activity flows linking activity patterns between neural populations (Ito *et al.*, 2017), revealing the generative contributions of upstream regions to downstream regions on the way to generating a given function of interest⁵ (Figure 1.3).

What if, instead of mapping the functionality of every neural population we instead mapped the connectivity between every neural population? Such a comprehensive map of a brain's connectivity is termed a *connectome* (Sporns, Tononi, and Kötter, 2005). While extremely useful (especially for mapping brain flows), having a brain's connectome would still not enable generative understanding of brain functionality. This is because simply knowing how parts connect to each other does not, by itself, reveal what functions their interactions generate.

To drive these points home, consider understanding how a car works. It is useful to know what each part of the car does, but just as useful is to know *how* it does those things. For example, we may know that the accelerator is associated with the car's movement, the engine somehow provides energy for moving the vehicle, the transmission is related to the car switching into different

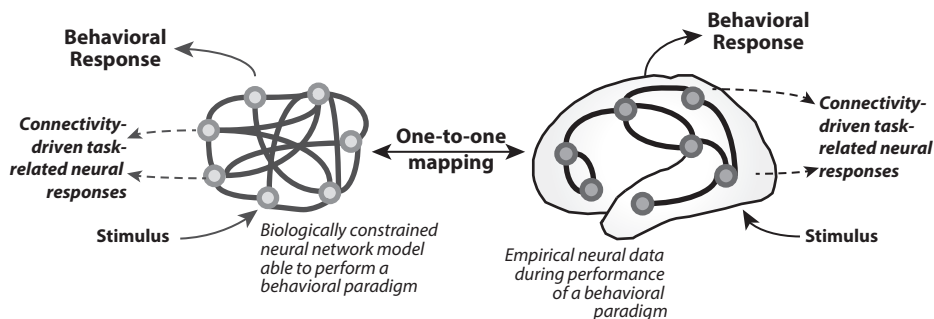


FIGURE 1.3. Activity flow models are generative empirically constrained network models of brain function.

At the aggregate level, activity flow models have a one-to-one mapping between model and neural features, at a given level of model detail/abstraction. This correspondence facilitates making inferences from model data to biological brain functionality. Adapted from Ito *et al.* (2020).

modes (e.g., reverse), and the wheels rotate to make the car move. We also know that these car parts are connected to each other. Stopping here would be equivalent to our current understanding of brain functions based on mapping functions and connections. Instead, we know that the “flow” of causal processes between these parts ultimately makes the car function: accelerator → engine → transmission → wheels. Further, we can ground the (sometimes vague) concept of emergence here—the useful properties of cars (e.g., moving fast and in a controlled manner with minimal human effort) *emerge* from the causal interactions of its parts. It is clear that complex mental abilities such as intelligence and creativity somehow emerge from brain processes. The brain flow framework is an approach to gain such generative causal accounts of brain functions.

In the case of face selectivity in the fusiform face area, we recently used functional magnetic resonance imaging (fMRI) data to model activity flows from their cortical origin (region V1 for visual information) to the fusiform face area. Remarkably, despite using relatively large-scale data (each fMRI data unit contains thousands of neurons), the resulting activity flow model accurately generated face selectivity (Cocuzza *et al.*, 2024). This revealed that flows directly from V1 produced some face selectivity, but that additional face selectivity was generated by subsequent brain-wide activity flows that eventually converged on the fusiform face area after multiple complex flow cycles. Similar insights were gained for other brain regions selective for other visual object categories, including places, tools, and non-facial body parts. Much is left to learn about the generative activity flow processes that produce face selectivity.

But now we may be finally on the path to obtaining the kind of mechanistic understanding of mental functions that neuroscience has been seeking since its inception.

A *raison d'être* for this book is to tell of the scientific journey I have been on since 2014, when I began to see the importance of activity flows for understanding brain function. To that point, like many neuroscientists, I had been mapping brain connections and attempting to link them to cognition using a variety of associative techniques. This involved identifying clusters of connections that matched clusters of task-evoked activations (Smith *et al.*, 2009; Power *et al.*, 2011), correlating individual differences between connectivity and task behavior (Cole *et al.*, 2012), or looking at task-related changes in functional connectivity (Cole *et al.*, 2014). While these are extremely useful, I began to realize that none of them yielded the kinds of deep insights into brain function that I craved. I believed computational models had the potential to provide such insights, but whenever I began using computational models I became overwhelmed with the number of assumptions I had to make to get them to generate interesting functionality. The key insight came when I began using empirical connectivity estimates as the parameters in neural network simulations (Cole *et al.*, 2016), with task-evoked activations as the to-be-generated target of those simulations. These models could then be immediately empirically validated by treating the model-generated task-evoked activations as predictions to be compared with the actual task-evoked activations. Notably, this approach partially mirrors machine learning approaches, which use one set of data to train a computer model to distinguish between data patterns (e.g., face vs. non-face images), then validates the model using independent/held-out data to make sure it generalizes (e.g., to new faces). Certainly there were assumptions with activity flow modeling, but now the computational models' assumptions were strongly empirically constrained and capable of yielding direct insights into those empirical results. Most essentially, I realized, these empirically constrained models could yield rigorous insights into the connectivity-related processes generating neurocognitive functions. Perhaps these models can help us understand complex mental processes like getting to Dad Day on time or building unicorns.

This origin story reveals that activity flow modeling is a distinctive unification of multiple approaches in neuroscience: connectivity (connectome) mapping, task-evoked activity mapping, computational modeling, and machine-learning-based prediction. Accordingly, the brain flow framework has the potential to gain advantages from each of these windows on brain function. Regarding connectivity, brain flows are based on the architectural elements that enable neural communication and are thought to encode learning (and therefore most functions). Considering task-evoked activity, brain flows are

grounded (due to experimenter control) in functions of interest. Focusing on computational modeling, brain flows can be tested for their ability to generate (not merely be statistically linked to) functions of interest. Finally, prediction of empirical brain activity brings the brain flow framework full circle, demonstrating the resulting models' ability to account for real (not just theoretical) brain function. This approach to investigating brain flows will be detailed in chapter 2.

In one sense this shift in focus from what a brain region (and each of its neurons) does to *how* it does it via activity flows is revolutionary for neuroscience, especially cognitive neuroscience. In another sense, the activity flow theoretical perspective may seem obvious—it follows from what most neuroscientists assume to be the central mechanism of the nervous system: the action potential.

To illustrate why it is not universally obvious, however, consider the chapter “Gain and Abstraction” in the 2019 book *The Brain from Inside Out* by the legendary neuroscientist György Buzsáki (Buzsáki, 2019). This chapter focused on coordinate transformations, such as between retinotopic and egocentric coordinate frames. A retinotopic frame is organized in the manner that our eyes take in information, with sections of the visual field arranged in the same manner as they stimulate our retinas. In contrast, an egocentric frame is organized according to where an individual is in space, such as my location in a room, rather than where my eyes are facing. Buzsáki's highly sophisticated account of how coordinate transformations (e.g., between retinotopic and egocentric coordinate frames) occur in the brain emphasizes which brain regions are involved, what each region does, and the role of gain modulation via neurotransmitters and local neuronal normalization mechanisms in implementing those transformations⁵. The details of Buzsáki's explanation don't matter for my point. This is because my point rests on the fact that, despite Buzsáki's book discussing connectivity in other contexts, connectivity is hardly mentioned here, let alone activity flows⁶. In contrast, the present account submits that such information transformations (as described below) are primarily the consequence of activity flow-driven transformations during the communication between brain regions and, to a lesser extent, within-region recurrent activity flows. While gain and normalization are important, I suggest such mechanisms primarily modulate activity to enhance the consequences of activity flows, rather than implementing core neuronal computations. Thus, the brain flow perspective has the potential to build on these more traditional approaches and revolutionize understanding of brain function.

The possibility that information processing in the brain may occur primarily via information transformations *between brain regions* rather than within them is strongly supported by the success of activity flow models (such as the V1-initiated

model of the face area described above). This contrasts with how most cognitive neuroscientists (including myself at one point) interpret task-evoked brain activity: as unknown within-region processing. The activity flow framework suggests that each brain region contains one or more information spaces—such as representations in a retinotopic or egocentric spatial coordinate system—with its activity flows to/from other brain regions actively converting information between these information spaces. Thus, the low-level visual feature information space in V1 containing an entire rich scene is transformed via activity flows into a face representation space in the fusiform face area (Cocuzza *et al.*, 2024). We can imagine taking a V1-eyed view and a parietal cortex-eyed view of a scene. While the V1-eyed view would be in a retinotopic frame and therefore jump around as you move your eyes, the parietal cortex-eyed view would be in an egocentric frame and therefore would not jump around as you move your eyes. A wide variety of parallel processing streams involving other brain regions, each with their own spaces, simultaneously process other aspects of the scene. Regarding our quest to understand how my daughter and I could conceptualize and perceive something like a unicorn (for now letting alone building one), there must be some activity-flow-based processing streams for rare or even novel visual categories like unicorns.

These visual category functions (e.g., selectivity for faces or unicorns) may seem arbitrary, and yet they are an instance of a fundamental form of activity flow transformation underlying flexible cognition and behavior: abstraction. The term *abstraction* is difficult to define in all its uses (psychologists and neuroscientists disagree on its exact definition), so I will operationally define it as the simplification of a set of representations—neural activity patterns consistently linked to a set of events—into a single representation that generalizes across those events. For example, a neural population represents the abstract category *faces* when it is especially active in response to many particular faces. This is a simplification relative to a neural population that is especially active only to a particular face (e.g., Albert Einstein's face), given the complexity of identifying the very specific set of features that make that particular face unique. Adding nuance to this definition of abstraction, it is possible to recombine simplifying abstractions (e.g., horse and horn) to create complex abstractions (e.g., a unicorn).

Abstraction underlies flexible cognition and behavior because it allows for generalization to novel situations. In the case of face perception, it is likely that abstraction in the fusiform face area facilitates perceiving novel faces. Similarly, abstraction of *horse* and *horn* categories likely facilitate perception of unicorns. But what about other kinds of abstraction? Abstract task rule representations, such as “if both items are the same,” have been identified within prefrontal cortex (and elsewhere). These representations can be used to generalize previous learning to novel situations, enabling rapid instructed task learning

(RITL), also known as zero-shot learning (Cole, 2009; Cole, Laurent and Stocco, 2013). Such learning of novel tasks from instructions allows for extremely flexible cognition and behavior. Importantly, my lab's recent work has shown how highly distributed and dynamic activity flows allow these abstract representations to mix with incoming sensory information to generate flexible behavior across a wide variety of tasks (Ito, Yang, *et al.*, 2022). I will be covering this and related studies in chapters 2 through 5. For now we dig deeper into the goal pursuit abilities these abstract representations and activity flow mechanisms enable.

1.3. The Brain as a Goal-pursuing Organ

Let's unpack this: That brains are organized primarily around goal pursuit, with cognitive control networks playing a prominent role in enabling our immense flexibility. These networks compositionally reuse brain flows to pursue goals—predicting action outcomes, planning, and adaptively searching for valuable (e.g., survival-promoting, hedonic) states.

One of the highlights of my career has been taking part in the discovery of dedicated brain systems that make it possible for us to flexibly pursue goals (Cole and Schneider, 2007; Cole, Pathak, and Schneider, 2010). These *cognitive control networks* are distributed sets of highly interconnected brain regions that help us seek valuable outcomes via our thoughts and actions^E. Reflecting this important and computationally challenging functionality, cognitive control networks take up a substantial portion of the brain (Figure 1.4A) (Dosenbach *et al.*, 2007; Duncan, 2010; Power *et al.*, 2011; Ji *et al.*, 2019)^E. A quick calculation from a recent brain atlas (Ji *et al.*, 2019) reveals that cognitive control networks take up 36% of cortex, and (when subcortex is included) 42% of the entire brain. In contrast, the language network takes up only 4% of the human brain. Regardless of why these networks take up so much space⁷, they are worth understanding, as they are important not only for our survival but also our ability to pursue personal goals.

Cognitive control is defined as goal-directed cognition—the pursuit of rewarding states and avoidance of punishing ones. Going beyond simple goals (e.g., bringing food to your mouth), cognitive control is typically associated with second-order goal pursuit processes such as inhibiting habits, solving novel problems, regulating emotions, and switching tasks/goals (Diamond, 2013). These large-scale networks (Figure 1.4) were termed cognitive control networks based on their consistently increased activity as a function of goal pursuit demand (Duncan and Owen, 2000; Wager, Jonides, and Reading, 2004; Chein and Schneider, 2005; Dosenbach *et al.*, 2006). While there is

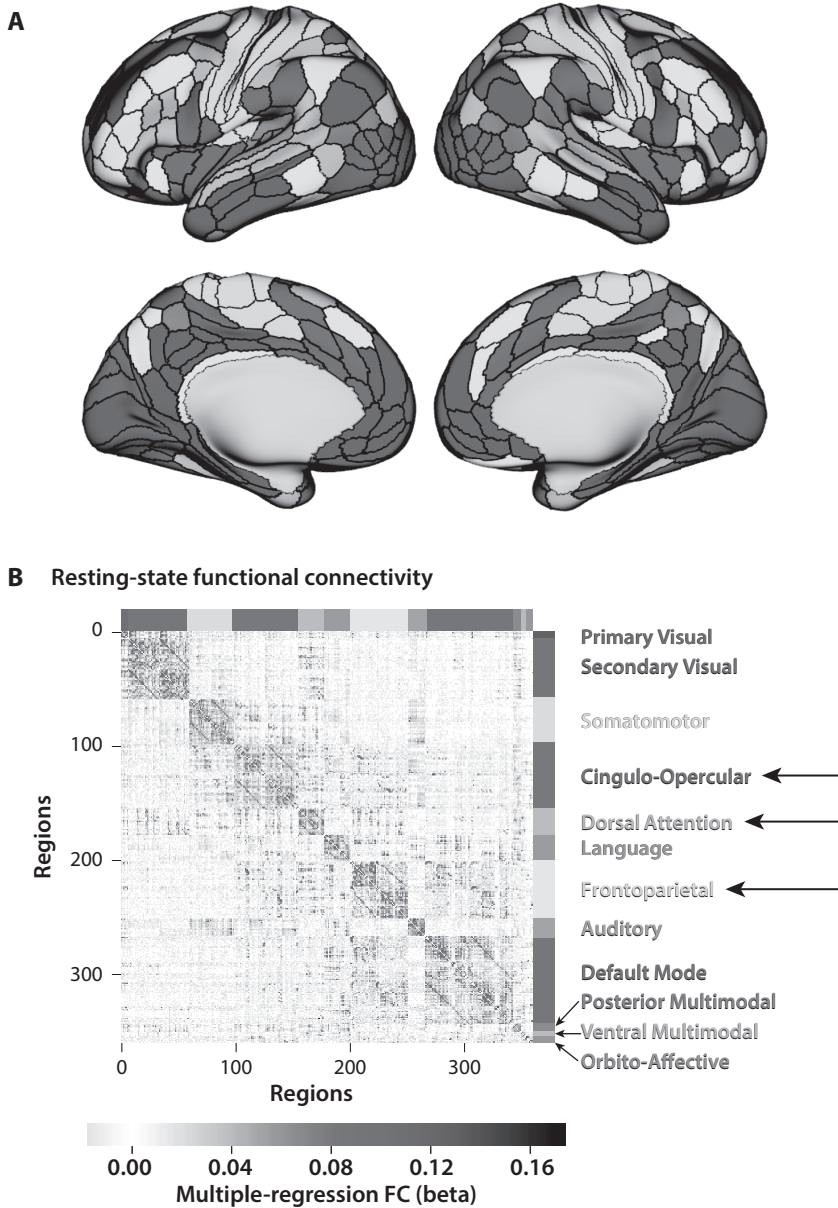


FIGURE 1.4. Large-scale functional networks of the human brain.

A) All functional brain networks in neocortex as identified by Ji *et al.* (2019), plotted as color-coded clusters of brain regions. Each color corresponds to the color of the network label in panel B. Brain regions are plotted on an inflated brain surface for ease of viewing inside cortical folds. These networks were also

evidence for specialization of function within cognitive control networks, they are notable for their *lack* of specialization. Indeed, these networks have been termed the “multiple demand system,” which is a set of regions that is involved in a wide variety of functions (Duncan, 2010). It doesn’t matter if you’re deciding what to have for dinner, taking an exam, or making up a silly song—most of your cognitive control networks’ brain regions are active as you pursue those goals (Wen and Egner, 2023). Thus, these networks really are in some sense large unitary functional units that dominate 42% of the entire human brain.

Shifting our focus back to brain flows (now in the context of goal pursuit), there is strong evidence that cognitive control networks flexibly reroute brain flows in other brain networks⁵. As will be detailed in chapter 3, cognitive control networks use flow rerouting to control other brain networks’ functionality and implement goal pursuit. I postulate that this flow rerouting is so central to cognitive control that we could replace all cognitive control functions in the brain with externally controlled flow rerouting and lose no functionality^p. For example, consider a car driving experiment, in which you (the experimenter) have control of *only activity flow routes* in a driver’s brain. Could you control the driver’s pursuit of a goal destination across myriad challenges?

Consider a simple stop light at an intersection. You could route red representation activity in visual cortex to foot representation activity in motor cortex (red→foot-to-brake), stopping the car at the intersection. You could even flip the rule in a special condition in which green meant stop: suppress any preexisting green-accelerator flows and instead route green representation activity in visual cortex to foot-to-brake representation activity in motor cortex (green→foot-to-brake; green→foot-to-accelerator). Speed and steering control would be a matter of enhancing flows from visual cortex to motor cortex that minimize the difference between the goal states of staying near the speed limit and on the correct side of the road⁸. Thus, while activity flows are essential to all cognitive functions, the (computationally intensive) goal-dependent rerouting of activity flows between neural representations is especially important for cognitive control functions.

Could such flow rerouting be the basis for how cognitive control networks flexibly implement such a wide variety of functions? A key to answering this question is the observation that cognitive control network regions have extensive

mapped to subcortical structures (not shown here). **B**) All pairwise functional connections between the 360 cortical regions estimated using multiple regression with fMRI data, ordered by their network labels. Network names are on the right, with cognitive control networks highlighted with arrows. See plates section for a version of this figure in color.

connectivity throughout the brain (Cole, Pathak, and Schneider, 2010; Power *et al.*, 2011)^S. This hub property may allow control networks to both monitor and control a variety of distal functions throughout the brain (Cole, Repovš, and Anticevic, 2014)^P. For example, during a visual task cognitive control networks must receive information from visual cortex (in the occipital and temporal lobes) to help select the proper motor response (in frontal cortex). When hearing a riddle, cognitive control networks would need to somehow switch to receiving information from auditory cortex to be able to solve the riddle. Consistent with such dynamic switching, we found evidence that the frontoparietal control network—one of the three cognitive control networks—dynamically shifts its brain-wide functional connectivity pattern depending on current task goals (Cole *et al.*, 2013; Cocuzza *et al.*, 2020)^S. We have termed these highly flexible and well-connected regions within cognitive control networks *flexible hubs*. I'll be covering flexible hubs and the possibility that their flexibility enables our brains' immensely flexible goal pursuits in Chapter 3.

Another possible explanation for how cognitive control networks implement such an apparently wide variety of functions may lie in how we define a function. It may be that some functions are more general (i.e., reused across diverse scenarios) than others. Stepping back, let's consider functions applied to individual cognitive control network regions. Cognitive control networks include regions primarily in lateral prefrontal cortex, anterior mid-cingulate cortex, anterior insula cortex, and posterior parietal cortex. Each of these regions has been linked to dozens of specific cognitive functions such as action observation (Molinari *et al.*, 2013) and arithmetic (Andin *et al.*, 2015). However, these regions' consistent involvement in highly general cognitive control functions such as task switching, attention switching, and working memory (Duncan, 2010) suggest many of these specific functions may be instantiations of more general functions. For example, anterior mid-cingulate cortex and anterior insula have been shown to have increased activity during response conflict, conceptual uncertainty, pain, and following errors (Bush, Luu, and Posner, 2000). Several theories appear to be able to account for these diverse responses, such as the expected value of control (Shenhav, Botvinick, and Cohen, 2013) or other predictive accounts (Alexander and Brown, 2019). Put simply, these theories suggest the anterior mid-cingulate cortex (and perhaps anterior insula) predicts the value of exerting effort in the interest of gaining predicted reward (i.e., the value of goal pursuit). This very general function may help explain how these brain regions can be involved in so many diverse tasks and show an apparent diversity of responses.

It is difficult to overestimate just how remarkably general the most fundamental cognitive control functions are. Working memory—the active maintenance and manipulation of information—is perhaps the most fundamental of

all cognitive control functions (Diamond, 2013), and it is so general that it can encode completely novel information. In my own experience, I have learned many new concepts in academic presentations that I had to quickly encode and maintain in working memory so I could follow the remainder of the talk. The brain flow processes underlying these general working memory processes will be discussed in chapters 3 and 5. Even more remarkably, working memory processes can *generate* novel interpretations and ideas. Let's focus on this topic next.

1.4. The Nature of Novelty in the World and the Computational Challenge of Pursuing Goals in the Face of Novelty

Let's unpack this: That the key computational challenge for goal pursuit is dealing with the fact that every moment is novel, with the brain overcoming this challenge via compositional reuse of neural activity flows.

In some sense every moment is novel for each of us. As we'll see in chapter 4, this is true in that we ourselves change in every moment (such as from forming new memories), and so we experience every new moment differently than all those that have come before. It is also true of the world outside ourselves, as it reconfigures in each moment. As the ancient philosopher Heraclitus said, "No man ever steps in the same river twice, for it's not the same river and he's not the same man." How our brains deal with this perpetual novelty has been one of the animating questions of my career.

The immense computational challenge of making sense of a constantly novel universe (and self) is all the more immense for children. With only limited life experience, meeting overwhelming novelty is the norm for them. Since all brains start with this massive computational problem, and given that humans navigate it well enough to survive and often thrive, the fundamental mechanisms of the brain must encompass solving this massive problem. No doubt evolution helps, setting up perceptual, cognitive, and motor systems to aid in our navigation of novelty. However, even in a static environment the combinatorics of possible configurations of predator-prey interactions (for example) is tremendous enough that evolution cannot account for all possibilities. In modern humans the fundamental role of novelty is more apparent, as much of the world society has built is highly distinct from our ancestral environment. All evolution can do is provide us with the computational capacity—such as cognitive control and associated general intelligence—to help us navigate this immense novelty.

From a broad perspective, the ability to understand novel situations is only possible due to a particular property of the universe and the environments we

Earth-dwellers are typically in. This property is the compositional nature of physical matter and energy (Heunen and Sadrzadeh, 2013), in which all objects and even object dynamics are built out of reusable components^P. Anything can be used as an example—for instance, consider that a tree is composed of parts reused in other trees, including leaves, branches, a trunk, and roots. Further, its trunk is composed of wood, which is composed of cells, which are composed of molecules, which are composed of atoms. Consider that the configuration or relationships among these components generate higher-level properties of the tree, and so long as those key configurations are present (e.g., the cells remain attached to one another) we call the object wooden and it has the properties of wood⁹. I can therefore see a tree I’ve never seen before and make inferences based on my previous experience and knowledge about how its trunk and branches would respond to various causal interventions. In the case of the various tree pieces given to my daughter and I on Dad Day, we could leverage the compositional nature of wood and nearby objects (e.g., nails, hammers, glue) to *cognitively decompose* the relevant objects into their properties and infer how we might interact with them to achieve our goal of making a unicorn figurine.

This rapid understanding of a novel situation, along with effective pursuit of a novel goal, depended not just on compositionality in the physical environment but also on a variety of cognitive abilities. It is likely that these cognitive abilities, in turn, depend on compositional neural representations^P. As will be discussed in chapter 4, recent studies have established the brain’s extensive use of highly distributed neural activity patterns that are compositional—systematically reused across contexts (Reverberi, Gorgen, and Haynes, 2012; Cole, Laurent, and Stocco, 2013). In my lab’s research we have found that previously-learned neural task rule representations (brain activity patterns and flows) are reused across a variety of situations, including when combined with other rules (Cole *et al.*, 2013; Ito, Yang, *et al.*, 2022)^S. This demonstrates that neural rule representations can be compositional, supporting the transfer of previously-learned information into novel scenarios (Cole, Laurent, and Stocco, 2013).

Digging deeper into the nature of neural processing, we may wonder how a compositional neural activity pattern—a spatial constellation of neural populations reactivated across similar situations—could help with general goal pursuit. Working backward from a goal-supporting action such as grasping a piece of wood, we can see that reused (i.e., compositional) muscle contraction patterns are helpful for accomplishing a novel goal like unicorn building. After all, a hand grasp reused from other similar objects would require little to no learning, making such novel goal pursuit much more efficient. Going back

further in the causal chain we can see that reused primary motor cortex neural activity patterns—which indirectly innervate muscles and cause muscle contractions—would be important for generating the action-implementing muscle patterns. Critically, this neural activity pattern is only efficacious due to brain flows; the movement of neural activity, in this case from primary motor cortex to muscles (via the spinal cord). The compositional nature of the neural activity (and flow) patterns thus has a direct translation to the compositionality of actions supporting goal pursuit.

How might we arrive at this compositional “hand grasp” activity pattern in primary motor cortex as we pursue our goal? With unlimited time we can imagine randomly trying various activity patterns until we get closer to our goal. However, the combinatorics of possible muscle movements in a single hand are vast. Even considering just five flex strengths for each of the five fingers we have $5^5 = 3,125$ possible finger movement configurations. If I had tried each possible hand grasp on Dad Day, and each hand grasp attempt took 1 second, it would have taken 52 minutes just to pick up the first piece of would-be-unicorn wood. Of course, there are multiple muscles per finger and many flex strength levels per muscle, such that the true computational challenge of finding the correct goal-supporting neural activity pattern would have been even worse.

Instead of a random search, it appears that activity patterns in other brain regions aid the search for goal-supporting actions. One such brain region is the posterior parietal cortex (a cognitive control network region), which supports visual-motor coordination (Freedman and Ibos, 2018) along with general goal pursuit (Alexander *et al.*, 2022)^E. It is therefore likely that goal representations maintained in other cognitive control network regions interact with sensory information via activity flows into posterior parietal cortex to select possible actions^P. Critically, just like more abstract task rules, the possible actions are likely represented compositionally within these brain regions. The selected reused compositional activity pattern (representation) would then flow to primary motor cortex for action execution.

In chapter 4 I will be diving deep into the nature of compositional representations and the role of brain flows in generating and utilizing those representations¹⁰ during goal pursuit. Fundamental to this story is the ability of complex brain activity flows to implement computations via transforming activity patterns between brain regions (Ito *et al.*, 2017). These transformations are not only useful for the central purpose of cognitive processes (goal pursuit) but also constrain and define the functions of each localized brain region (Passingham, Stephan, and Kötter, 2002; Mars, Passingham, and Jbabdi, 2018). We will next focus on this central role of activity flows in shaping neural representations.

1.5. Hierarchies in Map and Territory: A Representational Format for Effective Brain Flow Routing During Goal Pursuit

Let's unpack this: That a primary strategy used by the human brain for overcoming the challenge of navigating the myriad novel configurations generated by physical reality may be to mirror the very organizational structure that enables such organized novelty: composition hierarchies.

The previous subsection suggested that the productivity of both physical reality and brain representations—their capacity to generate new things (events, objects, etc.)—is dependent on their compositionality. Yet we also find that the relationships between elements are often more holistic or conjunctive, reducing reuse of elements and bringing configurational stability to how elements are organized. Clarifying what I mean by *conjunctive*, our wooden unicorn was built from compositional wooden elements, but what made it a unicorn was its particular conjunctive configuration. And so it is with most objects or events, from books to dances to planets. This suggests that much of physical reality (and brain representations) can be described in terms of a particular structure that relates compositional elements and their interactions: *composition hierarchy*^P.

A composition hierarchy is any set of compositional elements that interact to produce a conjunction (Figure 1.5A). For example, the wooden pieces interacted conjunctively to produce a unicorn figurine (Figure 1.5C). The compositional elements can be conceptualized as being at a lower level of a hierarchy, while the conjunction is at a higher level. Notably, the conjunction can be undone if we reconfigure a conjunction's elements, such as removing the wooden unicorn's horn (making it merely a horse). Echoing Plutarch's allegory of the Ship of Theseus—in which all of a ship's parts are replaced and we are asked if it remains the same ship—we can often replace each compositional element with another (functionally equivalent) element, maintaining the conjunctive configuration and therefore higher-level properties (Figure 1.5D and Figure 1.5E). Thus, we can swap out the wooden pieces and nails and nonetheless still have a wooden unicorn¹¹. These properties of (partial) independence across levels help establish that the compositional-conjunctive organization is indeed hierarchical.

Our thoughts often come in composition hierarchies as well^P. This is clear from how we built the wooden unicorn: We considered the conjunctive organization that defines a unicorn (a horse-like shape with a horn) and searched for a combination of compositional elements (wood pieces interacting via

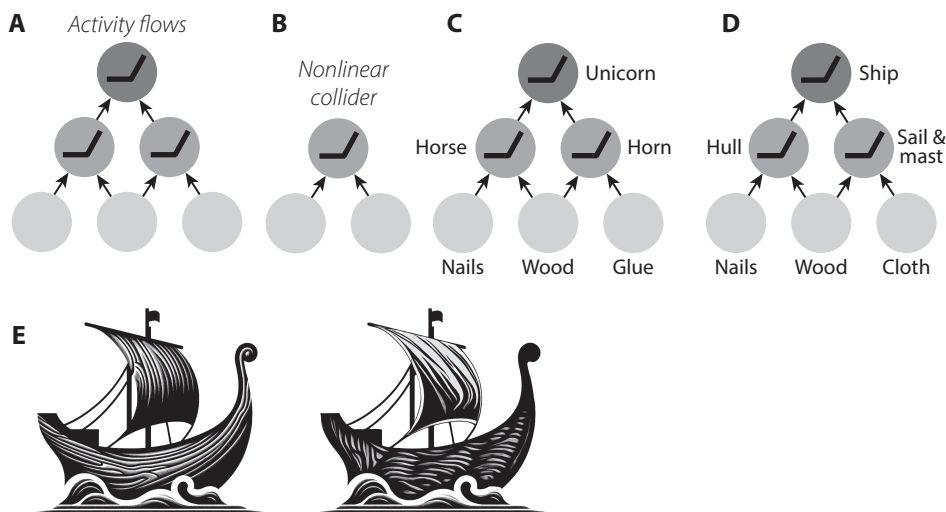


FIGURE 1.5. Composition hierarchies as an account of systematic emergence in brains and their environments.

A) Graph theoretical depiction of a flow composition hierarchy, with each circle as a node in a brain activity flow graph. Each node is a neural population that represents mental content, with lower nodes representing information compositionally and forming conjunctions up the hierarchy. Bent black lines indicate that the lower-level components' flows interact nonlinearly within each higher-level node, such that only certain configurations at the lower level activate that node. Notably, conjunctions at the second level are also compositional representations for the third level. Thus, novel compositional representations can be created and reused across multiple composition hierarchies. **B)** A graph theoretical motif—the nonlinear collider—which composes composition hierarchies. This is the simplest possible composition hierarchy. **C)** A wooden unicorn (see Figure 1.1) represented as a composition hierarchy. The activity flow graph is depicted, with one-to-one relations with physical reality in this case. **D)** A wooden ship represented as a composition hierarchy. Note the compositional reuse of wood and nail representations from panel C. **E)** The Ship of Theseus, rebuilt piece-by-piece from new compositional elements but maintaining its essential functionality via its conjunctive configuration. The ability to replace compositional elements but maintain its essential functionality is evidence for hierarchy. Illustration of ship generated by ChatGPT, then modified.

nails and glue) that fit that configuration. This example illustrates a direct link between composition hierarchies in physical reality and those in our neural “map” of reality. This suggests this particular representational organizational structure (i.e., representational geometry) could be due to evolution, learning, or both. However it arises, this representational geometry is likely particularly useful for both understanding and producing novel configurations of physical reality^P.

The best-established composition hierarchy in the human brain (so far) is the visual system^E. Just as physical reality builds molecules from atoms and organisms from molecules, so the visual system builds line orientations from center-surround fields and object edges from line orientations (Hubel and Wiesel, 1959; Mishkin, Ungerleider, and Macko, 1983; Ungerleider and Haxby, 1994). Extensive evidence suggests this process continues up a hierarchy—the ventral stream—to build integrated (conjunctive) representations of entire objects (Khaligh-Razavi and Kriegeskorte, 2014; Yamins *et al.*, 2014). This process was already covered briefly above (subsection 1.2) with a focus on face object representations, wherein we found evidence that face selectivity in the fusiform face area is a consequence of activity flows along the ventral stream (Cocuzza *et al.*, 2024). Importantly, we showed this is also the case for other visual categories (with their respective specialized regions along the ventral stream): scenes, body parts, and tools.

These findings suggest a metaphor, wherein the center-surround representations activated by photons hitting the retina are the (compositional) atoms of the visual system, with each (conjunctive) representational interaction up the hierarchy forming a unique level of organization, like atoms forming molecules. Belying a simple serial stream of activity flows, however, brains are extremely parallel in their organization. Thus, the visual system consists of many parallel and interacting hierarchies (e.g., the dorsal stream), with compositional representations taking part in forming many conjunctions further up the hierarchy. This organization may explain the extremely flexible nature of human vision, allowing us to perceive and understand novel visual objects and complex visual scenes with ease (Gauthier and Tarr, 2016). This and other features of the visual system will be covered in chapter 5.

Zooming out, we can see that the human brain’s overall network organization is hierarchical, with cognitive control networks sitting “above” sensory and motor networks (Huntenburg, Bazin, and Margulies, 2018)^E. This popular conceptualization of these brain systems is based on defining sensory motor networks as *lower* and these cognitive control networks as *higher*. But does this correspond to *lower* (compositional) and *higher* (conjunctive) in a composition hierarchy? From a pure information representation perspective perhaps not. This is based on there being highly abstract representations in prefrontal cortex

that are even simpler than those found in primary sensory-motor regions (e.g., V1 in the visual system), such as patterns activated in response to the same simple stimulus (e.g., a blue circle) (Wallis and Miller, 2003).

However, if we acknowledge that the processing started (for example) from spatially localized center-surround representations in the retina, then most simple representations in control networks actually result from a complex series of activity flow transformations. Considering a neural population representing a blue circle, we can imagine a series of conjunctions in the ventral visual stream generating a series of representations: (1) location-specific blue, (2) location-invariant blue (i.e., blue regardless of location in the visual field), (3) location-specific circle, and (4) location-invariant circle. These location-invariant representations could then converge via activity flows to sub-populations within control networks to generate location-invariant blue circle representations. This control network representation could then be used as a cue in a context-dependent task, such as indicating that water from a faucet will be cold, impacting your first step when you are trying to make tea (i.e., heating the water). Thus, while the brain's network organization may not be strictly hierarchical, it appears that it forms a composition hierarchy in the sense of building conjunctive representations from sensory-motor input and output representations.

Interactions are inherent to composition hierarchies, as conjunctions within these hierarchies are formed via interactions among compositional elements. How do these interactions occur in the brain? A recent study conducted in my lab found that these interactions are implemented in the brain via activity flows, here in the case of tasks formed out of instructed task rules (Ito, Yang, *et al.*, 2022)^S. Briefly, 12 rules were permuted into groups of 3, forming 64 distinct task sets that each participant performed during fMRI scanning. As will be discussed in chapter 5, we found that compositional activity flows representing rules merged in control networks with compositional activity flows representing sensory stimuli, forming conjunctions that were unique not only for each task but also each stimulus set. Activity flows from these conjunctive activity patterns then selected task-appropriate motor responses in primary motor cortex. Thus, abstract cognitive processes and even complex task behavior are also built out of composition hierarchies. Similar insights regarding the importance of both compositional (low dimensional) and conjunctive (high dimensional) neural representations have been obtained using invasive recordings of neural populations in nonhuman primates (Bernardi *et al.*, 2020), as well as using EEG in humans (Kikumoto and Mayr, 2020; Kikumoto, Mayr, and Badre, 2022).

Above, I suggested the possibility that the brain evolved to use composition hierarchies due to the demands of goal pursuit within a physical reality that also uses composition hierarchies. There are other compelling possibilities as

well. As discussed in chapter 4, it is also possible that the same physical constraints that led to composition hierarchies throughout physical reality led to the emergence of composition hierarchies in the brain. Though this seems plausible, it is important to note just how distinct these hierarchies are: Rather than local interactions binding atoms into objects, information-representing neural activity flows over long distances to bind representations in distal neurons' activity patterns. This seems to suggest that even if the same physical constraints led to hierarchies in brains and other objects, the forms of those hierarchies should nonetheless be highly distinct. This and other possibilities will be unpacked in chapter 4 and the second half of the book, revealing that the human brain is unique in its ability to build complex composition hierarchies that can, through sustained goal pursuit, generate elaborate novel configurations in the physical world (e.g., buildings, inventions, art).

Perhaps the most remarkable property of composition hierarchies is their extensibility to higher and higher levels of organization^P. In the case of physical reality, this is quite apparent in the case of evolution—a distributed algorithm that has resulted in multiple levels of organization, from organelles to cells to organs to organisms to families/groups to societies. As covered in chapters 4 and 5, compositional reuse at each level (e.g., cells) generates the next level (e.g., organs). Remarkably, these conjunctions (e.g., stomachs and kidneys) themselves then become compositionally reused and combined to generate the next level (e.g., individual humans). A very similar process occurs in brains, but now in a more abstract information domain carried by activity flows^P. Take, for instance, the invention of writing and reading. Having been invented only ~5000 years ago, there has been insufficient time for humans to have evolved dedicated neural circuits for reading, and yet with practice our brains are rewired to generate the activity flows needed to rapidly understand even novel text. In chapter 5 I cover how the generation of new conjunctive representations, along with neural pathways to rapidly activate those conjunctions, is likely the basis of automatic behavior such as reading familiar words, with flexible cognitive control facilitating conceptualization of complex novel concepts while reading (as you just did for this sentence).

1.6. How Brain Flows Generate the Hallmarks of the Human Mind

Let's unpack this: That the human mind is generated by millions of competing composition hierarchies, each produced and selected via distributed brain flows. The resulting global process is consciousness. These complex competitions are optimized via feedback during goal pursuit, leading to intelligence,

(partial) free will, mental health, and creativity. This grand process is woven from three key constraints: brain flows, representational transformations via composition hierarchies, and goal pursuits.

Unlike the earlier subsections of this chapter, this subsection introduces the topics for the entire second half of the book. I have taken inspiration from the book's central thesis in deciding how to organize it: The book's second half compositionally reuses content from the first half to generate novel explanations for how our brains' network dynamics generate key features of our minds.

The key ingredients developed in the first half of the book will be at three levels of organization: (1) brain flows, (2) representational transformations, and (3) goal pursuits. Each lower level generates the next-higher level, with the organization specified at the higher level also influencing the lower levels. This is like how your hand (higher level) is composed of cells (lower level), but the cells move through space as you wave your hand by virtue of being part of a hand. Brain flows are organized into specific representational transformations, and those transformations are organized into goal pursuits that influence future brain flows and representational transformations. Notably, these levels of organization roughly correspond to David Marr's famous levels of analysis for information-processing systems (Marr, 1982): (1) implementational, (2) algorithmic, and (3) computational. The implementational level composes the algorithmic processes that realize the system's computational goals. I will be making the case that brain flows implement (implementational level) representational transformations via flow composition hierarchies (algorithmic level), which are organized according to the computational goal of fulfilling rewards and avoiding punishments (computational level).

Importantly, while some cognitive scientists have used Marr's levels of analyses to claim that the implementational level (the brain) can be ignored when investigating the algorithmic and computational levels (the mind), the flow mechanisms focused on here reveal the limitations of this assumption. Brains/minds are composed of many levels of organization, with the lowest levels often providing constraints on the higher levels. This is good news for the science of the mind, as neuroscientific observations help reduce the combinatoric explosion of theoretical possibilities at the algorithmic and computational levels. Moreover, the activity flow modeling approach holds the promise of reading out details of the algorithmic level—the mechanisms of the mind—by using neuroimaging data to produce models at the implementation level. This data-driven bottom-up approach is becoming an important complement to the theory-driven top-down approach used in most of cognitive science.

A broad overview of the book's brain-flow-based generative network neuroscience framework is illustrated in Figure 1.6. I do not expect you to be able

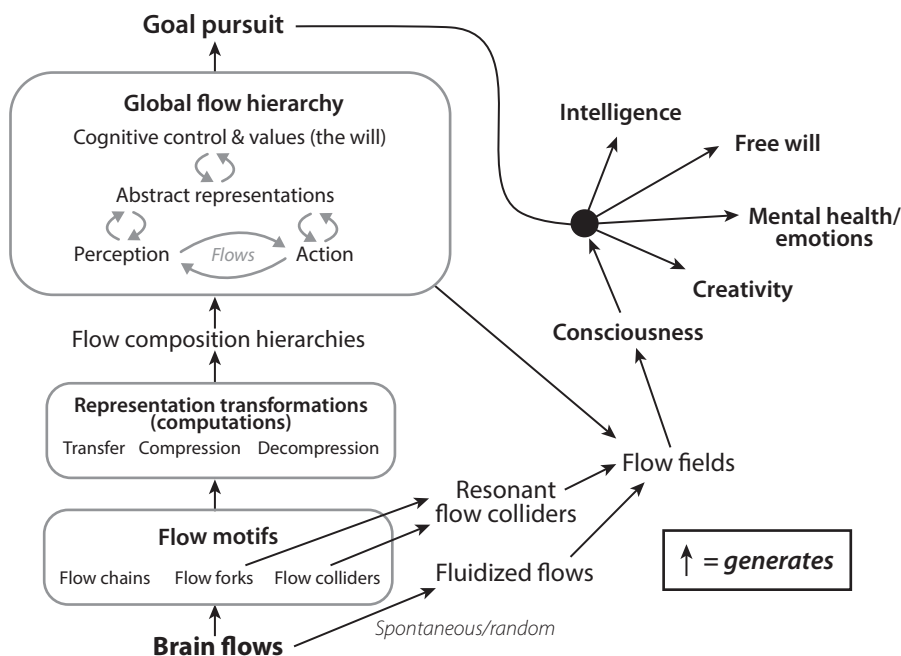


FIGURE 1.6. A map of the brain flow framework.

A guide to how the many levels of organization covered in this book are related to one another and to brain flows in particular. Brain flows are hypothesized to underlie every neural and psychological phenomenon covered here. This figure is not meant to be fully understood at this stage. Rather, it is meant as a high-level guide to how the many topics to come are related, with full comprehension of the figure becoming possible toward the end of the book. Note that *resonant flow colliders* are hypothesized to be embedded within the *global flow hierarchy*, but they are illustrated separately to emphasize their proposed role in compositionally generating *flow fields*.

to understand it at this point, but I am sharing it here as a map of what is to come. I expect that referring back to this figure will help you understand how the ideas throughout this book relate to one another. The figure also illustrates an important point: Despite some stretches of this book not overtly referring to brain flows, other than the processes underlying brain flows (e.g., action potentials) *every element of every brain process described in this book is conceptualized as being compositionally generated from brain flows*. Indeed, I considered rewriting sections describing psychological-level phenomena such as goals, values, and decisions in terms of equivalent brain flow processes.

I did not rewrite these sections in terms of flows for three reasons. First, such flow-based descriptions would be more difficult to read than their psychological-level equivalent, since these higher-level phenomena can be summarized with fewer words than their equivalent multi-level description. The reader is invited to apply mappings from psychological-level phenomena to underlying flows described in the first half of the book to any references to those psychological phenomena throughout the rest of the book. Second, just as I consider brain flows to be real phenomena despite emerging from lower-level molecular interactions, so I consider the psychological phenomena that emerge from lower-level brain flow processes to be real phenomena worthy of descriptions at their level of organization. Finally, I often found that I could not redescribe all psychological phenomena in terms of flows because we often do not know *exactly how* flows generate psychological phenomena. My goal here is to highlight the power of brain flows to fill knowledge gaps between the neural and the mental, but also to highlight the exciting research to be done to fill remaining gaps.

My Dad Day adventure was a massive computational challenge, and at the start of this chapter we were only able to partially account for my ability to overcome the challenges I faced. With each subsection we have gained new tools for understanding brain function. Perhaps we can begin fashioning an answer now, though I expect we won't get close to a satisfactory answer until the final chapter.

Dad Day required me to use thousands of visual, auditory, and motor representations. This is clear from all the stimuli I had to process and actions I had to take to accomplish my goals that day. Further, the sensory inputs had to somehow interact to allow me to infer features of the physical world (e.g., the appropriateness of each wooden piece for building a unicorn), assess progress toward my goals, and generate motor actions to improve that progress¹². As already discussed, these many processes were likely mediated by highly distributed brain flows. Strong evidence suggests that, despite their diverse functionality, these flows were highly organized across all the content domains and goals I processed that day⁵. For example, the first study quantifying brain-wide activity flows (Cole *et al.*, 2016) found that estimated activity flows could account for most of the variability in brain activity patterns across seven highly distinct tasks. These were tasks involving processes as varied as visual categorization, auditory story comprehension, abstract reasoning, emotion processing, and working memory manipulation. Further, we found evidence that all these distinct processes flowed via the same (highly complex) brain network architecture. As we'll see throughout the rest of the book, the precise organization of this brain network architecture is

central to generating the varied computations needed to accomplish goals such as those on Dad Day.

Ultimately, the flow-based interactions between goal representations, the incoming sensory information, and feedback from my actions drove me to successful goal pursuit. Each important piece of information did not come ready made for supporting my goals, however. Rather, my brain had to use its many flow composition hierarchies to process incoming information in parallel, allowing top-down goal priorities to select the highly processed information most relevant to my goals. As covered in chapter 6, we experience the brain-flow-mediated dynamic competition between these many hierarchies as consciousness. This information-rich cacophony of signals aided my goals of making it to my daughter's preschool event on time, giving her the support she deserved, and helping her create a wooden replica of a mythical creature. Knowing this is helpful but not sufficient for understanding my goal pursuit on Dad Day.

Perhaps digging into the fundamental nature of goal pursuit will help. Stepping back, as mentioned in subsection 1.3, it's clear that goal pursuit is an optimization process—the act of minimizing or maximizing one or more quantity, such as minimizing errors on a test or maximizing one's reward. Goal pursuit, the maximizing of reward (including minimization of punishment), is perhaps the ultimate optimization process, as it can take on any possible state as the reward-maximizing goal state. Consider, for instance, the arbitrariness of the most rewarding (i.e., winning) states at the end of a board game: the largest pile of fake cash in Monopoly; the correct (randomly selected) set of murderer, weapon, and location in Clue; or the capture of a chunk of wood or plastic (the opponent's king) in chess. Perhaps more properly adding optimization to our conceptual toolkit will help us make sense of how human brains can successfully pursue complex goals like I did on Dad Day.

Optimization as an algorithmic process was first directly characterized starting in the 19th century, though its roots go back at least to Newton's work on the calculus of variations in the 17th century (Newton, 1687). The calculus of variations is about finding extrema—maxima or minima—such that it fits nicely within our definition of optimization. More broadly, we can consider anything that maximizes or minimizes some quantity to be an optimization process. Centuries of work on optimization methods have resulted in advances in the calculus of variations in mathematics, control theory in engineering (Wiener, 1961), regression in statistics, reinforcement learning in psychology (Sutton and Barto, 1998), and utility maximization in economics, with recent advances in artificial neural network optimization using gradient descent and backpropagation of error (to minimize error during training) (Rumelhart, Hinton, and Williams, 1986). All of this sophistication pales in comparison to

the optimization potential of the human brain, which, of course, created these optimization tools as part of its myriad goal pursuit optimizations.

Let's see if we can use some concepts from these optimization tools to better understand our own goal pursuit abilities and their impact on our more general neural processes. Consider that many optimization approaches involve maximizing success by minimizing error. For example, the standard approach to regression (i.e., fitting a line to a cloud of data) in statistics involves reducing the summed error (the difference between the line and every data point) iteratively until it stops reducing. Similarly, goal pursuit minimizes errors over multiple attempts. Thus, every time a child attempts to create a dog figure using LEGO bricks he or she gets closer to it looking more like a dog. Importantly, if the errors being reduced with each attempt are along many parallel composition hierarchies, then it is likely there will be improvements to representations that generalize to a wide variety of similar situations. This could result in parallel improvements to: (1) the specific ability to create dogs with LEGO bricks, (2) the general perceptual-motor skill of working with LEGO bricks, (3) the general ability to create dog figures, and (4) the highly general ability to create any figure with any material. Thus, the simultaneous maintenance and goal-pursuit-based updating to many parallel flow composition hierarchies likely results in extensible abilities through learning that is both specific and generalizable.

Going back to the Dad Day case of our ability to build a unicorn figurine, we can now imagine my daughter's (and my own) experience playing with LEGO bricks, Play-Doh clay, and other media (e.g., sand) having a significant role in making that possible. I also had some limited experience with wood-working decades earlier I could leverage. Specifically, based on the properties of optimization processes, we can say that our earlier experience with related goals had sculpted a variety of composition hierarchies, some of which transferred that learning to unicorn building on Dad Day. Thus, while building with LEGO bricks mostly increased our ability to work with LEGO bricks¹³, there was some learning high in representational hierarchies that could transfer to any kind of building activity. Importantly, given that many representations in the hierarchies are compositional and therefore reused across diverse settings, many representations would have been meaningfully updated with even highly diverse learning episodes (e.g., playing with clay, sand, etc.). Our ability to build something novel from wood thus likely resulted from years of sculpting our many composition hierarchies through goal-driven feedback. There had to be some way to take advantage of this prior learning during the act of building a unicorn, however. This likely took the form of coordinating brain flows among these previously learned goal-relevant neural representations. As we'll see in later chapters, this coordination likely benefitted from the complex brain

(continued...)

INDEX

- abstraction, 16, 57, 60, 67, 124, 149, 158–167, 213; emotions and, 265; generative, 300–304; science and, 306–307
- activity flow modeling principles, 57–59
- activity flow network motifs, 62–64
- artificial intelligence, 4, 62, 110, 156–158, 170, 223–235, 228–235; artificial general intelligence and, 224; compositionality and, 126, 156, 170; future of, 234–235; large language models and, 225–228, 232–233
- attention, 20, 77, 81, 105, 123, 128, 193; conjunctions (binding) and, 133–146
- automaticity, 73, 78–85; biases/irrationality and, 231; conjunctions and, 173; development/learning and, 295; emotions and, 267; habits and, 78–79; intelligence and, 208, 218; intrinsic functional connectivity underlying, 85, 194
- brain flows, basis of, 8–10, 40, 138
- causality, 12–13, 22, 38; activity flow and, 58, 60, 62, 69; brain connectivity (effective/functional connectivity) and, 45–49, 52, 73 148–149; compositional determinism and, 217–220, 236–253; feedback loops and, 94–95; mental health treatments and, 272, 280–283; nested network framework consisting of, 109, 112–117; universal connectivity costs and, 132–133
- cerebellum, 85, 171–173, 189, 331n7, 339n1 and 8
- cognitive control (executive functions), 17–21, 23, 26, 28, 67, 78–85, 91–106, 119, 125, 158, 160–162; AI and, 227; consciousness and, 186, 190–193, 197; creativity and, 302; flexible hubs underlying, 183; free will and, 236, 250–253, 256; intelligence and, 206–220, 229; mental health and, 266, 274, 277–281, 288–291
- cognitive control networks. *See* cognitive control; prefrontal cortex
- composition-conjunction learning loop, 168–177; creativity and, 314; intelligence and, 206, 212, 220–223
- composition hierarchies, 24–34, 118, 140, 151–155, 160–161, 165–167; composition-conjunction learning loop and, 168–177; emotion and, 270; intelligence and, 220–223
- compositionality, 7, 22, 52, 88, 91–92, 107–116, 124–125, 130–133, 156–157, 162–164; free will and, 246–253
- conjunctions (conjunctivity balancing compositionality), 24–28, 54–55, 65–68, 115–116, 126; the binding problem and, 133–146; long-term memory and, 145
- connectivity (brain connectivity), 8, 11–12, 14–15, 44, 175; computation and, 62–69, 153–154; costs of, 124–133; excitatory-inhibitory balance of, 189; functional connectivity and, 44–45, 48–56, 83–84; graph theory and, 186; hierarchy

- connectivity (*continued*)
 - (gradient) and, 117–123; intrinsic
 - (resting-state) functional connectivity, 49–52, 83–85, 167, 208; task-state
 - functional connectivity, 85–93, 138, 187
- consciousness, 181–204; flow fields
 - underlying, 183–184, 194–200
- core values, 123, 166–167, 203, 209, 226, 230–235; free will and, 236–258
- creativity, 5–8, 204, 252, 287–318
- Dad Day story, 4–7, 14, 22–23, 31–34, 312, 315, 323; creativity during, 287–292; emotions during, 260–261; free will during, 237–240, 248, 251; goal pursuit during, 75, 125; representations during, 148–150; values and, 71; zone of proximal composition and, 177
- default-mode network, 18, 105, 119, 123–124, 193, 196, 209, 226, 237–238, 257, 273–276, 291
- dynamical systems theory (complexity theory, network science), 41–43, 94–101, 321; attractors and, 43, 98–100, 128, 193, 262, 266–267; basin boundaries and, 251–252; mental health and, 282; sand models and, 41–42, 188–189
- effective intelligence, 209–211, 229–230
- emergence, 13, 31, 41–42, 112, 115–116, 128–129, 131; consciousness and, 189, 204; creativity and, 290, 297; free will and, 245–246, 251
- emotions, 232, 234, 259–272, 322; brain dysfunctions and, 272–277
- feedback loops, 42, 94–101, 114, 119, 128, 142–143, 206, 212, 252; creativity and, 310; mental health and, 277, 279–280, 293
- flexible hub theory, 88
- flow adaptation, 294–300, 310
- flow colliders, 14, 60–66, 68, 74, 115, 134, 140–142, 174, 221
- flow field, 30, 142–146, 181–185, 194–204, 313–314; emotions and, 293–294
- flow motifs, 62–64, 134, 319, 323
- free will, 230, 236–258, 262, 272, 315–317
- functional networks, 17–18, 26, 49–50, 86, 92, 113, 119, 121, 130, 157, 186–187, 241–242
- generative abstraction, 299–304
- generative mental model, 219–223, 247, 270; creativity and, 289–290
- generative understanding, 11–12, 43, 59, 319; brain dysfunction treatments and, 261–262, 272; intelligence and, 220; research and, 71–74
- global flow hierarchy, 14, 84, 98, 119–124, 137, 319–322; creativity and, 288–299, 307, 311–317; consciousness and, 185, 196, 198, 200–204; effective intelligence and, 229–230; free will and, 236–240, 245–246, 273; intelligence and, 208–211
- goal-directed search, 85, 96
- goal pursuit, 17–21, 24–28, 32–33, 75–106; consciousness supporting, 200–204; intelligence and, 205–235; novelty and, 110–116; optimality and, 32–33, 255–257; reinforcement learning and, 254
- grounding, 35–39, 68–71
- hierarchy. *See* composition hierarchies; global flow hierarchy; nested hierarchy
- hippocampus, 67, 126–128, 141, 169, 171–175, 264, 339n9
- Holyoak, Keith, 158–159
- Hume, David, 70
- intelligence, 5, 205–235; cognitive control underlying, 211–217; compositionality enabling, 217–220; composition-conjunction learning loop generating, 220–228; consciousness and, 202; rationality and, 228–235

- James, William, 16; automaticity and, 78–80, 84; free will and, 250–251, 317
- Kahneman, Daniel, 104
- Lakoff, George, 36, 151–152
- mental health, 259–286; emotions and, 259–272; harmful dysfunction definition of, 260
- Mill, John Stuart, 115, 293–299
- modularity, 56, 113, 130–133, 185–187; free will and, 238, 241
- nested hierarchy, 110–116, 118, 122, 132–133, 241
- novelty, 16–17, 21–24, 86, 88–93, 107–146, 155, 158, 168–177, 181, 204, 220; cognitive control and, 82–83; creativity generated by, 287–318; free will and, 246, 253, 257; intelligence and, 205–211
- physics, 110–116, 313; free will and, 236, 239, 247–248
- prefrontal cortex (cognitive control networks), 16, 20, 27, 67, 86–88, 91–92, 100, 128; brain dysfunctions and, 273, 279; consciousness and, 191–192; free will and, 256–257
- rapid instructed task learning (RITL, zero-shot learning), 62, 81, 90–93, 171–177, 217, 223, 284, 303
- rationality, 101–105, 198, 228–235
- reinforcement learning, 32, 36, 37, 97, 219, 224–225, 227, 241, 252–254
- representation, 14, 16, 24, 30, 55, 66, 135, 140, 148, 161; geometry of, 65–69; goals and, 71, 86, 125; interactions of, 66, 158–167, 169; relational, 143–144, 147–150, 159–163, 166, 194, 199, 217, 221, 244
- resonant flow collider, 30, 134, 141–144, 192, 202, 248, 313–314, 323
- schema, 125, 143–144, 149, 155, 166, 265, 270, 301
- sensory memory (iconic memory and echoic memory), 194–195
- spontaneous neural activity, 41–42, 50, 78; consciousness generated by, 185–194, 197; creativity generated by, 292–299; free will generated by, 250–251
- Treisman, Anne, 133–136
- universal connectivity costs, 132–133
- values. *See* core values
- von Neumann, John, 65, 159, 223–224
- working memory, 20–21, 31, 80, 96, 100, 195; artificial intelligence and, 227; conjunctions within, 174–176; creativity generated by, 304, 314; free will and, 251–252; intelligence and, 214–216, 220–223; second-order relational reasoning and, 160–165, 167
- zone of proximal composition, 167–177, 252, 295–296, 323