4 The Metaphysics of Developing Cognitive Systems

Why the Brain Cannot Replace the Mind

Mark Fedyk and Fei Xu

1. Introduction

Where is the mind? One of the aims of cognitive science is to answer this question and in so doing provide an answer more substantive than the assertion that parts of it are somewhere behind the eyes and between the ears and that it involves, somehow, the brain. The chapter offers an answer to this question: We shall argue that the mind is a computational network that occupies a functional location in a more complex causal system formed out of various distinct but interacting neurological, physiological, and physical networks. The mind therefore has a functional location.

The idea of a functional location is important because only static entities have relatively fixed spatio-temporal properties. By their very nature, dynamic networks have constantly changing locations, even though states of these networks can usually be spatially located. Nevertheless, a network can be located by specifying how its endogenous processes and states interact with exogenous processes and states—as well as exogenous networks, systems, or, indeed, any other kind of causal process. The network itself can be located by describing, at least in part, its *functional role* within a larger web of cause and effect; again, this is to give the functional location of the network.

We begin with the distinction between functional and spatial location because Smith, Byrge, and Sporns (this volume) offer a different answer to the question of where the mind is located. According to them, the mind—in the non-reductive, Fodorean sense—is not real because it cannot be functionally located in relation to the brain. Their argument for this is novel and ingenious; they reason as follows: If cognition is real and cognition is computation, then the cognitive system must be composed out of a set of static states. But all of the parts of the brain are dynamic processes and entities. So, for any states of the cognitive system to also be components of the brain, they too would have to be dynamic; yet, since cognitive states are computational states, they must be static. It follows that there are no cognitive states. The brain is constituted in a way that prevents the mind from occupying a functional location amongst its many different neural processes and connective networks.

We disagree with Smith and her co-authors. There is a coherent sense in which systems composed of dynamic processes and computational systems can be *co-instantiated* (Siegelmann & Sontag, 1995; Whittle 2007). We therefore reject the major premise of Smith et al.'s argument. Indeed, there are many examples of this kind of co-instantiation. For instance, every electronic device with a computer chip in it is an example of a computational system (the electrical network running on the chip) that is co-instantiated with a physical system (the circuits etched onto different physical components linked together by a complex of wires and channels). More esoteric examples are provided by phenomena like membrane computation (roughly, those things which implement P-systems; cf. Păun & Rozenberg, 2002), biological processes that implement cellular automata, for example colour patterns of certain cephalopod molluscs (Packard, 2001; but see also Koutroufinis, 2017), and models of gene expression and regulation that conceptualize both as a form of "natural" computation (Istrail et al., 2007). These examples make it clear that it is neither logically impossible nor scientifically improbable that the mind is a computational network co-instantiated with a number of dynamic neural networks.

And yet, because almost all of the metaphysical theory of the development of dynamic systems that Smith and her co-authors offer is one that friends of cognition can, and should, accept, we shall use this chapter to pursue a deeper response to Smith et al. Our primary aim will be to extend the theory of the metaphysics of cognitive development that Smith et al. provide so that the extended theory can be used to give a (partial) account of the mind's functional location. Our extension of Smith et al.'s dynamic systems theory will also allow us to clarify exactly why there is no inference from the dynamism of the brain's networks to the non-existence of a cognitive system. But, in establishing this clarification, we will also prepare the ground for our secondary aim, which is to establish support for an argument that shows why, given a choice between our extended view of cognitive development and Smith et al.'s comparatively restricted view, the extended theory should be preferred. Here, the argument is simple: Our extended view can explain more scientific data than Smith et al.'s restricted theory—for that reason, the extended view should be preferred.

In more detail, here is how we shall proceed. We will start from an abstract examination of the metaphysics of dynamic systems (section 2.0). We begin here because many of the most scientifically interesting dynamic systems have two salient ontological features. First, they are made up of component systems and mechanisms that are organized at different levels of energy and constructed out of different forms of energy. Second, many of these systems can realize (sometimes extraordinarily complex) functions. Thus, our initial goal is to provide an explanation of these two observations. To do this, we introduce the concepts *causal buffering* and *metaphysical transduction* to explain how dynamic systems can be made up of systems that are able to pass information between themselves, thereby either establishing or maintaining the function of the system, without also transmitting so much energy as to cause the overall system to break apart. Then, we turn to innateness. Smith et al. are skeptical that the concept of innateness is compatible with a dynamic systems worldview, but we show (section 3.0) that there is a way of defining innateness in terms of developmental essentiality that is not only compatible with this worldview but also helpful for explaining how new systems can emerge as components, or byproducts, of existing dynamic systems. To wit: The existence of certain innate causal buffers (section 4.0) and certain innate metaphysical transducers (section 5.0) provides an elegant explanation of how mind can be a computational system that is co-instantiated with the brain's dynamic neurological networks. This conclusion provides us with the premise we need to argue that consilience considerations favor our extended view as compared to Smith et al.'s more restricted view (section 6.0). Lastly, we return to the question of the functional location of the mind in this chapter's final brief conclusion (section 7.0).

2. The Metaphysics of Systems of Systems

Smith et al. review in impressive detail the many ways that interactions with the proximate environment can both dynamically alter, and drive increases in the structural complexity of, various neural and physiological networks. Brain, body, and environment (BBE, hereafter) are constantly causing changes to one another. We think that cognitive mind needs to be added to this list of interacting networks. But in this section we will focus only on the metaphysical architecture of the complex, dynamic network formed out of brain, body, and environment—since it is not possible to disagree with Smith et al.'s contention that the BBE network plays a central role in structuring virtually all levels of development, including cognitive development.

Our starting point is an observation about the physical integrity of BBE networks—namely, that the integrity or stability of a BBE network cannot be taken for granted. There are no laws of nature which create or necessitate these networks, after all; BBE networks are not the automatic byproducts of nomological necessities. Instead, the physical integrity of a BBE network is normally a causal byproduct of the BBE network's endogenous structure. Yet a durable, functioning BBE network is nevertheless something like an unplanned, accidental circus act: Things as massive as an elephant and as small as a mouse, as ephemeral as a soundtrack and as abstract as a set of linguistic descriptions and commands, all must find a way of interacting in a sustained, coordinated, and causally integrated fashion.

The analogy is imperfect, of course, partly because it isn't clear what (beyond entertainment) the functions of circus acts include—but the analogy is nevertheless helpful. The analogy is helpful because it calls our attention to the fact that the components of a BBE network must interact only in very specific ways for the BBE network as a whole to both maintain its physical integrity and to consequently realize complex functions such as learning, progressively increasing task proficiency, or even just contextually appropriate behavior. This in particular is puzzling. BBE networks, like circus acts, can be made out of component subsystems that are themselves organized at very different levels of energy and constructed out of very different physical formats. But, unlike circus acts, BBE networks seem to be capable of degrees of a functional self-regulation, which requires the component systems of a BBE network to be able to share information (or at least signals) with one another. Consequently, the fact that a BBE network can maintain its endogenous structure—and thereby preserve its physical integrity as well as its functional capabilities—gives rise to two overlapping metaphysical questions. First, how is it that component physical systems of BBE networks that are frequently organized at very different magnitudes of energy are able to be components of the same overall system? Second, how is it that the component systems of any BBE network are able to send information between themselves without also transmitting so much energy as to cause the BBE network itself to break apart? Or, putting the questions a bit more abstractly: What is it about the metaphysics of BBE networks that explains both why they can maintain their physical integrity and why they can realize various complex functions?

We think the work on the metaphysics of mechanisms which has occurred in the philosophy of science over the last two decades contains the seeds of an answer to these two metaphysical questions (Cummins, 2000; Tabery, 2004; Craver & Bechtel, 2007; Glennan, 2017; Matthews & Tabery, 2017; Love, 2018). This is because, considered very abstractly, a BBE network is a system formed out of a number of complex component mechanisms, which themselves frequently take the form of causal systems. Indeed, a useful way of simplifying the import of this literature for our account of the metaphysics of BBE networks is to see this literature as providing the impetus for drawing a very general distinction between mechanisms and systems of mechanisms (cf., Craver, 2001)—or, as we shall say, a very general distinction between *simple systems* and *systems of systems*, as this choice of words makes our terminology a bit more straightforward.

Here is how we want to define this distinction. First of all, simple systems are closed networks of causally interacting components, where the energy transferred between the components is of roughly the same magnitude. This fact can explain why a simple system maintains its physical integrity: It is hard, or impossible, for the simple systems' endogenous causal processes to acquire sufficient power to be able to break or destroy the system as a whole. Most things in nature, though, are not simple systems. They are usually (dynamic, complex, and/or emergent) systems of systems. A system of systems is a system whose components are systems that would be simple systems if it were possible to extract them from the system of systems without breaking the simple system itself. The crucial difference, then, is that, unlike a simple system, a system of systems can have component systems that are organized (or are stable) at substantially different magnitudes of energy. Thus, the universe is one extremely big system of systems—and so too are all BBE networks. Furthermore, our view that the mind is a computational system is, metaphysically speaking, the proposal that the mind be conceptualized as one simple system amongst others in the system of systems formed by any real-life BBE network.

The question before us now, however, is only: What is it about the metaphysics of BBE networks, *given that they are systems of systems and not simple systems*, that explains how these systems maintain both their causal integrity and their ability to realize any number of different functions? An example will help us answer this question. Suppose that there exists a BBE network made of a child bouncing a red ball in a well-lit room with no one else present. This system of systems has amongst its constituent systems causal processes manifest in the forms of kinetic energy (the bouncing ball), radiant energy (photons reflected from the ball to the retina), and chemical energy (hydrolysis in the retina). Each of the component systems is stable and able to contribute to the smooth functioning of the system of systems despite the fact that the energy involved in each of these component systems is multiple orders of magnitude greater or smaller than in the other systems. Put somewhat baldly: If, for instance, the total mechanical energy in the bouncing ball were transmitted directly into the brain, it would destroy at least the cortex. The integrity of a system of systems, then, depends on the ability to shield, insulate, protect, or otherwise buffer the component simple systems from one another. Or, put more generally, systems of systems are stable because of *causal buffering*. In this case, causal buffering is provided by, inter alia, the flexibility of the child's arm, the child's perceptual coordination capacities, and perhaps even the child's skull itself.

Yet, it is important to observe that perfect causal buffering—causal buffers that block *all* energy transfer between systems—would prevent the system of systems from realizing even the simplest of functions. If no energy whatsoever could follow a loop running between the child's brain and the ball, then bouncing the ball would be impossible. *A fortiori*:

A body, itself a system of systems, would not be able to maintain homeostasis if it were impossible for its component simple systems to interact with one another. So, in addition to causal buffering, systems of systems must allow component systems to share information—which we will call *metaphysical transduction*. In our example, metaphysical transduction is realized by the mechanisms which implement, inter alia, the child's proprioception of her arm's location, the mechanisms which convert electromagnetic radiation into biochemical energy via the process of visual phototransduction and eventually generate the child's input visual cues, feedback from different clusters of striated muscles, and information from afferent neurons in the child's hands—all of which ensure that the child maintains a sense of the ball's location relative to her own body and its own path of spatial movement.

It is easy to find examples of causal buffers and metaphysical transducers that respectively hold together systems of systems and permit the system as a whole to have any number of uses and functions. Take, for example, any commercially produced car. When running, the engine produces vibrations which, if not absorbed, would shake the engine apart. The engine is buffered against itself by, inter alia, ensuring that its heaviest moving parts are in mechanical balance, increasing the mass of the engine block, placing the camshaft above the combustion chamber, and mounting the engine to the chassis using extremely durable rubber vibration dampeners. But vibration isn't the only kind of energy that threatens the physical integrity of the car. A separate array of buffers is used to control the heat generated by the engine; in most cases, this is the function of the radiator, but the radiator cannot perform its function if the oil which provides lubrication is not also buffering parts of the engine from the damaging effects of heat that is caused by friction.

Cars, of course, are meant to be driven; the steering system provides us with an elegant example of an interlocking chain of metaphysical transducers connecting the steering wheel to the front wheels. In a rack-and-pinion layout, for instance, the steering wheel turns a column which then turns a pinion gear that is meshed with a linear gear fixed atop a rack—the net effect of which is to turn the radial motion of the steering wheel into the horizontal motion of the rack itself. The rack is connected by way of tie rods (which act as both causal buffers and metaphysical transducers) to the king pin, which turns the wheels. If the strength of the driver is insufficient to produce enough torque to turn the wheels, the steering system will have hydraulic or electric actuators which amplify the steering inputs produced by the driver. Similar chains of metaphysical transducers connect the gas pedal with the throttle, the brake pedals with the brakes, and the numerous electronic control units (the PCU, TCU, and so on) with various subsystems endogenous to the system of systems that is any car. (And to foreshadow: We think it is not accidental that the computational systems embedded in the car's ECU, for instance, dramatically increase the number of scenarios that the car's engine can operate *optimally* within.)

So, just as the causal buffers and metaphysical transducers built into a car explain why the car does not explode, melt, or shake itself to bits, and also how signals are able to pass between different component systems such that the car is able to drive, so too will there be parts of BBE networks that function as causal buffers and metaphysical transducers, and which therefore explain why BBE networks can maintain their causal integrity while realizing different functions as complex as learning, or even comparatively simpler functions, such as planning, playing, mindreading, or just absentmindedly bouncing a ball.

3. Innateness as Developmental Essentiality

We have begun with a very general analysis of the metaphysics of systems of systems because it leads us to an important insight into the architecture of the cognitive system: Since the cognitive system is a simple system within a larger system of systems, it too should have its own causal buffers and metaphysical transducers. Moreover, at least some of the relevant buffers and transducers must be innate—for it is the innateness of at least some of the mind's buffers and transducers that explains how the cognitive system can develop. The cognitive system, just like the other component systems of BBE networks, does not just appear out of nowhere. All of these systems are causal byproducts of "precursor" systems, and our conception of innateness provides an explanation of how this is possible.

However, this line of reasoning depends upon a new conception of innateness; the difference in meaning between how we shall use the concept and how it is customarily used in philosophy, biology, and psychology is large enough to warrant a formal definition. Accordingly, we will start this section with an explanation of our conception of innateness, one that is designed to fit within the dynamic systems worldview, before turning to an explanation of how this concept can be used to explain the development of new simple systems within a larger system of systems.

The concept of innateness is customarily used to denote traits that are in some sense fixed or immalleable, such that what makes a trait innate is something like its invariance under different kinds of developmental, genetic, or environmental pressures.[1] We want to use, instead, a concept of innateness that expresses the idea that traits are innate because they are developmentally essential, and, for that, not *necessarily* fixed and immalleable over the full temporal duration of the system in which the trait is a part. This idea can be unpacked by returning to the question of how a new simple system can develop within an existing system of systems. Indeed, the biological world provides countless examples of systems of systems that have amongst their functions the power to, from time to time, produce a new simple system. When this happens, there will be a period of time during which the "parent" or "precursor" system overlaps with, and therefore shares some of the parts of, the "child" or "derivative" system—life, after all, does not begin or end; it is only selectively transmitted.

The idea, then, is that innate traits will be the components of the "child" systems that are byproducts of the operation of a "parent" system, which can, after a certain amount of time, become elements of the "child" system. These traits will also be essential to stabilizing the functions of the "child" system *because* they provide, at least initially, the causal buffering and metaphysical transduction needed for the "child" system to separate from the "parent" system, all without disrupting the functions realized by the overarching system of systems. Or, to put the same idea a different way, what makes a trait innate is time and system-relative: The innate traits of a simple system are just those traits which are amongst the initial parts of a new simple system and which are causally necessary for the new system to become a discrete system when the system is itself the causal byproduct of the operation of other simple systems in a system of systems.

That is the abstract outline of the concept; we can further clarify it by defining it explicitly. Thus, according to this new concept, a trait or mechanism is innate if and only if:

- The trait is *developmentally essential* to at least one of the simple systems that it is a part of; without this trait, the system in question cannot come into existence.
- The trait exists, proximately speaking, because it is a causal byproduct of one of the systems that it is *not* a developmentally essential part of.
- For at least a meaningful period of time after the trait comes into being, the trait can only modify, but not be modified by, causal processes that are endogenous to at least one of the systems that the trait is a developmentally essential part of.

Now, since our intent is only to use *this* concept of innateness—not to argue that it is something like the one *single* true concept of innateness—it will suffice to justify our characterization of certain cognitive mechanisms as innate using this concept by finding evidence that our tripartite definition is not empirically vacuous. Consider, thus, the genome of any organism—it will be innate by our definition. In sexually reproducing species, the processes of meiosis and fertilization that create a unique set of chromosomes occur in physiological systems that almost always lose the set of chromosomes as a part, but the same set of chromosomes is a developmentally essential component of a great number of different physiological systems. Finally, while complex feedback loops regulate the causal powers of a genome throughout much of its existence (Jablonka et al. 2014), the earliest stages of cell growth and differentiation are mostly biochemical effects of the genome itself (cf., Reik et al. 2001; Mizushima & Levine 2010).

As noted above, this concept of innateness allows us to say that some trait is innate *relative* to a particular system, but not innate relative to another system—even if the trait, for some non-trivial period of time, is a part of the second system. This is important because it is not possible for new systems to develop within existing systems of systems without the new system sharing most, if not all, of its component mechanisms with a precursor system for at least a short period of time. So, this conception of innateness allows us to distinguish between a trait being a byproduct of a parent system and thus not innate relative to the parent system yet nevertheless being developmentally essential to a child system and thus innate relative to this second system.

Because of its ability to mark out this distinction, this concept of innateness allows us to express the idea that the ability of new simple systems to develop within systems of systems is only possible because certain causal buffers and metaphysical transducers are innate—even if some of the buffers and transducers are either effects, or even parts of, the precursor system. Put more concretely, the idea is that, just as some of the brain's innate traits (e.g., the blood–brain barrier, which is itself a system of systems) explain how it emerges as a stable simple system within a system of systems constituted by bodily and environmental networks, some of the mind's innate traits can *explain* how a computational system emerges as a distinct system within a system of systems.

What might these innate transducers and buffers be? Amongst the transducers must be mechanisms that are able to convert streams of different non-cognitive signals into cognitive information, and also mechanisms which convert cognitive information into non-cognitive signals. Amongst the causal buffers, there must be mechanisms that permit a computational system to remain sufficiently insulated from potentially interfering forces for it to remain co-instantiated with the brain's neurological networks and systems.

Ultimately, we are most interested in the former—since digging a bit deeper into empirical theories of the mind's innate metaphysical transducers might lead to the intriguing conclusion that there are probably a number of innate concepts. However, before turning directly to the question of whether there are any innate concepts, we want to first return to Smith et al.'s argument that the dynamism of BBE networks is a reason to be skeptical of the existence of cognition. Exploring what it means to say that the cognitive system is co-instantiated with a variety of other systems sheds some light on what some of the mind's innate causal buffers may be.

4. Co-instantiation, Computation, and Degeneracy

Smith et al. are skeptical that the discrete and static states of any computational system can be functionally located within the brain. Turing also dealt with this problem. In the paper largely responsible for introducing the computational theory of cognition, Turing offers the following observations:

> [Discrete state machines] are the machines which move by sudden jumps or clicks from one quite definite [i.e. static] state to another. These states are sufficiently different for the possibility of confusion between them to be ignored. Strictly speaking there are no such machines. Everything really moves continuously. But there are many kinds of machines which can profitably be *thought of* as being discrete state machines. For instance in considering the switches for a lighting system it is a convenient fiction that each switch must be definitely on or definitely off. There must be intermediate positions, but for most purposes we can forget about them.
>
> (Turing, 1950, p. 439, emphasis in the original)

Turing's uses of "thought of" and "convenient fiction" are usefully ambiguous. One interpretation of what Turing means to say is that discrete state machines, and therefore digital computers, do not exist *simpliciter*. But this interpretation is contradicted by Turing's subsequent assertion that it is possible to build discrete state machines that are digital computers, but only if the physical system out of which both are built reduces to almost nil the chance that the continuously moving, dynamically interacting physical parts will cause the computer to depart from its programming. This suggests that the alternative interpretation which more accurately captures Turing's intended meaning is one which reads him as saying that discrete state machines, and therefore digital computers, and dynamic physical systems can be (and frequently are—think of all of the switches you have used today) co-instantiated. And one way of unpacking the meaning of the co-instantiation thesis is seeing that it implies that discrete state machines cannot actually be built *simpliciter*. We cannot build a physical system that is also a digital computer and which has exactly zero chance of departing from its programming. Nevertheless, we can build physical devices the operations of which are so *extremely* well-aligned with the operations of hypothetical zero-error computers that what gets built is a physical system that is co-instantiated with a non-zero-error (and therefore quasi-) computer.

Thus, there are two simple systems that are co-instantiated in any real-world digital computer: the continuous (or dynamic) physical components of the system and the static computational components of the (non-zero-error) computer system itself. The key point, then, is that the former so closely mirrors the operations of an entirely hypothetical zero-error computing machine that nothing is lost by thinking of the real non-zero-error quasi-computer system as if it is really the hypothetical zero-error computing machine. Turing is denying only the physical reality of zero-error digital computers, and asserting that non-zero-error computers can be co-instantiated with all sorts of physical systems.

This shows that it is conceptually possible for computational systems to be co-instantiated with larger dynamic systems. But this does not completely answer the question of how the static states of a computational network can be co-instantiated with the dynamic networks of the brain. The crux of the issue is that the dynamism of neural networks makes brains highly variable, both across individuals and over meaningful periods of developmental time. As Smith et al. stress, the dynamic properties of different brain networks, and the massively differential impact that variations in both behavior and environment can have on brain development, mean that patterns of neural connectivity are extremely variable from individual to individual, from behavioral context to behavioral context, and from environmental context to environmental context. Turing's observation that static systems can be co-instantiated with dynamic systems does not help us address the question of how a static system with the same functionality (say, implementing the inferential processes that infer edges from stereopsis) can be co-instantiated with a very large set of inherently different connective systems.

Put more precisely, however, this problem really just is the problem of explaining how there can be coincident causal realization of two systems without there being a homomorphism between the structures of the two systems. And this problem is solved by evidence that a one-to-many relationship holds between the functional organization of the computational mind and different neural networks. This, in turn, amounts to evidence that the brain has substantial amounts of what Edelman (1987; Tononi et al., 1999; Edelman & Gally, 2001) calls *degeneracy*:

> Degeneracy is the ability of elements that are structurally different to perform the same function or yield the same output. Unlike redundancy, which occurs when the same function is performed by identical elements, degeneracy, which involves structurally different elements, may yield the same or different functions depending on the context in which it is expressed. It is a prominent property of gene networks, neural networks, and evolution itself. Indeed, there is mounting evidence that degeneracy is a ubiquitous property of biological systems at all levels of organization.
>
> (Edelman & Gally, 2001)

Importantly, degeneracy is an empirical concept. With his collaborators, Edelman has shown that there are high levels of degeneracy in the brain's neural networks—a result that has been used by several subsequent researchers to explain how different computational functions can be co-instantiated with the different forms of connectivity inherent to any living brain (Eliasmith & Anderson, 2004; Eliasmith, 2007; Park & Friston, 2013).[2] Indeed, in an earlier article, Smith herself recognizes the importance of degeneracy: "The notion of degeneracy in neural structure means that any single function can be carried out by more than one configuration of neural signals and that different neural clusters also participate in a number of different functions" (Smith, 2005, p. 290). And finally, the brain is *innately* degenerate: Degeneracy is developmentally essential for the emergence of all neurological networks which share at least some physiological resources.

It should therefore be unsurprising that digital computers provide another example of degeneracy, albeit at the level of instruction set architecture. A microprocessor's instruction set specifies what computational functions that processor can perform—familiar architectures include the original x86 specification, extensions to it like SSE and AMD64, and the growing family of the ARM specifications. The circuits etched into silicon which implement these instruction sets can be radically different: There are thousands of very different microprocessor chips which have, for instance, the function of implementing the 32-bit variant of x86. (Technically, these circuits are non-zero-error implementations of the relevant instruction sets.) Transistors are a different example of a simpler form of degeneracy in electrical engineering: There are now thousands of different physical systems out of which transistors can be built. Finally, most field programmable gate arrays provide us with examples of degenerate computational systems that are co-instantiated with highly dynamic physical systems.

What's more, these observations show us something interesting about the notion that there are *literally* levels of analysis or levels of explanation that both sit within the domain of psychology and also separate psychology from other fields in the cognitive and behavioral sciences. The idea, put roughly, is that the theories of one field will not reduce to the theories of another field because they are about different metaphysically discrete *layers* or *planes* of reality. The disciplinary structure of the cognitive and behavioral sciences mirrors the layered organization of nature, or at least the cognitive and behavioral parts of nature: Each discipline studies a horizontally organized plane formed of phenomena that interact with phenomena on its plane only, and interact according to laws or generalizations that apply to that plane only (cf. Fodor, 1974; Fodor, 1997). Planes that are below provide some kind of ontological or metaphysical support for planes that are above, but, despite this, the laws of a lower or higher plane do not apply to any phenomena except those which occur on the plane itself. And there aren't "bridge laws" either—these would be "vertical" laws that connect the projectible terminology of one disciplinary vocabulary with the vocabulary of another discipline, where the vocabularies apply to different planes, and where the bridge laws serve to establish synonymous definitions for some of the concepts from the first discipline in terms of concepts from the second discipline. This is, we suggest, a rough sketch of the popular picture in the philosophy of mind (cf., Bermudez, 2007).

Yet, it is not a picture that we can wholly endorse. We are happy with the notion that there are levels of analysis, so long as this is taken only as a methodological metaphor (Boyd, 1993), i.e., a metaphor calling attention to certain facts from the research history of the cognitive sciences—facts such as that you cannot do all of the work that is interesting and projectible in cognitive psychology using the methods and concepts of neuroscience (and vice versa). But we have to stop at the point at which the metaphor of metaphysical levels of analysis gets turned into a theory of the fundamental ontological organization of reality according to which reality is literally organized into planes that have some kind of inherent or objective top-to-bottom geometry which allows us to order these planes in relation to one another. We cannot accept this theory—again, despite its apparent popularity amongst some philosophers of mind (cf., Kim, 1990; McLaughlin & Bennett, 2018)—because our commitment to the co-instantiation of computational systems with physical systems means that we are committed to a host of complex causal interactions between any (non-zero-error) computational system and the physical system with which it is co-instantiated. These causal interactions must occur in order for the computational system to be appropriately causally buffered, and for the computational system to play a role, with the help of certain metaphysical transducers, in supporting the functions realized by whatever overarching system of systems the computational system is a constituent of. Or, to put the same idea another way, we think that nature is a single plane—that of all physical stuff—and that, in some sense, almost everything is co-instantiated with something else. But there are also naturally occurring systems—and systems of systems, and systems of systems of those systems, etc.—that are sustained by all sorts of different kinds of causal buffering, and it is these complexes of causal buffers which, in turn, explain the persistence of systems like the cognitive system, but also the body's various physiological systems, and even large-scale systems like a national economy or a whole ecosystem. We think that scientific disciplines frequently succeed in their efforts to construct conceptual schemes which are mostly proprietary tools for referring to the endogenous causal activity of these systems—and that this is enough to explain why cognitive psychology is (literally) about a different set of phenomena than, inter alia, cognitive neuroscience, neuroanatomy, neurophysiology, and so on. Accordingly, we do not think that the recognition that the cognitive sciences are autonomous relative to one another implies a metaphysical theory according to which there are layers of reality organized in some top-to-bottom fashion according to some *a priori* metric of fundamentality (cf., Davidson, 1973).

But let us get back to the specific case concerning the co-instantiation of a (non-zero-error) computational system with different physical systems. This specific co-instantiation shows that it is *scientifically plausible* to adopt the position that there need not be a homomorphism between the structures of two or more physical systems that, in turn, are co-instantiated with computational systems that have the same function. Evidence that the brain's neural networks are extremely dynamic supports no meaningful *a priori* conclusions about the possible structures of any systems, computational or otherwise, that are co-instantiated with these networks. For all we know, the clusters of properties which constitute an interesting kind at one level of causal interaction (cognitive computation) may, at another level of causal interaction (neural connectivity), form no interesting clusters at all. As a purely conceptual matter, then, it is possible for a computational system to be co-instantiated with a dynamic neurological system. This conclusion is enough to refute the major premise of Smith et al.'s argument.

More important, however, is the observation that co-instantiation and degeneracy also explain some aspects of how the cognitive system is causally buffered, which thereby explains why cognitive psychology is an autonomous discipline. Co-instantiation means that the physical mechanisms and processes which (sometimes literally) insulate different neurological networks from external disruption and interference can confer the same benefit upon the cognitive system, too: Given that they are co-instantiated, whatever mechanisms buffer the brain's non-cognitive neurological networks also buffer the brain's cognitive networks. Furthermore, the brain's inherent degeneracy explains why dynamic changes in neurological networks need not induce changes in computational function: Degeneracy explains how there can be an island of (relative) computational stability in a sea of (again, relative) constant neurological and physiological change. Thus, degeneracy buffers computational function from change caused by ongoing patterns of change in the physical systems which realize the relevant (non-zero-error) computational systems. Or, put another way, the brain's *innate* degeneracy is likely amongst the most important causal buffers for the computational system that is any human mind.

5. The Innateness of the Initial Conceptual Repertoire

We turn now to the issue of whether there are innate concepts. We know that the mind must contain its own innate metaphysical transducers—the function of which is of course to convert into cognitive information various non-cognitive signals available to the mind as the output of the body's own suite of sensory transducers. And, here, it is important to keep metaphysical questions and scientific questions separate, because the conclusion that the mind must contain a number of innate metaphysical (or cognitive) transducers does not provide an answer to the many and much more difficult scientific questions about the specific empirical form that these transducers take.

That said, there are many sophisticated theories of the empirical form of the mind's innate cognitive transducers to choose from (cf., Samuels, 2000; Marcus, 2006; Smolensky & Legendre, 2006; Heyes, 2018; Schulz, 2018). We believe that one of the most conservative scientific accounts of the mechanisms likely responsible for the earliest forms of non-cognitive-to-cognitive-transduction comes from Susan Carey. According to Carey, the relevant mechanisms should be thought of as dedicated input analyzers: "A dedicated input analyzer computes representations of one kind of entity in the world and only that kind. All perceptual input analyzers are dedicated in this sense: the mechanism that computes depth from stereopsis does not compute color, pitch, or number" (Carey 2011a, p. 451). If (as we have argued is the case) some of these input analyzers are innate, it follows that there must also be a handful of innate concepts as well, namely whichever concepts are embedded in these dedicated input analyzers and which allow the analyzers to produce as output information that is richer—because it contains more structure, or is more abstract, or refers to an unobserved kind or process—than the information that is the input to the analyzer. Whatever else they are, concepts are what represent unobserved or unobservable properties and kinds.

There is compelling evidence that young children have abstract concepts for objects, agents, numbers, and probably also causes (Xu & Carey, 1996; Wang & Baillargeon, 2006; Carey, 2011a; Baillargeon et al., 2012). Our proposal, then, is that the hypothesis that there are innate input analyzers dedicated to generating conceptual representations of objects, agents, numbers, and causes represents the most empirically plausible way of cashing out the more abstract metaphysical conclusion that some metaphysical transduction must take place in order to transform non-cognitive information into cognitive (i.e., computationally tractable) information.

Carey characterizes the mind's innate conceptual resources the following way: "What I mean for a representation to be innate is for the input analyzers that identify the represented entities to be the product of evolution, not the product of learning, and for at least some of its computational role to also be the product of evolution" (Carey, 2011a, p. 453). But she also resists defining innateness in terms of static, fixed, or non-malleable properties:

> Some innate representational systems serve only to get development started. The innate learning processes (there are two) that support chicks' recognizing their mother, for example, operate only in the first days of life, and their neural substrate actually atrophies when the work is done. Also, given that some of the constraints built into core knowledge representations are overturned in the course of explicit theory building, it is at least possible that core cognition systems themselves might be overridden in the course of development.
>
> (Carey, 2011b, p. 117)

Her commitment to the view that at least some dedicated input analyzers are innate dovetails with the definition of innateness as developmental essentiality that we introduced above. Consequently, the view that some concepts are innate because they are embedded in innate metaphysical transducers avoids the difficulty of accounting for how the rich and complex conceptual repertoire of most adults' minds can be built out of innate concepts. On this view, the innate conceptual resources are needed only to get learning started, and not to provide the ingredients for all concepts learned over the whole of cognitive development.[3] Whether or not these resources persist, and if so for how long, are empirical problems left open by the definition of innateness as developmental essentiality and which remain, so far as we know, unresolved.

But, as a scientific matter only, Carey could be wrong. It could be that further research yields compelling reasons to be skeptical of the existence of a suite of innate, dedicated input analyzers. Nevertheless, were that to occur, there would still be good metaphysical reasons to remain committed to the existence of innate transducers which mediate information transfer between the cognitive system and the other systems in BBE networks.

That said, this line of reasoning does not touch on a deeper outstanding problem: Why should anyone posit cognition at all? Metaphysical transducers are necessary *only if* we must explain how the cognitive system plays a functional role within a larger system of systems forming a BBE network. If cognition is not real, then there is no reason to posit innate cognitive transducers that come equipped with at least a handful of endogenous conceptual representations.

6. Consilience and the Choice between the BBE and the CBBE View

So, with that, we now arrive at the argument for preferring our extended view over Smith et al.'s restricted view. The specific question here is whether a cognitive system should be posited along with the systems (of systems) constituting the body, the brain, and the environment. We believe that you should posit CBBE networks instead of only BBE networks because doing so allows you to explain more scientific data than is otherwise possible.

What kind of scientific data can only be explained by the extended CBBE view? This is any data that fits David Marr's operational definition of psychological computation: A psychological process is computational if the process can be "characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated" (Marr, 2010, p. 24). Note that Marr's definition refers to two logically distinct properties, properties that are jointly necessary in order to establish evidence of computational processes. The first is evidence of some kind of mapping between sets of information, such that the latter set can be treated as some kind of (possibly amplitative) transformation of the former set. The second is some kind of evidence that the relevant transformation is normatively appropriate for the situation or context: In some non-arbitrary sense, it is one a mind *should do*. Put another way, then, empirical evidence of psychological computation just is evidence of the rational processing of information (cf., Rumelhart & McClelland, 1985).

And there is a very large amount of exactly that kind of evidence. For example, consider two decades' worth of experiments that, taken together, demonstrate that both children and adults make inferences that seem to reflect unconscious knowledge of certain basic principles of logic and probability (Xu, 2007; Xu & Tenenbaum, 2007 Buchsbaum et al., 2011; Denison & Xu, 2012; Xu & Kushnir, 2012; Xu & Kushnir, 2013; Gopnik & Bonawitz, 2015; Wellman et al., 2016;). Indeed, by about the age of 4, children have the ability to recognize when information is relevant (Southgate et al., 2009), when information is supportive of generalizations (Sim & Xu, 2017), when information is evidence of causation (Gopnik et al., 2004; Sim et al., 2017), and when information can be expressed on an ordinal scale (Hu et al., 2015). Of course, the mind's sensitivity to these different kinds and uses of information is not neutral: Frequently, information is used as evidence—that is, the information is used to drive changes in belief or motivation, changes that are themselves consistent with certain deep principles of rationality (Xu, 2007; Xu, 2011; Fedyk & Xu, 2017). This information is used, that is to say, in roughly the way it *should* be used *if* it is to be used *rationally*, satisfying Marr's operational definition of computational processing.

This evidence demonstrates that there is a meaningful scientific choice between a theory of cognitive development that posits only components of BBE networks and a theory of cognitive development that posits, in addition, the existence of a *sui generis* cognitive system. Considerations of scientific consilience (cf., Wilson, 1999; Cantor et al., 2018) favor the latter theory of cognitive development—since only a theory which posits a cognitive system that is a computational system is able to explain *both* the impressive amount of empirical data that Smith et al. survey in their chapter, *and* the scientific data that is evidence of computational processing. More scientific data can be explained by our extended view of the metaphysics of cognitive development than Smith et al.'s more restricted ontology.

7. Conclusion

The picture of cognitive architecture that we want to endorse is, at bottom, this: There is a set of designated input analyzers that are innate to the cognitive system itself, plus a central reasoning system that scientists can study and understand by relying upon principles of rationality. The system as a whole is stable because the brain's innate degeneracy acts as one causal buffer—and not the only causal buffer—for the cognitive mind. The cognitive system is co-instantiated with the brain's dynamic networks—and, in this way, it is no different than all other real-world computational systems, given that (non-zero-error) computational systems are always, and can only be, co-instantiated with physical systems.

So, where is the mind? It is somewhere between the ears and behind the eyes, because it is co-instantiated with the brain. However, if we are instead asking about the functional location of the mind, then we can now be slightly more precise. The mind's functional location is given by asking how embedding a computational system within a BBE network extends the functional capabilities of the network. The most consequential of these increases seems to be to allow the resulting CBBE network to realize patterns of normative thought, which thereby dramatically amplify the range of context-appropriate behaviors available to the network. Or, put more simply, a cognitive system confers *rationality*—the capacity for different kinds of (epistemic, statistical, logical, moral, practical) principles to influence thought and regulate behavior. This dramatically increases the range of learning that is possible for our species (Tomasello, 2014), but it also dramatically deepens the sources of error, confusion, and mistakes as well. After all, it is only by having a mind that someone can seem to discover reasons to doubt the existence of the same.

Notes

[1] That said, Smith et al. are correct that there is no established definition of "innate"—for different examples see Kitcher (2001), Griffiths & Machery (2008), Griffiths (2002), and Ariew (1996). Of course, this shows neither that innateness is not real nor that the various definitions are confused or incoherent. In fact, we should expect a small family of potentially incommensurable concepts for some kind to develop as a byproduct of routine scientific investigation into the kind. Clusters of incommensurable concepts may sometimes be signs of inductive progress; we are, therefore, happy to add another concept to the cluster.

[2] See also Bullmore & Sporns (2012) and Dehaene & Changeux (2011) for richer analyses of how different mental functions may stand in a many-to-one relationship with various forms and instantiations of neuronal connectivity. See also Aizawa (2015) for discussion of several complementary philosophical issues.

[3] It is also helpful to point out that dedicated input analyzers and their innate conceptual resources are not necessarily Fodorian modules. Fodorian modules are a type of non-cognitive to cognitive metaphysical transducer, but they are not the only possible transducer which can perform that function. To see this, consider how Fodor describes a hypothetical module:

> A parser for [a language] L contains a grammar of L. What it does when it does its thing is, it infers from certain acoustic properties of a token to a characterization of certain of the distal causes of the token (e.g., to the speaker's intention that the utterance should be a token of a certain linguistic type). Premises of this inference can include whatever information about the acoustics of the token the

> mechanisms of sensory transduction provide, whatever information about the linguistic types in L the internally represented grammar provides, and nothing else.
>
> (Fodor, 1984, p. 37)

Separately, Fodor discusses cognitive transducers (Fodor, 1987). Furthermore, note that transduction is mentioned by Fodor, but it refers to processing prior to the module. But this language parser, too, is a metaphysical transducer: It converts acoustic information into lexical (or *lexicalizable*) information. It is an example, thus, of a transducer operating on the output of a transducer. So, again, Fodorian modules are a kind of metaphysical transducer, but they are not the only kind.

References

Aizawa, K. (2015). What is this cognition that is supposed to be embodied? *Philosophical Psychology*, 28(6), 755–75.

Ariew, A. (1996). Innateness and canalization. *Philosophy of Science*, 63, S19–S27.

Baillargeon, R., et al. (2012). Object individuation and physical reasoning in infancy: An integrative account. *Language Learning and Development: The Official Journal of the Society for Language Development*, 8(1), 4–46.

Bermudez, J. L. (2007). *Philosophy of Psychology: Contemporary Readings*. London: Routledge.

Boyd, R. N. (1993). Metaphor and theory change. In A. Ortony (Ed.), *Metaphor and Thought*, second edition. Cambridge: Cambridge University Press.

Buchsbaum, D. et al. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, 120(3), 331–40.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews: Neuroscience*, 13(5), 336–49.

Cantor, P. et al. (2018). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science*, 23, 1–31.

Carey, S. (2011a). *The Origin of Concepts*, reprint edition. Oxford: Oxford University Press.

Carey, S. (2011b). Precis of "The Origin of Concepts." *Behavioral and Brain Sciences*, 34(3), 113–24.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68(1), 53–74.

Craver, C. F. & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy*, 22(4), 547–63.

Cummins, R. (2000). "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and Cognition*, 117–45. Cambridge, MA: MIT Press.

Davidson, D. (1973). On the very idea of a conceptual scheme. *Proceedings and Addresses of the American Philosophical Association*, 4. http://www.jstor.org/stable/3129898

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–27.

Denison, S., & Xu, F. (2012). Probabilistic inference in human infants. *Advances in Child Development and Behavior*, 43, 27–58.

Edelman, G. M. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 98(24), 13763–8.

Eliasmith, C. (2007). How to build a brain: From function to implementation. *Synthese*, 159(3), 373–88.

Eliasmith, C., & Anderson, C. H. (2004). *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. Cambridge, MA: MIT Press.

Fedyk, M., & Xu, F. (2018). The epistemology of rational constructivism. *Review of Philosophy and Psychology*, 9(2), 343–62. https://doi.org/10.1007/s13164-017-0372-1

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28(2), 97–115.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (1984). Observation reconsidered. *Philosophy of Science*, 51(1), 23–43.

Fodor, J. A. (1987). Why paramecia don't have mental representations. *Midwest Studies in Philosophy*. http://onlinelibrary.wiley.com/doi/10.1111/j.1475–4975.1987.tb00532.x/full

Fodor, J. A. (1997). Special sciences: Still autonomous after all these years. *Noûs*, 31, 149–63.

Glennan, S. (2017). *The New Mechanical Philosophy*. Oxford: Oxford University Press.

Gopnik, A. et al. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111(1), 3–32.

Gopnik, A., & Bonawitz, E. (2015). Bayesian models of child development. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(2), 75–86.

Griffiths, P. E. (2002). What is innateness? *Monist*, 85(1), 70–85.

Griffiths, P. E., & Machery, E. (2008). Innateness, canalization, and "biologicizing the mind." *Philosophical Psychology*, 21(3), 397–414.

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge, MA: Belknap Press, an imprint of Harvard University Press.

Hu, J. et al. (2015). Preschoolers' understanding of graded preferences. *Cognitive Development*, 36, 93–102.

Istrail, S., De-Leon, S. B.-T., & Davidson, E. H. (2007). The regulatory genome and the computer. *Developmental Biology*, 310(2), 187–95.

Jablonka, E., Lamb, M. J., & Zeligowski, A. (2014). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, revised edition. Cambridge, MA: MIT Press.

Kim, J. (1990). Supervenience as a philosophical concept. *Metaphilosophy*, 21(1–2), 1–27.

Kitcher, P. (2001). Battling the undead: How (and how not) to resist genetic determinism. *Thinking About Evolution: Historical, Philosophical, and Political Perspectives*, 2, 396–414.

Koutroufinis, S. A. (2017). Organism, machine, process: Towards a process ontology for organismic dynamics. *Organisms: Journal of Biological Sciences*, 1(1), 23–44.

Love, A. (2018). Developmental mechanisms. In S. Glennan & P. Illari (Eds.), *The Routledge Handbook of the Philosophy of Mechanisms*. New York: Routledge.

Marcus, G. F. (2006). Cognitive architecture and descent with modification. *Cognition*, 101(2), 443–65.

Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.

Matthews, L. J., & Tabery, J. (2017). Mechanisms and the metaphysics of causation. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, 115. London: Routledge.

McLaughlin, B., & Bennett, K. (2018). Supervenience. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2018/entries/supervenience/

Mizushima, N., & Levine, B. (2010). Autophagy in mammalian development and differentiation. *Nature Cell Biology*, 12(9), 823–30.

Packard, A. (2001). A "neural" net that can be seen with the naked eye. In W. Backhaus (Ed.), *Neuronal Coding of Perceptual Systems*, 397–402. Singapore: World Scientific Publishing Co.

Park, H.-J., & Friston, K. (2013). Structural and functional brain networks: From connections to cognition. *Science*, 342(6158), 1238411.

Păun, G., & Rozenberg, G. (2002). A guide to membrane computing. *Theoretical Computer Science*, 287(1), 73–100.

Reik, W., Dean, W., & Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science*, 293(5532), 1089–93.

Rumelhart, D. E., & McClelland, J. L. (1985). Levels indeed! A response to Broadbent. *Journal of Experimental Psychology: General*, 114(2), 193–7.

Samuels, R. (2000). Massively modular minds: Evolutionary psychology and cognitive architecture. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Modularity, Language and Meta-cognition*, 13–46. Cambridge: Cambridge University Press.

Schulz, A. W. (2018). *Efficient Cognition: The Evolution of Representational Decision Making*. Cambridge, MA: MIT Press.

Siegelmann, H. T., & Sontag, E. D. (1995). On the computational power of neural nets. *Journal of Computer and System Sciences*, 50(1), 132–50.

Sim, Z. L., Mahal, K. K. & Xu, F. (2017). Learning about causal systems through play. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.* https://mindmodeling.org/cogsci2017/papers/0210/paper0210.pdf

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2- and 3-year-old children. *Developmental Psychology*, 53(4), 642–51.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review: DR*, 25(3), 278–98.

Smolensky, P., & Legendre, G. (2006). *The Harmonic Mind: Cognitive Architecture*. Cambridge, MA: MIT Press.

Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, 12(6), 1013–19.

Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science*, 71(1), 1–15.

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), 3257–62.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, 59(236), 433–60.

Wang, S.-H., & Baillargeon, R. (2006). Infants' physical knowledge affects their change detection. *Developmental Science*, 9(2), 173–81.

Wellman, H. M. et al. (2016). Infants use statistical sampling to understand the psychological world. *Infancy: The Official Journal of the International Society on Infant Studies*, 21(5), 668–76.

Whittle, A. (2007). The co-instantiation thesis. *Australasian Journal of Philosophy*, 85(1), 61–79.

Wilson, E. O. (1999). *Consilience: The Unity of Knowledge*. New York: Vintage Books.

Xu, F. (2007). Rational statistical inference and cognitive development. *The Innate Mind: Foundations and the Future*, 3, 199–215.

Xu, F. (2011). Rational constructivism, statistical inference, and core cognition. *Behavioral and Brain Sciences*, 34(03), 151–2.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, 30(2), 111–53.

Xu, F., & Kushnir, T. (2012). *Rational Constructivism in Cognitive Development*. Cambridge, MA: Academic Press.

Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, 22(1), 28–32.

Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–72.

Further Readings for Part II

Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.
Highly influential book uses work from developmental psychology to argue for the existence of innate concepts within core cognitive systems as well as the ability to acquire novel concepts via a process of Quinean bootstrapping.

Cowie, F. (1999). *What's Within? Nativism Reconsidered*. New York: Oxford University.
Draws on work from across cognitive science to critique arguments for nativist theses regarding language acquisition and innate concepts from Noam Chomsky and Jerry Fodor, respectively.

Fodor, J. A. (1981). The present status of the innateness controversy. In J. A. Fodor (Ed.), *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
The strongest statement of Fodor's notorious argument for radical concept nativism.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, 60, 115–40.
Draws on work in developmental psychology to argue that innate capacities, social input, and direct observation are all important influences on children's acquisition of concepts.

Gross, S., & Rey, G. (2012). Innateness. In E. Margolis, R. Samuels, & S. Stich (Eds.), *Oxford Handbook of Philosophy of Cognitive Science*, 318–60. New York: Oxford University Press.
A lengthy survey of the contemporary debate on what innateness is and whether concepts are innate.

Prinz, J. J. (2004). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.
Draws on philosophical and empirical literature to argue for a new form of concept empiricism and against nativist theses from Chomsky and Fodor.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89–96.
Draws on research on non-human primates as well as human infants, children, and adults to argue for the existence of domain-specific systems of core cognition that represent objects, actions, numbers, places, and potentially social partners.

Study Questions for Part II

1) According to Smith and colleagues, what is the brain–behavior–environment network, and how can it explain human behavior without positing the existence of concepts?
2) According to Smith and colleagues, how have recent advances rendered moot traditional questions about innateness?
3) According to Smith and colleagues, which questions should we investigate in place of questions about the origins of concepts?
4) According to Fedyk and Xu, which concepts are innate, and what makes them innate?
5) According to Fedyk and Xu, how does positing innate concepts help explain phenomena that could not otherwise be adequately explained?
6) Why do Fedyk and Xu disagree with Smith and colleagues about the existence of a *sui generis* cognitive system?