9 A Category Theory Explanation for Systematicity: Universal Constructions

Steven Phillips and William H. Wilson

1 Introduction

When, in 1909, physicists Hans Geiger and Ernest Marsden fired charged particles into gold foil, they observed that the distribution of deflections followed an unexpected pattern. This pattern afforded an important insight into the nature of atomic structure. Analogously, when cognitive scientists probe mental ability, they note that the distribution of cognitive capacities is not arbitrary. Rather, the capacity for certain cognitive abilities correlates with the capacity for certain other abilities. This property of human cognition is called systematicity, and systematicity provides an important clue regarding the nature of cognitive architecture: the basic mental processes and modes of composition that underlie cognition—the structure of mind.

Systematicity is a property of cognition whereby the capacity for some cognitive abilities implies the capacity for certain others (Fodor and Pylyshyn 1988). In schematic terms, systematicity is something's having cognitive capacity c_1 if and only if it has cognitive capacity c_2 (McLaughlin 2009). An often-used example is one's having the capacity to infer that John is the lover from *John loves Mary* if and only if one has the capacity to infer that Mary is the lover from *Mary loves John*.

What makes systematicity interesting is that not all models of cognition possess it, and so not all theories (particularly, those theories deriving such models) explain it. An elementary theory of mind, atomism, is a case in point: on this theory, the possession of each cognitive capacity (e.g., the inferring of John as the lover from *John loves Mary*) is independent of the possession of every other cognitive capacity (e.g., the inferring of Mary as the lover from *Mary loves John*), which admits instances of having one capacity without the other. Contrary to the atomistic theory, you don't find (English-speaking) people who can infer John as the lover (regarding the above example) without being able to infer Mary as the lover

(Fodor and Pylyshyn 1988). Thus, an atomistic theory does not explain systematicity.

An atomistic theory can be augmented with additional assumptions so that the possession of one capacity is linked to the possession of another. However, the problem with invoking such assumptions is that any pair of capacities can be associated in this way, including clearly unrelated capacities such as being able to infer John as the lover and being able to compute 27 as the cube of 3. Contrary to the augmented atomistic theory, there are language-capable people who do not understand such aspects of number. In the absence of principles that determine which atomic capacities are connected, such assumptions are ad hoc—"free parameters," whose sole justification is to take up the explanatory slack (Aizawa 2003).

Compare this theory of cognitive capacity with a theory of molecules consisting of atoms (core assumptions) and free parameters (auxiliary assumptions) for arbitrarily combining atoms into molecules. Such auxiliary assumptions are ad hoc, because they are sufficiently flexible to account for any possible combination of atoms (as a data-fitting exercise) without explaining why some combinations of atoms are never observed (see Aizawa 2003 for a detailed analysis).

To explain systematicity, a theory of cognitive architecture requires a (small) coherent collection of assumptions and principles that determine only those capacities that are systematically related and no others. The absence of such a collection, as an alternative to the classical theory (described below), has been the primary reason for rejecting connectionism as a theory of cognitive architecture (Fodor and Pylyshyn 1988; Fodor and McLaughlin 1990).

The classical explanation for systematicity posits a cognitive architecture founded upon a combinatorial syntax and semantics. Informally, the common structure underlying a collection of systematically related cognitive capacities is mirrored by the common syntactic structure underlying the corresponding collection of cognitive processes. The common semantic structure between the John and Mary examples (above) is the *loves* relation. Correspondingly, the common syntactic structure involves a process for tokening symbols for the constituents whenever the complex host is tokened. For example, in the *John loves Mary* collection of systematically related capacities, a common syntactic process may be $P \to$ *Agent loves Patient*, where *Agent* and *Patient* subsequently expand to *John* and *Mary*. Here, tokening refers to instantiating both terminal (no further processing) and nonterminal (further processing) symbols. The tokening principle seems to support a much needed account of systematicity, because all

capacities involve one and the same process; thus, having one capacity implies having the other, assuming basic capacities for representing constituents *John* and *Mary*.

Connectionists, too, can avail themselves of an analogous principle. In neural network terms, computational resources can be distributed between task-specific and task-general network components (e.g., weighted connections and activation units) by a priori specification and/or learning as a form of parameter optimization. For instance, an intermediate layer of weighted connections can be used to represent common components of a series of structurally related tasks instances, and the outer connections (the input–output interface) provide the task-specific components, so that the capacity for some cognitive function transfers to some other related function, even across completely different stimuli (see, e.g., Hinton 1990). Feedforward (Rumelhart, Hinton, and Williams 1986), simple recurrent (Elman 1990), and many other types of neural network models embody a generalization principle (see, e.g., Wilson, Marcus, and Halford 2001). In connectionist terms, acquiring a capacity (from training examples) transfers to other capacities (for testing examples).

Beyond the question of whether such demonstrations of systematicity, recast as generalization (Hadley 1994), correspond to the systematicity of humans (Marcus 1998; Phillips 1998), there remains the question of articulating the principle from which systematicity (as a kind of generalization) is a necessary, not just possible consequence. To paraphrase Fodor and Pylyshyn (1988), it is not sufficient to simply show existence—that there exists a suitably configured model realizing the requisite capacities; one also requires uniqueness—that there are no other models not realizing the systematicity property. For if there are other such configurations, then further (ad hoc) assumptions are required to exclude them. Existence/uniqueness is a recurring theme in our explanation of systematicity.

Note that learning, generally, is not a principle that one can appeal to as an explanation of systematicity. Learning can afford the acquisition of many sorts of input–output relationships, but only some of these correspond to the required systematic capacity relationships. For sure, one can construct a suitable set of training examples from which a network acquires one capacity if and only if it acquires another. But, in general, this principle begs the question of the necessity of that particular set of training examples. Connectionists have attempted to ameliorate this problem by showing how a network attains some level of generalization for a variety of training sets. However, such attempts are far from characteristic of what humans actually get exposed to.

Some authors have claimed to offer "alternative" nonclassical compositionality methods to meet this challenge whereby complex entities are tokened without tokening their constituents. The tensor product network formalism (Smolensky 1990) is one (connectionist) example. Another (nonconnectionist) example is Gödel numbering (van Gelder 1990). However, it's unclear what is gained by this notion of nontokened compositionality (see Fodor and McLaughlin 1990). For example, any set of localized orthonormal vectors (a prima facie example of classical tokening) can be made nonlocal (i.e., nonclassical, in the above sense) with a change of basis vectors. Distributed representations are seen as more robust against degradation than local representations—the loss of a single unit does not result in the loss of an entire capacity. In any case, the local–distributive dimension is orthogonal to the classical–nonclassical tokening dimension—a classical system can also be implemented in a distributed manner simply by replicating representational resources.

Classicists explicitly distinguish between their symbols and their implementation via a *physical instantiation function* (Fodor and Pylyshyn 1988, n. 9). Though much has been made of the implementation issue, this distinction does not make a difference in providing a complete explanation for systematicity, as we shall see. Nonetheless, the implementation issue is important, because any explanation that reduces to a classical one suffers the same limitations. We mention it because we also need to show that our explanation (presented shortly) is not classical—nor connectionist (nor Bayesian, nor dynamicist), for that matter.

The twist in this tale of two theories is that classicism does not provide a complete explanation for systematicity, either, though classicism arguably fares better than connectionism (Aizawa 2003). That the classical explanation also falls short seems paradoxical. After all, the strength of symbol systems is that a small set of basic syntax-sensitive processes can be recombined in a semantically consistent manner to afford all sorts of systematically related computational capacities. Combinatorial efficacy notwithstanding, what the classical theory fails to address is the many-to-many relationship between syntax and computational capacity: more than one syntactic structure gives rise to closely related though not necessarily identical groups of capacities. In these situations, the principle of syntactic compositionality is not sufficient to explain systematicity, because the theory leaves open the (common) possibility of constructing classical cognitive models possessing some but not all members of a collection of systematically related cognitive capacities (for examples, see Aizawa 2003; Phillips and Wilson 2010, 2011, 2012). For instance, if we replace the production $P \rightarrow$ *Agent loves Patient* with productions $P_1 \rightarrow$ *Agent loves John* and $P_2 \rightarrow$ *John loves Patient*, then this alternative classical system no longer generates the instance *Mary loves Mary*. The essential challenge for classicism echoes that for connectionism: explain systematicity without excluding models admitted by the theory just because they don't support all systematically related capacities—why are those models not realized (cf.: why don't some combinations of atoms form molecules)?

So far, none of the major theoretical frameworks in cognitive science—classicist, connectionist, Bayesian, nor dynamicist—has provided a theory that fully explains systematicity. This state of affairs places cognitive science in a precarious position—akin to physics without a theory of atomic structure: without a stable foundation on which to build a theory of cognitive representation and process, how can one hope to scale the heights of mathematical reasoning? In retrospect, the lack of progress on the systematicity problem has been because cognitive scientists were working with "models" of structure (i.e., particular concrete implementations), where systematicity is a *possible* consequence, rather than "theories" of structure from which systematicity *necessarily* follows. This diagnosis led us (Phillips and Wilson 2010, 2011, 2012) to *category theory* (Eilenberg and Mac Lane 1945; Mac Lane 2000), a theory of structure par excellence, as an alternative approach to explaining systematicity.

The rest of this chapter aims to be, as much as possible, an informal, intuitive discussion of our category theory explanation as a complement to the formal, technical details already provided (Phillips and Wilson 2010, 2011, 2012). To help ground the informal discussion, though, we also include some standard formal definitions (see Mac Lane 2000 for more details) as stand-alone text, and associated *commutative diagrams*, in which entities (often functions) indicated by paths with the same start point and the same end point are equal. An oft-cited characteristic feature of category theory is the focus on the directed relationships between entities (called *arrows*, *morphisms*, or *maps*) instead of the entities themselves—in fact, categories can be defined in arrow-only terms (Mac Lane 2000). This change in perspective is what gives category theory its great generality. Yet, category theory is not arbitrary—category theory constructs come with formally precise conditions (axioms) that must be satisfied for one to avail oneself of their computational properties. This unique combination of abstraction and precision is what gives category theory its great power. However, the cost of taking a category theory perspective is that it may not be obvious how category theory should be applied to the problem at hand, nor what benefits are afforded when doing so. Hence, our purpose in this chapter is threefold: (1) to provide an intuitive understanding of our category theory explanation for systematicity; (2) to show how it differs from other approaches; and (3) to discuss the implications of this explanation for the broader interests of cognitive science.

2 What Is Category Theory?

Category theory was invented in the mid-1940s (Eilenberg and Mac Lane 1945) as a formal means for comparing mathematical structures. Originally it was regarded as a formal language for making precise the way in which two types of structures are to be compared. Subsequent technical development throughout the twentieth century has seen it become a branch of mathematics in its own right, as well as placing it on a par with set theory as a foundation for mathematics (see Marquis 2009 for a history and philosophy of category theory). Major areas of application, outside of mathematics, have been computer science (see, e.g., Arbib and Manes 1975; Barr and Wells 1990) and theoretical physics (see, e.g., Baez and Stay 2011; Coecke 2006). Category theory has also been used as a general conceptual tool for describing biological (Rosen 1958) and neural/cognitive systems (Ehresmann and Vanbremeersch 2007), yet applications in these fields are relatively less extensive.

Category theory can be different things in different contexts. In the abstract, a category is just a collection of *objects* (often labeled A, B, ...), a collection of *arrows* (often labeled f, g, ...) between pairs of objects (e.g., $f : A \to B$, where A is called the *domain* and B the *codomain* of arrow f), and a *composition operator* (denoted $\circ$) for composing pairs of arrows into new arrows (e.g., $f \circ g = h$), all in a way that satisfies certain basic rules (axioms). When the arrows are functions between sets, the composition is ordinary composition of functions, so that $(f \circ g)(x) = f(g(x))$. To be a category, every object in the collection must have an *identity arrow* (often denoted as $1_A : A \to A$); every arrow must have a domain and a codomain in the collection of objects; for every pair of arrows with matching codomain and domain objects there must be a third arrow that is their composition (i.e., if $f : A \to B$ and $g : B \to C$, then $g \circ f : A \to C$ must also be in the collection of arrows); and composition must satisfy *associativity*, that is, $(h \circ g) \circ f = h \circ (g \circ f)$, and *identity*, that is, $1_B \circ f = f = f \circ 1_A$, laws for all arrows in the collection. Sets as objects and functions as arrows satisfy all of this: the resulting category of all ("small") sets is usually called **Set**. In this regard, category theory could be seen as an algebra of arrows (Awodey 2006).

For a formal (abstract) category, the objects, arrows, and composition operator need no further specification. A simple example is a category whose collection of objects is the set $\{A,B\}$ and collection of arrows is the set $\{1_A : A \to A, 1_B : B \to B, f : A \to B, g : B \to A\}$. Since there are no other arrows in this category, compositions $f \circ g = 1_A$ and $g \circ f = 1_B$ necessarily hold. Perhaps surprisingly, many important results pertain to this level and hence apply to anything that satisfies the axioms of a category.

For particular examples of categories, some additional information is provided regarding the specific nature of the objects, arrows, and composition. Many familiar structures in mathematics are instances of categories. For example, a partially ordered set, also called a *poset*, $(P,\leq)$ is a category whose objects are the elements of the set P, and arrows are the order relationships $a \leq b$, where a, $b \in P$. A poset is straightforwardly a category, since a partial order $\leq$ is *reflexive* (i.e., $a \leq a$, hence identities) and *transitive* (i.e., $a \leq b$ and $b \leq c$ implies $a \leq c$, hence composition is defined). Checking that identity and associativity laws hold is also straightforward. The objects in a poset considered as a category have no internal parts. In other categories, the objects may also have internal structure, in which case the arrows typically preserve that structure. For instance, the category **Pos** has posets now considered as objects and order-preserving functions for arrows, i.e., $a \leq b$ implies $f(a) \leq f(b)$. For historical reasons, the arrows in a category may also be called *morphisms*, *homomorphisms*, or *maps* or functions when specifically involving sets.

Definition (Category). A *category* **C** consists of a class of objects $|\mathbf{C}|=\{A, B, \ldots\}$; and for each pair of object A, B in **C**, a set $\mathbf{C}(A,B)$ of morphisms (also called arrows, or maps) from A to B where each morphism $f : A \to B$ has A as its *domain* and B as its *codomain*, including the *identity* morphism $1_A : A \to A$ for each object A; and a composition operation, denoted $\circ$, of morphisms $f : A \to B$ and $g : B \to C$, written $g \circ f : A \to C$ that satisfies the laws of:

- *identity*, where $f \circ 1_A = f = 1_B \circ f$, for all $f : A \to B$; and
- *associativity*, where $(h \circ g) \circ f = h \circ (g \circ f)$, for all $f : A \to B$, $g : B \to C$ and $h : C \to D$.

In the context of computation, a category may be a collection of types for objects and functions (sending values of one type to values of another, or possibly the same, type) for arrows, where composition is just composition of functions. In general, however, objects need not be sets, and arrows need not be functions, as shown by the first poset-as-a-category example. For our purposes, though, it will often be helpful to think of objects and arrows as sets and functions between sets. Hence, for cognitive applications, one can think of a category as modeling some cognitive (sub)system, where an object is a set of cognitive states and an arrow is a cognitive process mapping cognitive states.

2.1 "Natural" Transformations, Universal Constructions

To a significant extent, the motivations of category theorists and cognitive scientists overlap: both groups aim to establish the principles underlying particular structural relations, be they mathematical structures or cognitive structures. In this regard, one of the central concepts is a natural transformation between structures. Category theorists have provided a formal definition of "natural," which we use here. This definition builds on the concepts of *functor* and, in turn, category. We have already introduced the concept of a category. Next we introduce functors before introducing natural transformations and universal constructions.

Functors are to categories as arrows (morphisms) are to objects. Arrows often preserve internal object structure; functors preserve category structure (i.e., identities and compositions). Functors have an object-mapping component and an arrow-mapping component.

Definition (Functor). A *functor* $F:\mathbf{C}\to\mathbf{D}$ is a map from a category **C** to a category **D** that sends each object A in **C** to an object $F(A)$ in **D**; and each morphism $f:A\to B$ in **C** to a morphism $F(f):F(A)\to F(B)$ in **D**, and is structure-preserving in that $F(1_A)=1_{F(A)}$ for each object A in **C**, and $F(g\circ_{\mathbf{C}} f)=F(g)\circ_{\mathbf{D}} F(f)$ for all morphisms $f:A\to B$ and $g:B\to C$, where $\circ_{\mathbf{C}}$ and $\circ_{\mathbf{D}}$ are the composition operators in **C** and **D**.

There is an intuitive sense in which some constructions are more natural than others. This distinction is also important for an explanation of systematicity, as we shall see. The concept of a *natural transformation* makes this intuition formally precise.

Natural transformations are to functors as functors are to categories. We have already seen that functors relate categories, and similarly, natural transformations relate functors. Informally, what distinguishes a natural transformation from some arbitrary transformation is that a natural transformation does not depend on the nature of each object A. This independence is also important for systematicity: basically, the cognitive system does not need to know ahead of time all possible instances of a particular transformation.

$$\begin{tikzcd} F(A) \arrow[r, "\eta_A"] \arrow[d, "F(f)"'] & G(A) \arrow[d, "G(f)"] \\ F(B) \arrow[r, "\eta_B"'] & G(B) \end{tikzcd}$$

Figure 9.1
Commutative diagram for natural transformation.

$$\begin{tikzcd} X \arrow[r, "\psi"] \arrow[dr, "f"'] & F(B) \arrow[d, dashed, "F(k)"] & B \arrow[d, dashed, "k"'] \\ & F(Y) & Y \end{tikzcd}$$

Figure 9.2
Commutative diagram for universal construction.

Definition (Natural transformation). A *natural transformation* $\eta:F\to G$ between a domain functor $F:\mathbf{C}\to\mathbf{D}$ and a codomain functor $G:\mathbf{C}\to\mathbf{D}$ consists of **D**-maps $\eta_A:F(A)\to G(A)$ for each object A in **C** such that $G(f)\circ\eta_A=\eta_B\circ F(f)$. (See figure 9.1.)

For the purpose of explaining systematicity, we need something more than just that constructions are natural in this sense; constructions are also required to be *universal*, in a technical sense to be introduced next. A universal construction is basically an arrow that is "part" of every arrow in the category that models the (cognitive) domain of interest.

Definition (Universal construction). Given an object $X\in|\mathbf{C}|$ and a functor $F:\mathbf{B}\to\mathbf{C}$, a *couniversal morphism* from X to F is a pair (B,Ψ) where B is an object of **B**, and Ψ is a morphism in **C**, such that for every object $Y\in|\mathbf{B}|$ and every morphism $f:X\to F(Y)$, there exists a unique morphism $k:B\to Y$ such that $F(k)\circ\psi=f$. (See figure 9.2.)

A universal construction is either a couniversal morphism, or (its dual) a universal morphism (whose definition is obtained by reversing all the arrows in the definition of "couniversal").

At first it may not seem obvious how universal constructions are related to natural transformations. Note that, given a category of interest **D** and an object X in **D**, X corresponds to a constant functor $X:\mathbf{C}\to\mathbf{D}$ from an arbitrary category **C** to the category of interest **D**, where functor X sends every object and arrow in C to the object X and identity 1_X in **D**, thus yielding a natural transformation $\eta : X \to F$.

3 Systematicity: A Category Theory Explanation

All major frameworks assume some form of compositionality as the basis of their explanation for systematicity. In the classical case, it's syntactic; for the connectionist, it's functional (as we have already noted in section 1). In both cases, systematic capacity is achieved by combining basic processes. However, the essential problem is that there is no additional constraint to circumscribe only the relevant combinations. Some combinations are possible that do not support all members of a specific collection of systematically related capacities. So, beyond simply stipulating the acceptable models (i.e., those consistent with the systematicity property), additional principles are needed to further constrain the admissible models.

Our category theory explanation also relies on a form of compositionality, but not just any form. The additional ingredient in our explanation is the formal category theory notion of a universal construction. The essential point of a universal construction is that each and every member of a collection of systematically related cognitive capacities is modeled as a morphism in a category that incorporates a common universal morphism in a unique way. From figure 9.2, having one capacity $f_1 : A \to F(Y_1)$ implies having the common universal morphism $\psi : X \to F(B)$, since $f_1 = F(k_1) \circ \psi$. And, since the capacity-specific components $F(k_i)$ are uniquely given (constructed) by functor $F : \mathbf{B} \to \mathbf{C}$, and the arrows k_i in **B**, one also has capacity $f_2 : A \to F(Y_2)$, since $f_2 = F(k_2) \circ \psi$. That is, all capacities (in the domain of interest) are systematically related via the couniversal morphism ψ. Our general claim is that each collection of systematically related capacities is an instance of a universal construction. The precise nature of this universal morphism will depend (of course) on the nature of the collection of systematically related capacities in question. In this section, we illustrate several important examples.

3.1 Relations: (Fibered) Products

We return to the *John loves Mary* example to illustrate our category theory explanation. This and other instances of relational systematicity are captured by a categorical product (Phillips and Wilson 2010). A categorical product provides a universal means for composing two objects (A and B) as a third object (P) together with the two arrows (p_1 and p_2) for retrieving information pertaining to A and B from their composition P. The require-

$$
\begin{tikzcd}
 & & A \\
Z \arrow[urr, bend left, "z_1"] \arrow[drr, bend right, "z_2"'] \arrow[r, dashed, "\langle z_1,z_2\rangle"] & P \arrow[ur, "p_1"] \arrow[dr, "p_2"'] & \\
 & & B
\end{tikzcd}
$$

Figure 9.3
Commutative diagram for product.

ment that such a construction be universal is critical for explaining the systematicity property.

Definition (Product). A *product* of objects A and B in category C is an object P (also denoted $A{\times}B$) together with two morphisms (sometimes called *projections*) $p_1 : P \to A$ and $p_2 : P \to B$, jointly expressed as (P, p_1, p_2) such that for every object $Z \in |\mathrm{C}|$ and pair of morphisms $z_1 : Z \to A$ and $z_2 : Z \to B$ there exists a unique morphism $u : Z \to P$, also denoted $\langle z_1, z_2 \rangle$, such that $z_1 = p_1 \circ u$ and $z_2 = p_2 \circ u$. (See figure 9.3).

Suppose A and B correspond to the set of representations for the possible agents and patients that can partake in the *loves* relation, which includes instances such as *John loves Mary* and *Mary loves John*, and the product object P, which is the Cartesian product $A{\times}B$ (in the case of products of sets, as in this example), corresponds to the representations of those relational instances. Then any requirement to extract components A (B) from some input Z necessarily factors through $p_1(p_2)$ uniquely.

As a universal construction, products are constructed from the *product functor*. Informally, the product functor sends pairs of objects (A, B) to the product object $A{\times}B$ and pairs of arrows (f, g) to the product arrow $f{\times}g$. In this way, all possible combinations must be realized. The universality requirement rules out partial constructions, such as a triple (Q, q_1, q_2) where the object Q contains just three of the four possible pair combinations of *John* and *Mary* (and q_1 and q_2 return the first and second item of each pair), because this triple does not make the associated diagram (see figure 9.3) commute. Thus, no further assumptions are needed to exclude such cases, in contrast with the classical (or connectionist) explanation, which admits such possibilities.

One can think of the product functor as a way of constructing new wholes (i.e., $A{\times}B$) from parts (A, B). The product functor is seen as the "conceptual inverse" to the *diagonal functor* which makes wholes into parts by making copies of each object and morphism: that is, object A and morphism f are sent to pairs (A, A) and (f, f). Together, diagonal and product functors constitute an *adjunction*, a family of universal constructions (Mac Lane 2000). Adjunctions are important to our explanation of systematicity because they link representation and access (inference) without relying on the assumption that such processes be compatible. A classical explanation simply assumes compatibility between syntactically compositional representations and the processes that operate on them. Yet, as we've seen, there is more than one way of syntactic composition, and not all of them support systematicity. By contrast, in our category theory explanation, the commutativity property associated with an adjunction enforces compatibility (Phillips and Wilson 2010, 2011).

We further contrast our explanation with a proposed alternative illustrated by Gödel numbering (van Gelder 1990). This scheme depends on careful construction of a suitable transformation function that depends on the values of all possible elements (past, present, and future) that can partake in the relation. However, in general, a cognitive system cannot have knowledge of such things. At best, a cognitive system can update a set of representations to accommodate new instances to maintain correct transformation. But such allowances admit nonsystematicity: at the point prior to update, the cognizer is in a state of having some but not all systematically related capacities. Thus, such schemes do not account for systematicity.

Products address systematic capacity where there is no interaction between constituents A and B. Another case addresses quasi systematicity, where capacity extends to some but not all possible combinations of constituents. In this situation, the interaction between A and B is given by two arrows $f : A \to C$ and $g : B \to C$ to a common (constraint) object C. The universal construction for this situation is called a *pullback* (or *fibered product*, or *constrained product*). The explanation for systematicity in this case essentially parallels the one given for products: replace product with pullback (see Phillips and Wilson 2011).

3.2 Recursion: F-(co)algebras

Our explanation for systematicity with regard to recursive domains also employs a universal construction, albeit with a different kind of functor, called an *endofunctor*, where the domain and codomain are the same category (Phillips and Wilson 2012), hence its importance for recursion. A motivating example is that the capacity to find the smallest item in a list implies the capacity to find the largest item, assuming a basic capacity for distinguishing the relative sizes of items. For example, you don't find people who can select the lowest card from a deck without being able to select the highest card, assuming they understand the relative values of cards. Yet, classical theory admits recursive and nonrecursive compositional methods, for realizing these two capacities, without there being any common component processes. If one capacity is realized by a recursive method and the other by a nonrecursive method, then the two capacities are not intrinsically connected in any way—the tokening principle on which classical theory depends is no longer in play. Thus, classical theory also fails to fully explain systematicity with regard to recursively definable capacities (Phillips and Wilson 2012).

In recent decades, computer scientists have turned to category theory to develop a systematic treatment of recursive computation (see, e.g., Bird and Moor 1997). We have adapted this theory for an explanation of systematicity in regard to recursive cognitive capacities. Conceptually, recursive capacities are decomposed into an invariant component, the recurrent part, and a variant component, with the capacity-specific computation taking place at each iterative step. The invariant component corresponds to the underlying recursive data structure (e.g., stepping through a deck of cards) underpinning the group of systematically related capacities. The variant component corresponds to the computation at each step (e.g., comparing cards for the smaller or larger card). In category theory terms, every recursive capacity is an algebra, called an *F-algebra*, built using an endofunctor F. Under very general conditions, a category of such F-algebras has a universal construction called an initial F-algebra, and hence provides an explanation for systematicity with regard to recursive capacities.

Definition (F-algebra, initial algebra, catamorphism). For an endofunctor $F : \mathbf{C} \to \mathbf{C}$, an *$F$-algebra* is a pair (A,α), where A is an object and $\alpha : F(A) \to A$ is a morphism in $\mathbf{C}$.

An *initial algebra* (A, in) is an *initial object* in the category of F-algebras $\mathbf{Alg}(F)$. That is, $in : F(A) \to A$ is a morphism in $\mathbf{C}$, and there exists a unique F-algebra homomorphism from (A, in) to every F-algebra in $\mathbf{Alg}(F)$.

A *catamorphism* $h : (A, in) \to (B, \beta)$ is the *unique* F-algebra homomorphism from initial F-algebra (A, in) to F-algebra $(B,)$. That is, $h \circ in = \beta \circ F(h)$ and the uniquely specified h for each such is denoted *cata* (i.e., h = *cata*). (See figure 9.4.)

The dual constructions: *F-coalgebra*, *final coalgebra*, and *anamorphism* are also used to explain related instances of systematicity (see Phillips and Wilson 2012 for details).

$$
\begin{tikzcd}
F(A) \arrow[r, "in"] \arrow[d, "F(\mathrm{cata}\ \beta)"'] & A \arrow[d, "\mathrm{cata}\ \beta"] \\
F(B) \arrow[r, "\beta"'] & B
\end{tikzcd}
$$

Figure 9.4
Commutative diagram for catamorphism.

In each case, the recursive capacity depends only on the common arrow $in : F(A) \to A$ and the unique arrow $\mathrm{cata}\beta : A \to B$ (see figure 9.4). In outline, searching for the smallest number in a list of numbers is given by *fold*(∞, *lower*) *l*, where *fold* is the common recursive part, (∞, *lower*) is the task-specific component, ∞ applies to empty lists, *lower* returns the lower of two numbers, and *l* is a list of numbers. For example, *fold*(0, *lower*)[3,2,5] = *lower*(3, *lower*(2, *lower*(5, ∞))) = 2. Searching for the largest number in a list is given by *fold*(0, *higher*) *l*, where (0, *higher*) is the task-specific component, 0 applies to empty lists, and *higher* returns the higher of two numbers. For example, *fold*(0, *higher*)[3,2,5] = *higher*(3, *higher*(2, *higher*(5, 0))) = 5.

4 Category Theory in Context

The abstract and abstruse nature of category theory may make it difficult to see how our explanation relates to other theoretical approaches, and how it should make contact with a neural level of analysis. For all of the theoretical elegance of category theory, constructs must also be realizable by the underlying brain system. The relationship between our category theory explanation and a classical or connectionist one is analogous to the relationship between, say, an abstract and concrete specification of a group in mathematics: particular classical or connectionist architectures may be models of our theory. Here, we sketch some possibilities.

4.1 Classical Models

Our category theory explanation overlaps with the classical one in the sense that the common constituent of a collection of complex cognitive capacities is "tokened" (i.e., imparted or executed) whenever each complex cognitive capacity is. Notice, however, that tokening in the category theory sense is the tokening of arrows, not objects; analogously, tokening is of processes, not symbols. Classical systems also admit symbols as processes.

Nonetheless, our explanation goes significantly beyond the classical one in that we require not just any arrow, but rather, an arrow derived from a universal construction. Thus, no further assumptions are required to guarantee that each and every capacity is uniquely constructed from it. Moreover, the virtue of this arrow-centric perspective is that, contra classicism, our explanation extends to nonsymbolic domains, such as visual cognition, without further adjustment to the theory.

Universal constructions such as adjunctions and *F*-(co)algebras as the basis for a theory of cognitive architecture are unique to our theory, and go significantly beyond the widespread use of isomorphism (cf. analogy models, including Suppes and Zinnes 1963; Halford and Wilson 1980) in cognitive science generally. From a category theory perspective, two systems that are isomorphic are essentially the "same" up to a change of object and arrow labels. An adjunction is more general and potentially more useful: two systems (involving different sorts of processes) in an adjoint relationship need not be isomorphic, while still being in a systematic relationship with each other.

4.2 Nonclassical Models

For conciseness, we treat connectionist and Bayesian models in the same light, despite some significant advances in Bayesian modeling (see, e.g., Kemp and Tenenbaum 2008). By regarding connectionist and Bayesian networks as graphs (with additional structure), one could consider a category of such graphs as objects and their homomorphisms as morphisms. That additional structure could include "coloring" graph nodes to distinguish corresponding input, output and internal network activation units, and functions corresponding to propagation of network activity. The category **Grph** of graphs and graph homomorphisms has products, suggesting that a suitable category of graphs with additional structure can be devised that also has products. In these cases, systematicity is realized as a functor from the product (as a category) into a category of connectionist/Bayesian networks, thus guaranteeing an implementation of the systematicity property within a connectionist/Bayesian-style framework.

A similar approach also applies to dynamic systems models. In a simple (though not exhaustive) case, one class of dynamic systems can be treated as a finite state machine. A category of such machines and the structure-preserving morphisms also has products (see Arbib and Manes 1975). Thus, again, systematicity is implemented as a functor from the product (as a category) into this category of finite state machines.

5 Testing the Theory

Our theory can be tested using a series of tasks, where each task instance is composed of an invariant component corresponding to the (co)universal arrow and a variant component corresponding to the unique arrow of the underlying universal construction (see Phillips and Wilson 2012, text S2). An example of this kind of design is familiar in the form of so-called *learning set* paradigms (see, e.g., Kendler 1995).

5.1 Nonrecursive Example

A series of simple classification tasks illustrates one kind of experimental design that can be used to test for systematicity in terms of universal constructions. For the first task, suppose participants are given stimuli to be classified into one of two classes. Let S be the set of stimuli, and R the set of (two) responses associated with each class. Hence, the morphism $t : S \to R$ is the stimulus-response process for the first task instance. Next, suppose the task is modified (for the next task instance), say by changing the responses to each stimulus class from left and right to up and down. Let R' be the set of responses associated with the new task, and $t' : S \to R'$ the associated stimulus-response process. Since the responses are determined by the classes rather than directly by the stimuli, each task instance t decomposes into the task-invariant classification component $c : S \to C$, where C is the set of classes, and the task-variant response mapping component $r : C \to R$ uniquely. That is, c corresponds to Ψ, the couniversal arrow, in figure 9.2 for a universal construction.

A test of systematicity for this example is whether participants can correctly predict the stimulus response classification on new task trials after receiving sufficient trials to determine r, the response mapping. In the general case that there are n possible responses, so n trials are needed to determine the correct mapping (one trial for each possible response), but no more. Thus, systematicity is evident on correct prediction for the remaining $m - n$ stimulus-response trials, assuming there are more stimuli than responses (i.e., $m > n$). (See Phillips and Wilson 2010 for a further example.)

5.2 Recursive Example

Phillips and Wilson (2012) provide an example of systematicity with respect to recursively definable concepts in the form of finding the smallest/largest item in a list. Here, we illustrate how this kind of systematicity can be directly tested. Suppose participants are given pairs of stimuli (e.g., shapes) from which they must predict the "preferred" (or rewarded) shape—essentially, a discrimination task. This preliminary task allows participants to learn the total order associated with the set of stimuli. Upon completion of the discrimination task, participants are presented with a list of stimuli, selected from the set used in the preliminary task, and asked to find the most preferred stimulus in that list. Upon completion of this second task, participants are then required to identify the unpreferred stimulus from a pair of stimuli, and then the least preferred stimulus in a list. Evidence of systematicity in this example is correct determination of the least preferred stimulus without further feedback. This paradigm could be further extended by changing the set of stimuli between task instances. (See Phillips and Wilson 2012 for another example.)

6 Beyond Systematicity

Beyond systematicity are other questions that a general (category) theory of cognitive architecture should address. We round out this chapter by considering how a categorial theory of cognitive architecture may address such issues.

6.1 Systematicity and Nonsystematicity: Integration

Not all cognitive capacities are systematic, as we mentioned in the first section. The classical proposal (Fodor and Pylyshyn 1988), and so far ours too, speak only to the systematic aspects of cognition, while leaving nonsystematic aspects unaddressed. Fodor and Pylyshyn (1988) recognized the possibility of some kind of hybrid theory: say, a classical architecture fused with some nonclassical (e.g., connectionist) architecture to address cognitive properties beyond the scope of (or unaccounted for by) classical theory. Aizawa (2003) warns, however, that hybrid theories require a higher explanatory standard: not only must each component theory account for their respective phenomena, but there must also be a principled account for why and when each component theory is invoked. Here, we sketch how our category theory approach could be extended to incorporate nonsystematic aspects of cognition.

As Phillips and Wilson (2012) suggest, if we regard a category as a model of a cognitive subsystem, then combining two categories by taking a fibered (co)product can be regarded as the integration of two subsystems into a larger combined system. Clark, Coecke, and Sadrzadeh (2008) provide an example of how subsystems can be combined categorically for modeling aspects of language. Their example is a hybrid symbolic-distributional model of grammar as the fibered product of a symbolic and a distributed (vector-based) component. In our case, one category realizes systematicity, another realizes nonsystematicity, and the category derived from their fibered (co)product realizes both.

For our purposes, though, we must also consider the principle dictating which component is to be employed and under what circumstances. We have suggested (Phillips and Wilson 2012) that a cost–benefit trade-off may be the basis of such a principle. For instance, there are at least two ways to add numbers such as 3 and 5. One can employ a systematic counting procedure by counting from the first number (3) the number of increments indicated by the second number (5). This procedure has the benefit of working for any two numbers (systematicity), but at the cost of being slow when the numbers are large. Alternatively, one can simply recall from memory the sum of the given numbers. This second procedure has the benefit of speed, but the cost of unreliability (unsystematic): the sums of some pairs may not have been memorized, and moreover, time and effort are required to memorize each pair.

To accommodate such possibilities, we further suggest here that our category theory approach can be extended by associating a cost with each morphism within the framework of *enriched category theory* (Kelly 2005). Enriched category theory considers categories whose hom-sets (i.e., sets of arrows between pairs of objects) have additional structure. For example, by defining a partial-order over a set of arrows each hom-set becomes a poset (i.e., a set with the extra order structure)—the category is enriched over the category of posets, **Pos**, the category of partially ordered sets and order-preserving functions. In this way, a choice between arrows (alternative cognitive strategies) can be based on an order principle—choose the strategy with the lower associated cost, when the alternatives are comparable.

Note, however, that we are not yet in a position to provide such principles. One could, of course, simply fit data by assigning adjustable parameters to each morphism, akin to a connectionist network. However, this maneuver merely affords compatibility with the data. What we really require is a principle necessitating when a particular subsystem is employed, lest we also succumb to the kinds of ad hoc assumptions that have bedeviled other approaches to the systematicity problem (Aizawa 2003).

6.2 Category Theory, Systematicity, and the Brain

Any theory of cognitive architecture must ultimately be reconcilable with the underlying neural architecture. Ehresmann and Vanbremeersch (2007)

have provided a general description of how biological, neural, and cognitive systems may be cast within a category theory framework, though their work was not intended to address the systematicity problem.

Our category theory approach to the systematicity problem suggests an intriguing connection between the implied components of a categorial cognitive architecture (universal constructions) and brain structure. Note that the universal constructions we have employed to address various instances of systematicity involve endofunctors. The composition of adjoint functors is necessarily an endofunctor, and F-(co)algebras are based on endofunctors. An analogue of recurrency in the brain is the reciprocating neural connections within and between brain regions. Thus, one place to look for a correspondence between cognitive and neural architectures, at least in regard to the systematic aspects of cognition, are recurrent neurally connected brain regions. These kinds of connections are prevalent throughout the brain. Conversely, brain regions lacking such connections suggest corresponding cognitive capacities lacking systematicity. Of course, reciprocal neural connections may have other functional roles, and the computational connection to adjunctions is only speculation at this point.

We have begun investigating the relationship between category theory constructs and the brain (Phillips, Takeda, and Singh 2012). A pullback (fibered product), which featured as the kind of universal construction in our explanation of quasi systematicity (Phillips and Wilson 2011), also corresponds to integration of stimulus feature information in visual attention. By varying the "arity" (unary, binary, ternary) of the fibered product matching the number of feature dimensions (color, frequency, orientation) needed to identify a target object, we observed significantly greater EEG synchrony (phase-locking) between frontal and parietal electrodes with increasing arity. These results also provide a category theory window into development, discussed next.

6.3 Development and Learning

Cognitive development, in some cases, can also be seen as instances of systematicity. The capacity for inferential abilities, such as *transitive inference* and *class inclusion* are consistently enabled around the age of five years (see Halford, Wilson, and Phillips, 1998). Children who have the capacity to make transitive inferences typically also have the capacity to make inferences based on class inclusions. Conversely, children who fail at class inclusion also fail at transitive inference. Thus, we can see this equivalence as another instance of the systematicity schemata: capacity c_1 if and only if capacity c_2, all else being equal (McLaughlin 2009).

We have given a category theory explanation for these data (Phillips, Wilson, and Halford 2009) to overcome some difficulties with our earlier *relational complexity* approach (Halford et al. 1998). This common inferential capacity was explained in terms of the arity of the underlying (co) product. Older children (above age five years) have the capacity for binary (co)products, whereas younger children do not. Thus, as an instance of a universal construction, and a special case of our systematicity explanation, having the capacity for transitive inference implies class inclusion because the underlying categorical structures are dual to each other. In the dual case, the constructions are related by reversal of arrow directions via *contravariant* functors, where each object is mapped to itself, and each arrow to an arrow whose domain and codomain are respectively the codomain and domain of the source arrow.

Learning is also of central importance to cognitive science. A universal construction is a kind of optimal solution to a problem: a (co)universal morphism is an arrow that is a factor (in the sense of function composition) of all arrows to/from a particular construction (functor). To the extent that learning (and evolution) is a form of optimization, universal constructions may provide an alternative perspective on this aspect of cognition.

7 Conclusion

We began this chapter with the distribution of deflected charged particles affording an important insight into the structure of the atom as an analogy to the importance of the distribution of cognitive capacities to understanding the nature of cognitive architecture. We end this chapter with the insight that systematicity affords cognitive science: the atomic components of thought include universal constructions (not symbols, connections, probabilities, or dynamical equations, though these things may be part of an implementation) insofar as the systematicity property is evident.

Acknowledgment

This work was supported by a Japanese Society for the Promotion of Science (JSPS) Grant-in-Aid (Grant No. 22300092).

References

Aizawa, K. 2003. *The Systematicity Arguments*. New York: Kluwer Academic.

Arbib, M. A., and E. G. Manes. 1975. *Arrows, Structures, and Functors: The Categorical Imperative*. London: Academic Press.

Awodey, S. 2006. *Category Theory*. New York: Oxford University Press.

Baez, J. C., and M. Stay. 2011. Physics, topology, logic and computation: A Rosetta Stone. In *New Structures in Physics*, ed. B. Coecke, 95–172. New York: Springer.

Barr, M., and C. Wells. 1990. *Category Theory for Computing Science*. New York: Prentice Hall.

Bird, R., and O. de Moor. 1997. *Algebra of Programming*. Harlow: Prentice Hall.

Clark, S., B. Coecke, and M. Sadrzadeh. 2008. A compositional distributional model of meaning. In *Proceedings of the Second Symposium on Quantum Interaction*, ed. P. Bruza, W. Lawless, K. van Rijsbergen, D. Sofge, B. Coecke, and S. Clark, 133–140. Oxford: College Publications.

Coecke, B. 2006. Introducing categories to the practicing physicist. In *What Is Category Theory?* vol. 3, 45–74. Monza, Italy: Polimetrica.

Ehresmann, A. C., and J.-P. Vanbremeersch. 2007. *Memory Evolutive Systems: Hierarchy, Emergence, Cognition*, vol. 4. Oxford: Elsevier.

Eilenberg, S., and S. Mac Lane. 1945. General theory of natural equivalences. *Transactions of the American Mathematical Society* 58:231–294.

Elman, J. L. 1990. Finding structure in time. *Cognitive Science* 14:179–211.

Fodor, J. A., and B. P. McLaughlin. 1990. Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition* 35 (2):183–204.

Fodor, J. A., and Z. W. Pylyshyn. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition* 28 (1–2):3–71.

Hadley, R. F. 1994. Systematicity in connectionist language learning. *Mind and Language* 9 (3):247–272.

Halford, G. S., and W. H. Wilson. 1980. A category theory approach to cognitive development. *Cognitive Psychology* 12:356–411.

Halford, G. S., W. H. Wilson, and S. Phillips. 1998. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences* 21 (6):803–831.

Hinton, G. E. 1990. Mapping part-whole hierarchies in connectionist networks. *Artificial Intelligence* 46 (1–2):47–76.

Kelly, G. M. 2005. Basic concepts of enriched category theory. *Reprints in Theory and Applications of Categories 10*.

Kemp, C., and J. B. Tenenbaum. 2008. The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America* 105 (31): 10687–10692.

Kendler, T. S. 1995. *Levels of Cognitive Development*. Mahwah, NJ: Erlbaum.

Mac Lane, S. 2000. *Categories for the Working Mathematician*, 2nd ed. New York: Springer.

Marcus, G. F. 1998. Rethinking eliminative connectionism. *Cognitive Psychology* 37 (3):243–282.

Marquis, J.-P. 2009. *From a Geometrical Point of View: A Study of the History and Philosophy of Category Theory*. New York: Springer.

McLaughlin, B. P. 2009. Systematicity redux. *Synthese* 170:251–274.

Phillips, S. 1998. Are feedforward and recurrent networks systematic? Analysis and implications for a connectionist cognitive architecture. *Connection Science* 10 (2):137–160.

Phillips, S., Y. Takeda, and A. Singh. 2012. Visual feature integration indicated by phase-locked frontal-parietal EEG signals. *PLoS ONE* 7 (3):e32502.

Phillips, S., and W. H. Wilson. 2010. Categorial compositionality: A category theory explanation for the systematicity of human cognition. *PLoS Computational Biology* 6 (7):e1000858.

Phillips, S., and W. H. Wilson. 2011. Categorial compositionality II: Universal constructions and a general theory of (quasi-)systematicity in human cognition. *PLoS Computational Biology* 7 (8):e1002102.

Phillips, S., and W. H. Wilson. 2012. Categorial compositionality III: F-(co)algebras and the systematicity of recursive capacities in human cognition. *PLoS ONE* 7 (4):e35028.

Phillips, S., W. H. Wilson, and G. S. Halford. 2009. What do Transitive Inference and Class Inclusion have in common? Categorical (co)products and cognitive development. *PLoS Computational Biology* 5 (12):e1000599.

Rosen, R. 1958. The representation of biological systems from the standpoint of the theory of categories. *Bulletin of Mathematical Biophysics* 20:317–341.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning representations by back-propagation of error. *Nature* 323:533–536.

Smolensky, P. 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46 (1–2):159–216.

Suppes, P., and J. L. Zinnes. 1963. Basic measurement theory. In *Handbook of Mathematical Psychology*, ed. R. D. Luce, R. R. Bush, and E. Galanter. London: Wiley.

van Gelder, T. 1990. Compositionality: A connectionist variation on a classical theme. *Cognitive Science* 14:355–384.

Wilson, W. H., N. Marcus, and G. S. Halford. 2001. Access to relational knowledge: a comparison of two models. In *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, ed. J. D. Moore and K. Stenning, 1142–1147. Austin, TX: Cognitive Science Society.