A category theory explanation for systematicity

Steven Phillips (steve@ni.aist.go.jp)
National Institute of Advanced Industrial Science and Technology (AIST),
Tsukuba, Ibaraki 305-8568 JAPAN

William H. Wilson (billw@cse.unsw.edu.au)
School of Computer Science and Engineering, The University of New South Wales,
Sydney, New South Wales, 2052 AUSTRALIA

Keywords: systematicity; classicism; connectionism; compositionality; category theory; product; functor; adjunction

Abstract

Classical and Connectionist theories of cognitive architecture "explain" systematicity, whereby the capacity for some cognitive behaviors is intrinsically linked to the capacity for others, as a consequence of syntactically and functionally combinatorial representations, respectively. However, both theories depend on *ad hoc* assumptions to exclude specific architectures—grammars, or Connectionist networks—that do not account for systematicity. By analogy with the Ptolemaic (i.e., geocentric) theory of planetary motion, although either theory can be made to be consistent with the data, both nonetheless fail to explain it (Aizawa, 2003b). Category theory provides an alternative explanation based on the formal concept of adjunction, which consists of a pair of structure preserving maps, called functors. A functor generalizes the notion of a map between representational states to include a map between state transformations (processes). In a formal sense, systematicity is a necessary consequence of a "higher-order" theory of cognitive architecture, in contrast to the "first-order" theories derived from Classicism or Connectionism. Category theory offers a re-conceptualization for cognitive science, analogous to the one that Copernicus provided for astronomy, where representational states are no longer the center of the cognitive universe—replaced by the relationships between the maps that transform them.

Introduction

For more than two decades since Fodor and Pylyshyn's seminal paper on the foundations of a theory of cognitive architecture (Fodor & Pylyshyn, 1988), the problem of explaining systematicity remains unresolved (Aizawa, 2003b) despite numerous Classicist and Connectionist claims to the contrary (Fodor & McLaughlin, 1990; van Gelder, 1990; Smolensky, 1987).

The problem of systematicity for a theory of cognition is to explain why the capacity for some cognitive behaviours is intrinsically linked to some other cognitive capacities. The systematicity problem is actually three problems:

1. *Systematicity of representation*—why is it the case that the capacity to generate some representations (e.g., the representation `John loves Mary`) is intrinsically linked to the capacity to generate some other representations (e.g., the representation `Mary loves John`)?

2. *Systematicity of inference*—why is it the case that the capacity to make some inferences (e.g., that John is the lover in the proposition John loves Mary) is intrinsically linked to the capacity to make some other inferences (e.g., that Mary is the lover in the proposition Mary loves John)?

3. *Compositionality of representation*—why is it the case that the capacity for some semantic content (e.g., the thought that John loves Mary, however that thought may be represented) is intrinsically linked to the capacity for some other semantic context (e.g., the thought that Mary loves John, however that thought may also be represented)?

These problems are logically independent—one does not necessarily follow from another (Aizawa, 2003a), and so a theory is required it explain all three.

Classicists and Connectionists employ some form of combinatorial representations to explain systematicity. For Classicists, representations are combined in such a way that tokening of representations of complex entities entails tokening of representations of their constituent entities, so that the syntactic relationships between the constituent representations mirror the semantics ones—systematicity is a result of a combinatorial syntax and semantics (Fodor & Pylyshyn, 1988). For Connectionists, representations of complex entities are constructed more generally so that their tokening does not necessarily imply tokening constituent entity representations (van Gelder, 1990; Smolensky, 1987). We refer to the former as *classical compositionality*, and the latter as *functional compositionality*.

In general, a Classical or Connectionist architecture can demonstrate systematicity by having the "right" collection of grammatical rules, or functions such that one capacity is indivisibly linked to another. Suppose, for example, a Classical system with the following three production rules:

G1: `P` → `Agent loves Patient`
`Agent` → `John | Mary`
`Patient` → `John | Mary`.

The capacities to generate all four representations (i.e., `John loves John`, `John loves Mary`, etc.) are indivisibly linked, because the presence of all three, or absence of any one of those rules means the system is only capable of generating either all or none of those representations. In no case can the

system generate one without being able to generate the other. So, this Classical architecture has the systematicity of representation property with respect to this group of four propositions. Tensor products (Smolensky, 1990), or Godel numbers (van Gelder, 1990) are functionally compositional analogues to this explanation. Systematicity of inference follows from having additional processes that are sensitive to the structure of these representations. For Classical architectures, compositionality of representation also follows, because the semantic content of a complex representation is built up from the semantic contents of the constituents and their syntactic relationships (Aizawa, 2003a). Aizawa (2003a, 2003b) disputes whether a Connectionist architecture can also demonstrate compositionality of representation. Regardless, though, neither Classicism, nor Connectionism can derive theories that provide a full account of systematicity (Aizawa, 2003b).

A demonstration of systematicity is not an explanation for it. In particular, although grammar G1 has the systematicity of representation property, the following grammar:

G2: `P` → `John loves Patient` | `Agent loves Mary`
`Agent` → `John` | `Mary`
`Patient` → `John` | `Mary`

does not. This architecture cannot generate a representation of the proposition Mary loves John even though it can generate representations of both John and Mary as agents and patients, and the John loves Mary proposition. The essential problem for Classical theory—likewise Connectionist theory—is that syntactic compositionality by itself is not sufficient without some additional assumptions that admit grammars such as G1 that have the systematicity property, but exclude grammars such as G2 that do not. An explanation for systematicity in these cases turns on the nature of those additional, possibly *ad hoc* assumptions.

Ad hoc assumptions

Aizawa (2003b) presents an explanatory standard for systematicity and the problem of *ad hoc* assumptions by analogy with the Ptolemean (geocentric) versus Copernican (heliocentric) explanations for the motions of the planets (see Phillips, 2007, for a review). The geocentric explanation for planetary motion places the Earth at the center of the other planets' circular orbits. Although this theory can roughly predict planetary position, it fails to predict periods of apparent retrograde motion for the superior planets (i.e. Mars, Jupiter, etc.) across the night sky. To accommodate this data, the geocentric theory was augmented with the assumption that the other planets revolve around points that revolve around the Earth. This additional assumption is *ad hoc* in that it is unconnected with the rest of the theory and motivated only by the need to fit the data—the assumption could not be confirmed independently of confirming the theory. The heliocentric explanation, having all planets move around the Sun, eschews this *ad hoc* assumption. Retrograde motion falls out as a natural consequence of the positions of the Earth and other planets relative to the Sun. Tellingly, as more accurate data became available, the geocentric theory had to be further augmented with epicycles on epicycles to account for planetary motion; not so for the heliocentric theory.

The problem for Classical and Connectionist theories is that they cannot explain systematicity without recourse to their own *ad hoc* assumptions (Aizawa, 2003b). For Classicism, having a combinatorial syntax and semantics does not differentiate between grammars such as G1 and G2. For Connectionism, a common recourse to learning also does not work, whereby systematicity is acquired by adjusting network parameters (e.g., connection weights) to realize some behaviours—training set—while generalizing to others—test set. Learning also requires *ad hoc* assumptions, because even widely used learning models, such as feedforward (Rumelhart, Hinton, & Williams, 1986) and simple recurrent networks (Elman, 1990), fail to achieve systematicity (Marcus, 1998; Phillips, 2000) when construed as a degree of generalization (Hadley, 1994; Niklasson & Gelder, 1994). Hence, neither Classical nor Connectionist proposals satisfy the explanatory standard laid out by Aizawa, or Fodor and Pylyshyn for that matter.

Our category-theory based approach addresses the problem of *ad hoc* assumptions because the concept of an adjunction, which is central to our argument, ensures that the construct we seek (a) exists, and (b) is unique. That is to say, from this core assumption and category theory principles, the systematicity property necessarily follows for the particular cognitive domains of interest, because in each case the one and only collection of cognitive capacities derived from our theory is the systematic collection, without further restriction by additional (*ad hoc*) assumptions.

Basic category theory

Category theory is a theory of structure *par excellence* (see Awodey, 2006; Mac Lane, 2000, for introductions). It was developed out of a need to formalize commonalities between various mathematical structures (Eilenberg & Mac Lane, 1945), and has been used extensively in computer science for the analysis of computation (see, e.g., Pierce, 1991; Walters, 1991). Yet, applications to cognitive psychology have been almost non-existent (but, see Halford & Wilson, 1980; Phillips, Wilson, & Halford, 2009, for two examples). Our explanation of systematicity with respect to binary relational propositions is based on the concept of an *adjunction*. In this section, we provide definitions of this and other formal concepts that it depends.

Category

A *category* $\mathbf{C}$ consists of a class of objects $|\mathbf{C}| = (A, B, \dots)$; 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 satisfy the laws of:

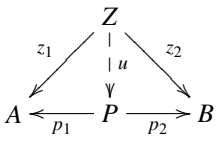
- *unity*, where $f \circ 1_A = f = 1_B \circ f$, for all $f : A \to B$; and
- *associativity*, where $h \circ (g \circ f) = (h \circ g) \circ f$, for all $f : A \to B$, $g : B \to C$ and $h : C \to D$.

The most familiar example of a category is **Set**, which has sets for objects and functions for morphisms, where the identity morphism 1_A is the identity function and the composition operation is the usual function composition operator "$\circ$".

A morphism $f : A \to B$ is an *isomorphism* if there exists a $g : B \to A$, such that $g \circ f = 1_A$ and $f \circ g = 1_B$. In this case, A is said to be isomorphic to B, written $A \cong B$.

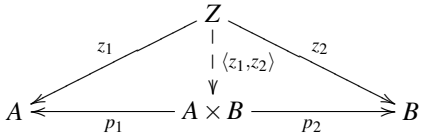
Product

A *product* of two objects A and B in a category $\mathbf{C}$ is an object P together with two morphisms $p_1 : P \to A$ and $p_2 : P \to B$, such that for any pair of morphisms $z_1 : Z \to A$ and $z_2 : Z \to B$, there is a unique morphism $u : Z \to P$, such that the following diagram commutes:



(1)

where a broken arrow indicates that there exists exactly one morphism making the diagram commute. That is, the compositions along any two paths with the same start object and the same finish object are the same. So, in this diagram, $z_1 = p_1 \circ u$ and $z_2 = p_2 \circ u$, where p_1 and p_2 are sometimes called projection morphisms. A product object P is *unique up to a unique isomorphism*. That is, for any other product object P' with morphisms $p'_1 : P' \to A$ and $p'_2 : P' \to B$ there is one and only one isomorphism between P and P' that makes a diagram like this one commute. Hence, P is not unique, only unique with respect to another product object via isomorphism. In **Set**, P is (up to isomorphism) the Cartesian product $A \times B$, $p_1 : A \times B \to A$, $p_2 : A \times B \to B$, where p_1 and p_2 are the projection maps to A and B, i.e., $p_1 : (a,b) \mapsto a$, and $p_2 : (a,b) \mapsto b$, and u is the function $\langle z_1, z_2 \rangle : Z \to A \times B$, sending x to tuple $(z_1(x), z_2(x))$, so that $p_1 \circ u = z_1$ and $p_2 \circ u = z_2$. (The $\mapsto$ arrow, often read as "maps to", indicates the action of a function on a domain element. Thus $f(a) = b$ is equivalent to $f : a \mapsto b$.) Since u is uniquely determined by z_1 and z_2, u is often written as $\langle z_1, z_2 \rangle$, and the diagram used in defining a product then becomes



(2)

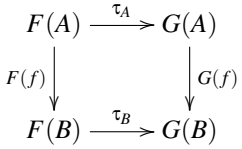
Functor

A functor $F : \mathbf{C} \to \mathbf{D}$ is a structure-preserving map between categories $\mathbf{C}$ and $\mathbf{D}$ that associates each object A in $\mathbf{C}$ to an object $F(A)$ in $\mathbf{D}$; and each map $f : A \to B$ in $\mathbf{C}$ to a map $F(f) : F(A) \to F(B)$ in $\mathbf{D}$, such that $F(1_A) = 1_{F(A)}$ for each object A in $\mathbf{C}$; and $F(g \circ_{\mathbf{C}} f) = F(g) \circ_{\mathbf{D}} F(f)$ for all maps $f : A \to B$ and $g : B \to C$ for which compositions $\circ_{\mathbf{C}}$ and $\circ_{\mathbf{D}}$ are defined in categories $\mathbf{C}$ and $\mathbf{D}$, respectively. The object and arrow components of a functor are sometimes explicitly distinguished as F_0 and F_1, respectively. Otherwise, the functor component is implicitly identified by its argument.

Functor composition and isomorphism are defined analogously to maps (above). That is, the composition of functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{E}$ is the functor $G \circ F : \mathbf{C} \to \mathbf{E}$, sending all objects A in $\mathbf{C}$ to objects $G \circ F(A)$ in $\mathbf{E}$; and maps $f : A \to B$ in $\mathbf{C}$ to maps $G \circ F(f) : G \circ F(A) \to G \circ F(B)$, such that identity and composition are respected. That is, $G \circ F(1_A) = 1_{G \circ F(A)}$; and $G \circ F(g \circ_{\mathbf{C}} f) = (G \circ F(g)) \circ_{\mathbf{E}} (G \circ F(f))$. A functor $F : \mathbf{C} \to \mathbf{D}$ is an *isomorphic functor*, if and only if there exists a functor $G : \mathbf{D} \to \mathbf{C}$ such that $G \circ F = 1_{\mathbf{C}}$ and $F \circ G = 1_{\mathbf{D}}$, where $1_{\mathbf{C}}$ and $1_{\mathbf{D}}$ are the identity functors sending objects and maps to themselves in the respective categories.

Natural transformation

A *natural transformation* $\tau : F \to G$ is a structure-preserving map from domain functor $F : \mathbf{C} \to \mathbf{D}$ to codomain functor $G : \mathbf{C} \to \mathbf{D}$ that consists of $\mathbf{D}-$maps τ_A for each object A in $\mathbf{C}$, such that $G(f) \circ \tau_A = \tau_B \circ F(f)$, as indicated by the following commutative diagram in the category $\mathbf{D}$:

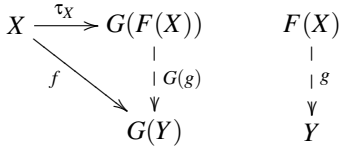


(3)

A natural transformation is a *natural isomorphism*, or *natural equivalence* if and only if each τ_A is an isomorphism. That is, for each $\tau_A : F(A) \to G(A)$ there exists a $\tau_A^{-1} : G(A) \to F(A)$ such that $\tau_A^{-1} \circ \tau_A = 1_{F(A)}$ and $\tau_A \circ \tau_A^{-1} = 1_{G(A)}$. Natural transformations also compose, and the composition of two natural transformations is also a natural transformation.

Adjunction

An *adjunction* consists of a pair of functors $F : \mathbf{C} \to \mathbf{D}$, $G : \mathbf{D} \to \mathbf{C}$ and a natural transformation $\tau : 1_{\mathbf{C}} \to (G \circ F)$, such that for every $\mathbf{C}-$object X and every $\mathbf{C}-$map $f : X \to G(Y)$ there exists a unique $\mathbf{D}-$map $g : F(X) \to Y$, such that the following diagram commutes:



(4)

where the functors are implicitly identified by (co)domain categories $\mathbf{C}$ (left subdiagram) and $\mathbf{D}$ (right subdiagram). The two functors are called an *adjoint pair*, (F, G), where F is the *left adjoint* of G, and G is the *right adjoint* of F; and natural transformation τ is called the *unit* of the adjunction.

Category theory explanation: Adjoint functors

We develop our adjoint functors explanation of systematicity in three movements. First, we show that a categorical product provides an account of systematicity of representation and systematicity of inference. However, a product of two objects may afford many isomorphic product objects that do not also account for compositionality of representation. Second, we show that the product functor provides the principled means for constructing only those products that also have the compositionality of representation property. However, there may be more than one product that has the compositionality property, but differs in semantic content by having different syntactic relationships between identical sets of constituents. So, a principled choice is needed to determine *the* product. Third, we show that the diagonal functor, which is left adjoint to the product functor, provides that principled choice. For concreteness, we refer to the category **Set**, but our explanation does not depend on this category.

First, suppose objects A (say, agents) and B (patients) are sets containing representations of John and Mary, denoted as $\{\mathtt{J},\mathtt{M}\}$. Although A and B are the same set in this example they may not be in the general case. Since our argument does not depend on equality, we maintain distinct names for generality, and for conceptual clarity. A categorical product of these two sets is the Cartesian product of A and B, which is the set of all pairwise combinations of elements from A and B, together with maps p_1 and p_2 for retrieving the first and second constituent in each case. That is, $A \times B = \{(\mathtt{J},\mathtt{J}),(\mathtt{J},\mathtt{M}),(\mathtt{M},\mathtt{J}),(\mathtt{M},\mathtt{M})\}$, $p_1 : (a,b) \mapsto a$, and $p_2 : (a,b) \mapsto b$. By definition, the Cartesian product, $A \times B$, generates all pairwise combinations of elements from A and B, therefore the Cartesian product has the systematicity of representation property. Moreover, by definition, the categorical product, $(A \times B, p_1, p_2)$, affords the retrieval of each constituent from each representation (otherwise it is not a product), therefore the categorical product also has the systematicity of inference property. In this case, Z from the categorical product definition takes the role of input, so inferring John as the lover from John loves Mary is just $z_1(\mathtt{JM}) = p_1 \circ u(\mathtt{JM})$, where JM is the input and u is the input-to-product object map, whose unique existence is guaranteed.

The Cartesian product, however, is not the only product object that satisfies the definition of a categorical product of A and B. An alternative product has $P = \{1,2,3,4\}$ as the product object, and $p'_1 : 1 \mapsto \mathtt{J}, 2 \mapsto \mathtt{J}, 3 \mapsto \mathtt{M}, 4 \mapsto \mathtt{M}$ and $p'_2 : 1 \mapsto \mathtt{J}, 2 \mapsto \mathtt{M}, 3 \mapsto \mathtt{J}, 4 \mapsto \mathtt{M}$ as the projections. However, this alternative does not have the compositionality of representation property: the semantic contents of these representations, whatever they may be, are not systematically related to each other, or the semantic content of John, or Mary. Hence, categorical products, in themselves, are not sufficient for an explanation of systematicity.

Second, for any category $\mathbf{C}$ that has products (i.e. every pair of objects in $\mathbf{C}$ has a product), one can define a product functor $\Pi : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$, that is from the Cartesian product of categories, $\mathbf{C} \times \mathbf{C}$, itself a category, to $\mathbf{C}$, where $\Pi_0 : (A,B) \mapsto A \times B$, $\Pi_1 : (f,g) \mapsto f \times g$, as indicated by the following diagram:

$$\begin{array}{ccc} (A,B) & \xmapsto{\Pi_0} & A \times B \\ {\scriptstyle (f,g)}\downarrow & & \downarrow{\scriptstyle f \times g} \\ (C,D) & \xmapsto[\Pi_0]{} & C \times D \end{array} \tag{5}$$

omitting $\Pi_1 : (f,g) \mapsto f \times g$ for clarity. In this case, the semantic contents of these elements are systematically related to each other and their constituents John and Mary. This categorical construction is an instance of Classical compositionality, whereby the constituents $a_i \in A$, $b_j \in B$ are tokened wherever the compositions $(a_i, b_j) \in A \times B$ are tokened. As such, it has the compositionality of representation property.

Although the product functor explanation accounts for compositionality of representation, it introduces a new problem: $(B \times A, p'_2, p'_1)$, where $p'_2 : (b,a) \mapsto a$ and $p'_1 : (b,a) \mapsto b$ is also a valid product, but the semantic content of (a,b) is not the same as (b,a). That is because they have different order relationships between their constituents even though the corresponding constituents are identical. Thus, a principled choice is required to determine whether, for example, John loves Mary should map to (John, Mary), or (Mary, John). Otherwise, one can define an architecture that does not have the systematicity of inference property by employing both products to correctly infer John as the lover in John loves Mary via $(A \times B, p_1, p_2)$, yet incorrectly infer John as the lover in Mary loves John via $(B \times A, p'_2, p'_1)$, where position within the product triple identifies the relevant projection. The assumption that architectures employ only the first product is *ad hoc* just like the assumption that Classical architectures employ grammars such as G1, but not G2. So, a principled choice is needed to determine *the* product.

Third, and finally, the left adjoint to the product functor is the *diagonal* functor $\Delta : \mathbf{C} \to \mathbf{C} \times \mathbf{C}$, where $\Delta_0 : A \mapsto (A,A)$, $\Delta_1 : f \mapsto (f,f)$ as indicated by the following diagram:

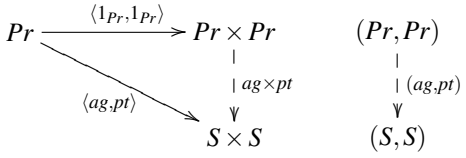
$$\begin{array}{ccc} A & \xmapsto{\Delta_0} & (A,A) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle (f,f)} \\ B & \xmapsto[\Delta_0]{} & (B,B) \end{array} \tag{6}$$

The (diagonal, product) adjoint pair is indicated by the following commutative diagram:

$$\begin{array}{ccccc} C & \xrightarrow{\tau_C = \langle 1_C, 1_C \rangle} & C \times C & \qquad & (C,C) \\ & \searrow^{\langle s,t \rangle} & \downarrow{\scriptstyle s \times t} & & \downarrow{\scriptstyle (s,t)} \\ & & M \times N & & (M,N) \end{array} \tag{7}$$

(see Pierce, 1991, Example 2.4.6). In this manner, the John loves Mary family of cognitive capacities is specified by the

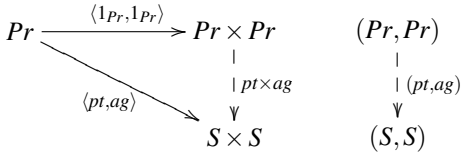
commutative diagram



(8)

where *ag* and *pt* are the *agent* and *patient* maps from the set of proposition inputs *Pr* into the set $S \supseteq A \cup B$ containing all the possible constituent representations. Given $\langle ag, pt \rangle$ as the morphism used by the architecture to map proposition inputs to their corresponding internal representations, then as mentioned (Introduction) the definition of an adjunction guarantees that $ag \times pt$ is unique with respect to making Diagram 8 commute. That is, $ag \times pt \circ \langle 1_{Pr}, 1_{Pr} \rangle(\texttt{JM}) = ag \times pt(\texttt{JM}, \texttt{JM}) = (\texttt{John}, \texttt{Mary}) = \langle ag, pt \rangle(\texttt{JM})$, where JM is the input for proposition John loves Mary. The alternative construction $pt \times ag$ is excluded because $pt \times ag \circ \langle 1_{Pr}, 1_{Pr} \rangle(\texttt{JM}) = pt \times ag(\texttt{JM}, \texttt{JM}) = (\texttt{Mary}, \texttt{John}) \neq (\texttt{John}, \texttt{Mary}) = \langle ag, pt \rangle(\texttt{JM})$. Having excluded $pt \times ag$ by the commutativity property of the adjunction, the only two remaining ways to map the other inputs (i.e., $\langle ag, pt \rangle$ and $ag \times pt \circ \langle 1_{Pr}, 1_{Pr} \rangle$) are equal. So, given that the architecture can represent John loves Mary as $(\texttt{John}, \texttt{Mary})$ via $\langle ag, pt \rangle$ and infer John as the lover via p_1 from the product $(A \times B, p_1, p_2)$, then necessarily it can represent Mary loves John and infer Mary as the lover using the same maps. That is, $p_1 \circ \langle ag, pt \rangle(\texttt{MJ}) = p_1(\texttt{Mary}, \texttt{John}) = \texttt{Mary}$, or $p_1 \circ ag \times pt \circ \langle 1_{Pr}, 1_{Pr} \rangle(\texttt{MJ}) = p_1 \circ ag \times pt(\texttt{MJ}, \texttt{MJ}) = p_1(\texttt{Mary}, \texttt{John}) = \texttt{Mary}$.

This explanation works regardless of whether proposition John loves Mary is represented as (John, Mary) via $\langle ag, pt \rangle$, or (Mary, John) via $\langle pt, ag \rangle$. In the latter case, the adjunction picks out the construction $pt \times ag$, because it is the one and only one that makes the following diagram commute:



(9)

$pt \times ag \circ \langle 1_{Pr}, 1_{Pr} \rangle(\texttt{JM}) = pt \times ag(\texttt{JM}, \texttt{JM}) = (\texttt{Mary}, \texttt{John}) = \langle pt, ag \rangle(\texttt{JM})$, but $ag \times pt \circ \langle 1_{Pr}, 1_{Pr} \rangle(\texttt{JM}) = ag \times pt(\texttt{JM}, \texttt{JM}) = (\texttt{John}, \texttt{Mary}) \neq (\texttt{Mary}, \texttt{John}) = \langle pt, ag \rangle(\texttt{JM})$. Given that the architecture can represent John loves Mary as $(\texttt{Mary}, \texttt{John})$ via $\langle pt, ag \rangle$ and infer John as the lover via p'_2 from the product $(B \times A, p'_2, p'_1)$, then necessarily it can do so for Mary loves John using the same maps. That is, $p'_2 \circ \langle pt, ag \rangle(\texttt{MJ}) = p'_2(\texttt{John}, \texttt{Mary}) = \texttt{Mary}$, or $p'_2 \circ pt \times ag \circ \langle 1_{Pr}, 1_{Pr} \rangle(\texttt{MJ}) = p'_2 \circ pt \times ag(\texttt{MJ}, \texttt{MJ}) = p'_2(\texttt{John}, \texttt{Mary}) = \texttt{Mary}$.

Importantly, the unit of the adjunction, $\langle 1_{Pr}, 1_{Pr} \rangle$, is not a *free parameter* of the explanation; it defines the adjunction. Also, there is no choice in representational format (i.e. left-right, or right-left constituent order)—the given capacity to represent a proposition fixes the same order for all the other propositions. Hence, systematicity is a necessary consequence of this adjoint pair without recourse to additional (*ad hoc*) assumptions, and so meets the explanatory standard set by Aizawa, and Fodor and Pylyshyn.

Explanatory levels: *n*-category theory

A generalization of category theory, called *n*-category theory (see Leinster, 2003) is used to formally contrast our category theory explanation against Classical and Connectionist approaches. Notice that the definitions of functor and natural transformation are very similar. In fact, they are morphisms at different levels of analysis. For *n*-category theory, a category such as **Set** is a 1-category, with 0-objects (i.e. sets) for objects and 1-morphisms (i.e. functions) for arrows. A functor is a morphism between categories. The category of categories, **Cat**, has categories for objects and functors for arrows. Thus, a functor is a 2-morphism between 1-objects (i.e. 1-categories) in a 2-category. A natural transformation is a morphism between functors. The functor category, **Fun**, has functors for objects and natural transformations for arrows. Thus, a natural transformation is a 3-morphism between 2-objects (i.e. functors) in a 3-category. (A 0-category is just a *discrete* category, where the only arrows are identities, which are 0-morphisms.) In this way, the order *n* of the category provides a formal notion of explanatory level.

Classical or Connectionist compositionality is essentially a lower levels attempt to account for systematicity. For the examples we used that level is perhaps best described in terms of a 1-category. Indeed, a context-free grammar defined by a graph is modeled as the *free* category on that graph containing sets of terminal and non-terminal symbols for objects and productions for morphisms (Walters, 1991). By contrast, our category theory explanation involves higher levels of analysis, specifically functors and natural transformations, which live in 2-categories and 3-categories, respectively. Of course, one can also develop higher-order grammars that take as input or return as output other grammars. Similarly, one can develop higher-order networks that take as input or return as output other networks. However, the problem is that neither Classical nor Connectionist compositionality delineates those (higher-order) grammars or networks that have the systematicity property from those that do not.

Discussion

In addition to explaining systematicity, our category theory approach has further implications. According to our explanation, systematicity with respect to binary relational propositions requires a category with products. Phillips et al. (2009) also provided a category theory account of the strikingly similar profiles of development for a suite of reasoning abilities that included *Transitive Inference* and *Class Inclusion*, among others—all abilities are acquired around the age of five years. The difference between the failures of younger children and the successes of older children (relative to age five) across all these reasoning tasks was explained as their capacity to compute (co)products. (A *coproduct* is related to a product

by arrow reversal—see, e.g., Pierce, 1991, for a formal definition.) Therefore, our explanation implies that systematicity is not a property of younger children's cognition. Some support for this implication is found on memory tasks that require binding the background context of memorized items (Lloyd, Doydum, & Newcombe, 2009), though further work is needed to test this implication directly.

Our explanation does not depend on **Set**, it only requires a category with products. For example, the categories **Top** of topological spaces and continuous mappings, and **Vec** of vector spaces and linear mappings (see, e.g., Awodey, 2006) could also be used. These possibilities imply that an explanation of systematicity does not depend on a particular (discrete symbolic, or continuous subsymbolic) representational format. Thus, a further benefit is that our approach opens the way for integration of other (sub/symbolic) levels of analysis.

For reasons of space, we have only sketched our category theory approach to systematicity. More detailed explanation and justification are given in Phillips and Wilson (in prep.), where we also address other examples of systematicity, such as multiple relations, and relational schemas. In our approach, we have not dealt with domains that are quasi-systematic, which appear to be particularly prevalent in language (see Johnson, 2004). For these cases, we would also need category theory-derived principled restrictions to products. *Pullbacks* (see Phillips, Wilson, & Halford, 2009, for an application to cognitive development) are one way to restrict product objects, in the same arrow-theoretic style.

From a category theory perspective, we now see why cognitive science lacked a satisfactory explanation for systematicity—cognitive scientists were working with lower-order theories in attempting to explain an essentially higher-order property. Category theory offers a re-conceptualization for cognitive science, analogous to the one that Copernicus provided for astronomy, where representational states are no longer the center of the cognitive universe—replaced by the relationships between the maps that transform them.

Acknowledgment. We thank the reviewers for extensive comments to help clarify the presentation of this work.

References

Aizawa, K. (2003a). Cognitive architecture: The structure of cognitive representations. In S. P. Stich & T. A. Warfield (Eds.), *The Blackwell guide to philosophy of mind* (pp. 172–189). Cambridge, MA: Blackwell.

Aizawa, K. (2003b). *The systematicity arguments*. New York: Kluwer Academic.

Awodey, S. (2006). *Category theory*. New York, NY: Oxford University Press.

Eilenberg, S., & Mac Lane, S. (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., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, *35*, 183–204.

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

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

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

Johnson, K. (2004). On the systematicity of language and thought. *The Journal of Philosophy*, *101*(3), 111–139.

Leinster, T. (2003). *Higher operads, higher categories*. Cambridge: UK: Cambridge University Press.

Lloyd, M. E., Doydum, A. O., & Newcombe, N. S. (2009). Memory binding in early childhood: evidence for a retrieval deficit. *Child Development*, *80*(5), 1321–1328.

Mac Lane, S. (2000). *Categories for the working mathematician* (2nd ed.). New York, NY: Springer.

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

Niklasson, L., & Gelder, T. van. (1994). Systematicity and connectionist language learning. *Mind and Language*, *9*(3), 288–302.

Phillips, S. (2000). Constituent similarity and systematicity: The limits of first-order connectionism. *Connection Science*, *12*(1), 1–19.

Phillips, S. (2007). Kenneth Aizawa, The systematicity arguments, Studies in brain and mind. *Minds and Machines*, *17*(3), 357–360.

Phillips, S., & Wilson, W. H. (in prep.). *Categorial compositionality: A category theory explanation for the systematicity of human cognition.*

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

Pierce, B. C. (1991). *Basic category theory for computer scientists*. Cambridge, UK: MIT Press.

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

Smolensky, P. (1987). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *Southern Journal of Philosophy*, *26*, 137–161.

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

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

Walters, R. F. C. (1991). *Categories and computer science*. Cambridge, UK: Cambridge University Press.