Categorial compositionality continued: A category theory explanation for quasi-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: quasi-systematicity; classicism; connectionism; compositionality; category theory; product; pullback; functor; natural transformation; adjunction; universal construction

Abstract

The classical account for systematicity of human cognition supposes: (1) syntactically compositional representations; *and* (2) processes that are sensitive to their structure. The problem with this account is that there is no explanation as to *why* these two components must be compatible, other than by *ad hoc* assumption (convention) to exclude nonsystematic variants that, e.g., mix prefix and postfix concatenative compositional schemes. Recently, we proposed an alternative explanation (Phillips & Wilson, 2010) without *ad hoc* assumptions, using a branch of mathematics, called *category theory*. In this paper, we extend our explanation to domains that are *quasi-systematic* (e.g., language), where the domain includes some but not all possible combinations of constituents. The central category-theoretic construct is an *adjunction* involving *pullbacks*, where the focus is on the relations between processes, rather than the representations. In so far as cognition is systematic, the basic building blocks of cognitive architecture are adjunctions by our theory.

Introduction

A complete theory of human cognition must explain why our mental abilities are organized into particular groups of behaviours rather than just some arbitrary, random collection of cognitive capacities. This property of cognitive architecture (i.e., the collection of basic processes and modes of composition that together generate cognitive behaviour) is called *systematicity* (Fodor & Pylyshyn, 1988), and the problem posed for a theory of cognition is to explain why systematicity is a necessary consequence of the assumptions and principles embodied by the architecture that the proposed theory posits (Fodor & Pylyshyn, 1988; Aizawa, 2003).

The classical explanation derives from the principle of *classical compositionality*, which says that cognitive representations and processes are constructed from a combinatorial syntax and semantics, whereby semantic relations between constituents of the complex entities represented by a cognitive system are mirrored by syntactic relations between the corresponding constituent representations—that is, syntactically structured representations *and* processes that are sensitive to (i.e., compatible with) those structures (Fodor & Pylyshyn, 1988).

To account for systematicity, the two parts of the classical compositionality principle, i.e. (1) combinatorial syntax and semantics, and (2) structure-sensitive processes, must be compatible. However, classical theory does not explain *why* they must be compatible, other than by assumption. By convention, one may assume an *infix* mode of classical concatenative compositionality, whereby *John loves Mary* is represented by [John Loves Mary]. Yet, by convention, one may also choose a *prefix* mode, e.g., [Loves John Mary], or a *postfix* mode, e.g., [John Mary Loves], as employed in some (programming) languages, or even where argument order is reversed, e.g., [Mary Loves John]. All these possibilities are valid forms of classical compositionality, but an architecture that employs incompatible combinations will not exhibit systematicity. Classical compositionality does not fully explain systematicity because of the *ad hoc* assumption that only certain combinations are permitted, which is enforced by the cognitive scientist not the cognitive system. For an extended discussion on the problem of *ad hoc* assumptions in science generally, and classical/connectionist explanations of systematicity specifically, see Aizawa (2003).

At this point, modellers may think to augment their theory with some sort of learning principle, such as is commonly incorporated into connectionist (Rumelhart, Hinton, & Williams, 1986), and Bayesian modeling (Tenenbaum, Griffiths, & Kemp, 2006). However, connectionist and Bayesian approaches suffer the same shortcoming: while both are capable of configuring architectures with the desired form of systematicity, they are likewise able to configure architectures without that form of systematicity from that same learning principle (see also Phillips & Wilson, 2010, on this point).

Recently, we presented an alternative explanation for systematicity without recourse to such *ad hoc* assumptions (Phillips & Wilson, 2010) that employed a branch of mathematics called *category theory* (Mac Lane, 2000), where the theoretical focus is on the relationships between structure-sensitive processes, rather than the representations on which they operate. In particular, the category theory notion of *functor* maps (generalizations of) functions to (generalized) functions, as well as mapping objects to objects. The central explanatory element in Phillips and Wilson (2010) is the formal category theory concept of *adjunction*: an adjunction relates two functorial constructions so that of the *possibly* systematic capacity realizing constructions there is one and only one construction that realizes all systematically related capacities via the adjunction. Hence, no further, *ad hoc*, assumptions are required to distinguish systematic from unsystematic architectures, thus meeting the explanatory standard for systematicity in human cognition originally specified in Fodor and Pylyshyn (1988), and subsequently clarified in Aizawa (2003). In our theory, basic building blocks of human cognitive architecture involve the adjunctive relationships between functorial constructions.

Our explanation of systematicity was applied in two domains that involved cognitive capacities pertaining to (1) a common relation, and (2) a common relational schema. With respect to these domains, human cognition exhibits what we may call "full" systematicity, in the sense that capacity is extended to each and every combination of the possible constituents that may partake in the relation or schema. For example, suppose one has the capacity to represent entities *John*, *Mary*, *Sue*, *Tom*, and *loves*, and the relational proposition that *John loves Mary*, then one has the capacity to represent all combinations, such as *Sue loves Tom*, *Tom loves John*, *Mary loves Mary*, and so on.

Quasi-systematicity

Not all domains are fully (completely) systematic. Additional constraints relevant to the domain of interest preclude some combinations. In particular, linguistic constructions often incorporate different types of constraints, including syntactic, phonetic, semantic, and pragmatic constraints that may further restrict the group of capacities that are intrinsically connected (Johnson, 2004). For example, English-speakers say *John put his gear down*, but not *John stowed his gear down*, even though they say *John put his gear away*, or *John stowed his gear away* (see Johnson, 2004). In this sense, we say cognition is quasi-systematic with respect to this domain, where quasi-systematicity is just a further refinement to a more specialized collection of systematic (intrinsically connected) capacities. That systematicity and quasi-systematicity are just differences in degrees of the same basic phenomenon motivates our proposal for a general theory explaining both. Our purpose in this paper is to show how our category theory explanation of systematicity (Phillips & Wilson, 2010) generalizes to include quasi-systematicity.

Basic category theory

We introduce the category theory definitions used to explain quasi-systematicity. Numerous introductions to category theory are available: for mathematicians, see Awodey (2006); Mac Lane (2000); for computer scientists, see Pierce (1991); and for general interest, see Lawvere and Schanuel (1997). Further motivation and background for our use of category theory in cognition is given in Phillips and Wilson (2010), and also Phillips, Wilson, and Halford (2009).

Category

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

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

One may think of a category as modeling a cognitive domain, where objects are sets of cognitive states, and morphisms are cognitive processes mapping possible cognitive state transitions. In this case, the category is **Set** having sets for objects and functions for morphisms, where the identity morphism is the identity function sending elements to themselves and composition is the usual composition of functions. The category theory methods that we apply to systematicity are not specifically limited to **Set** and could be used with other categories. For example, the category **Met** of metric spaces (objects) and continuous functions (morphisms) may be appropriate for cognitive domains concerning continuous instead of discrete entities.

Functor

A *functor* $F : \mathbf{C} \rightarrow \mathbf{D}$ is a structure-preserving map from a domain category $\mathbf{C}$ to a codomain category $\mathbf{D}$ that sends each object $A \in |\mathbf{C}|$ to an object $F(A) \in |\mathbf{D}|$; and each morphism $f : A \rightarrow B \in \mathbf{C}(A,B)$ to a morphism $F(f) : F(A) \rightarrow F(B) \in \mathbf{D}(F(A), F(B))$, such that $F(1_A) = 1_{F(A)}$ for each object A; and $F(g \circ_{\mathbf{C}} f) = F(g) \circ_{\mathbf{D}} F(f)$ for all morphisms $f : A \rightarrow B$ and $g : B \rightarrow C$ for which compositions $\circ_{\mathbf{C}}$ and $\circ_{\mathbf{D}}$ are defined in categories $\mathbf{C}$ and $\mathbf{D}$, respectively.

Functors preserve structure in that every morphism in the domain category is associated with just one morphism in the codomain category, though this association does not have to be unique. Functors also provide a means for constructing new categories from old. In our context, one may think of functors as a means for constructing new cognitive representations and processes from existing ones. Thus, functors provide the formal starting point for a theory about the systematicity of cognitive capacities.

Natural transformation

A *natural transformation* $\eta : F \dot{\rightarrow} G$ from a functor $F : \mathbf{C} \rightarrow \mathbf{D}$ to a functor $G : \mathbf{C} \rightarrow \mathbf{D}$ consists of $\mathbf{D}$−maps $\eta_A : F(A) \rightarrow G(A)$ for each object $A \in |\mathbf{C}|$, such that for every morphism $f : A_1 \rightarrow A_2$ in $\mathbf{C}$ we have $G(f) \circ \eta_{A_1} = \eta_{A_2} \circ F(f)$, as indicated by the following *commutative* diagram (here "commutative" means that paths with the same start/end object yield the same morphism):

$$\begin{tikzcd}
F(A_1) \arrow[r, "\eta_{A_1}"] \arrow[d, "F(f)"'] & G(A_1) \arrow[d, "G(f)"] \\
F(A_2) \arrow[r, "\eta_{A_2}"'] & G(A_2)
\end{tikzcd} \tag{1}$$

A natural transformation is a *natural isomorphism*, or *natural equivalence* if and only if each η_A is an isomorphism. That is, for each $\eta_A : F(A) \rightarrow G(A)$ there exists a morphism $\eta_A^{-1} : G(A) \rightarrow F(A)$ such that $\eta_A^{-1} \circ \eta_A = 1_{F(A)}$ and $\eta_A \circ \eta_A^{-1} = 1_{G(A)}$.

Adjunction

The formal concept of adjunction is central to our explanation for (quasi-)systematicity. If we interpret functors as constructing cognitive representations and processes, then an adjoint relationship between two functors is a relationship between cognitive constructions. Except to explain systematicity (Phillips & Wilson, 2010), adjunctions do not appear to have been used for cognitive modeling, but see Magnan and Reyes (1995) for a conceptual introduction; see also Mac Lane (2000) for adjunctions in mathematics; in other fields see, e.g., Goguen (1972) in the context of general systems theory of abstract machines and behaviours. For some orientation, one may think of classical/connectionist approaches as primarily focussed on the processes that transform representations, at the expense of being unable to guarantee a systematic relationship between those processes. In contrast, an adjunction guarantees that the only pairings of functors modeling such processes are the systematic ones. Thus, systematicity follows without further, *ad hoc* assumptions.

An *adjunction* consists of a pair of functors $F : \mathbf{C} \rightarrow \mathbf{D}$, $G : \mathbf{D} \rightarrow \mathbf{C}$ and a natural transformation $\eta : 1_{\mathbf{C}} \dot{\rightarrow} (G \circ F)$, such that for every $\mathbf{C}-$object X, $\mathbf{D}-$object Y, and $\mathbf{C}-$map $f : X \rightarrow G(Y)$, there exists a unique $\mathbf{D}-$map $g : F(X) \rightarrow Y$, such that $G(g) \circ \eta_X = f$, as indicated by the following commutative diagram:

$$\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[dr, "f"'] & G \circ F(X) \arrow[d, dashed, "G(g)"] & F(X) \arrow[d, dashed, "g"] \\
 & G(Y) & Y
\end{tikzcd} \tag{2}$$

The two functors are called an *adjoint pair*, denoted (F, G), where F is the *left adjoint* of G (written, $F \dashv G$), and G is the *right adjoint* of F, and η is the *unit* of the adjunction.

An equivalent definition of adjunction is in terms of the *counit*, which presents the adjunction from the perspective of the second category (i.e., $\mathbf{D}$). An adjunction is an instance of a *universal construction*, and the unit and counit are *universal* (*mediating*) *arrows*. That is, every construction factors through them. Hence, the universal arrow accounts for the indivisible nature of systematic capacities without *ad hoc* assumptions, because all capacities factor through it uniquely.

Pullback (product)

A *pullback* of two morphisms $f : A \rightarrow C$ and $g : B \rightarrow C$ in category $\mathbf{C}$ is, up to unique isomorphism, an object P (also denoted $A \times_C B$) together with two morphisms $p_1 : P \rightarrow A$ and $p_2 : P \rightarrow B$, jointly expressed as (P, p_1, p_2), such that for every object $Z \in |\mathbf{C}|$ and pair of morphisms $z_1 : Z \rightarrow A$ and $z_2 : Z \rightarrow B$ there exists a unique morphism $u : Z \rightarrow P$, also denoted $\langle z_1, z_2 \rangle$, such that the following diagram commutes:

$$\begin{tikzcd}
 & Z \arrow[d, dashed, "\langle z_1, z_2 \rangle"] \arrow[ddl, bend right, "z_1"'] \arrow[ddr, bend left, "z_2"] & \\
 & A \times_C B \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\
A \arrow[dr, "f"'] & & B \arrow[dl, "g"] \\
 & C &
\end{tikzcd} \tag{3}$$

A pullback may be thought of as a product of objects A and B constrained by C and the morphisms f and g. In the category **Set**, for example, $A \times_C B$ is, up to unique isomorphism, the subset of the Cartesian product $A \times B$ that includes just those pairs of elements (a, b) satisfying the constraint that $f(a) = g(b) \in C$. With this intuition in mind, we can begin to see how pullbacks may pertain to quasi-systematicity of relations, which we address next. In a category $\mathbf{C}$, a *terminal object*, denoted 1, is an object, such that there is a morphism to it from every object in $\mathbf{C}$. Note that a pullback with $C = 1$ (i.e., a terminal) is equivalent to a product, effectively there is no constraint on the product. Thus, our explanation of quasi-systematicity subsumes our explanation of systematicity given in Phillips and Wilson (2010).

Explanation for quasi-systematicity

We apply our category theory approach to explaining quasi-systematicity with respect to natural relations and some aspects of language.

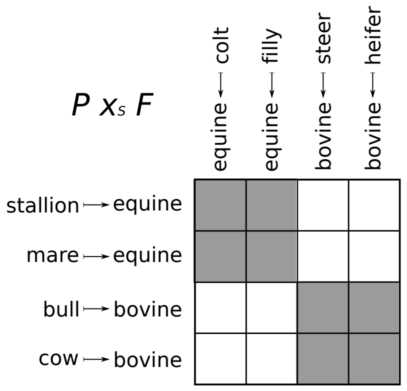
Natural relations

We use the relation *parent* (e.g., *mares parent colts*) to illustrate our explanation of quasi-systematicity in terms of pullbacks. If one knows that *mares parent colts* and *stallions parent fillies* then one also knows that *mares parent fillies* and *stallions parent colts*. Likewise, if one knows that *cows parent steers* and *bulls parent heifers*, then one also knows that *cows parent heifers* and *bulls parent steers*. Yet, one would not also think that *mares parent steers*, or *bulls parent fillies*. One also would not think that *colts parent stallions*, or *heifers parent bulls*. An architecture based only on a product is inadequate. Instead, the quasi-systematic capacities associated with this relation derive from a pullback.

The pullback diagram associated with the *parent* relation is an instantiation of Diagram 3. In particular, the pullback for this relation is given in the following commutative diagram:

$$\begin{tikzcd}
 & Pr \arrow[d, dashed, "\langle pg{,}os\rangle"] \arrow[ddl, bend right, "pg"'] \arrow[ddr, bend left, "os"] & \\
 & P\times_S F \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\
P \arrow[dr, "s_p"'] & & F \arrow[dl, "s_f"] \\
 & S &
\end{tikzcd} \tag{4}$$

where Pr is the set of valid propositions, with pg and os as the progenitor and offspring maps (respectively), P is the set of progenitors, F is the set of offspring, S is the set of species, s_P (s_F) map the progenitors (offspring) to their species, and $P\times_S F = \{(p,f)|s_p(p) = s_f(f)\}$. Suppose $P = \{stallion, mare, bull, cow\}$, $F = \{colt, filly, steer, heifer\}$, and $S = \{equine, bovine\}$, so $s_p : stallion \mapsto equine$, $mare \mapsto equine$, $bull \mapsto bovine$, $cow \mapsto bovine$, and $s_f : colt \mapsto equine$, $filly \mapsto equine$, $steer \mapsto bovine$, $heifer \mapsto bovine$. Then, $P\times_S F$ is the set {(*stallion*, *colt*), (*stallion*, *filly*), (*mare*, *colt*), (*mare*, *filly*), (*bull*, *steer*), (*bull*, *heifer*), (*mare*, *steer*), (*mare*, *heifer*)}, which contains just the elements in the *parent* relation (see Figure 1).



P x_s F	equine ⟵ colt	equine ⟵ filly	bovine ⟵ steer	bovine ⟵ heifer
stallion ⟶ equine	▨	▨		
mare ⟶ equine	▨	▨		
bull ⟶ bovine			▨	▨
cow ⟶ bovine			▨	▨

Figure 1: Matrix representation of *parent* pullback.

The adjoint for this example is $(\Delta_{\circledast}, \Pi_S)$, where S is the constraining object, and $\circledast$ refers to constructs that are specific to pullbacks. The adjunction is indicated by the following diagram:

$$\begin{tikzcd}
Pr \arrow[r, "\langle 1_{Pr}{,}1_{Pr}\rangle"] \arrow[dr, "\langle pg{,}os\rangle"'] & Pr\times_{Pr}Pr \arrow[d, dashed, "pg\times os"] & & Pr \arrow[r, "1_{Pr}"] \arrow[d, dashed, "pg"'] & Pr \arrow[d, dashed, "s_p\circ pg \mid s_f\circ os"] & Pr \arrow[l, "1_{Pr}"'] \arrow[d, dashed, "os"] \\
 & P\times_S F & & P \arrow[r, "s_p"'] & S & F \arrow[l, "s_f"]
\end{tikzcd} \tag{5}$$

where $s_p \circ pg = s_f \circ os$ and $Pr\times_{Pr}Pr \cong Pr$. Diagram 5 simplifies to:

$$\begin{tikzcd}
Pr \arrow[r, "\langle 1_{Pr}{,}1_{Pr}\rangle"] \arrow[dr, "\langle pg{,}os\rangle"'] & Pr\times_{Pr}Pr \arrow[d, dashed, "pg\times os"] & (1_{Pr}, 1_{Pr}) \arrow[d, dashed, "(pg{,}os)"'] \\
 & P\times_S F & (s_p, s_f)
\end{tikzcd} \tag{6}$$

where a composition such as $P \rightarrow S \leftarrow F$ in Diagram 5 is identified by the morphisms (i.e., s_p and s_f), and a map between such compositions by the corresponding morphisms between the *outer* objects. For example, P and F are the outer objects in $P \rightarrow S \leftarrow F$, and S is the *inner* object. Reference to the morphism between inner objects is omitted, because it is determined by the other morphisms.

The adjunction is also given from the perspective of the counit in the following diagram:

$$\begin{tikzcd}
Pr \arrow[d, dashed, "\langle pg{,}os\rangle"'] & (1_{Pr}, 1_{Pr}) \arrow[d, dashed, "(\langle pg{,}os\rangle{,}\langle pg{,}os\rangle)"'] \arrow[dr, "(pg{,}os)"] & \\
P\times_S F & (1_{P\times_S F}, 1_{P\times_S F}) \arrow[r, "(p_1{,}p_2)"'] & (s_p, s_f)
\end{tikzcd} \tag{7}$$

where mediating arrow $(p_1, p_2) : (P\times_S F, P\times_S F) \rightarrow (P,F)$ is the counit.

The explanation for quasi-systematicity comprises two parts: one part pertains to the constraints on allowable elements; and the other part pertains to universal construction, and is essentially the same explanation as that for full systematicity, except that the universal construction is associated with a pullback.

Regarding the constraints part of the explanation, there are two sources of constraints in the form of the sets containing the possibly related elements (i.e. P and F in this example), and the requirement that Diagram 4 commutes. That P contains only progenitors and F only offspring precludes pairs corresponding to *colts parent mares*, for example. The fact that Diagram 4 must commute (to be a pullback) precludes instances corresponding to *stallions parent steers*, for example, because stallion and steer belong to different species.

The universal construction part of the explanation parallels the explanation for full systematicity: given a cognitive capacity for a relation realized as a particular pullback, then the commutativity property of the adjunction ensures that there is one and only one way to realize the other capacities, obviating the need for an *ad hoc* assumption stipulating which pullback. In particular, $(F\times_S P, p'_1, p'_2)$, where $p'_1 : F\times_S P \rightarrow F$ and $p'_2 : F\times_S P \rightarrow P$, is also a pullback. Thus, from pullbacks alone an architecture can be constructed whereby mare is correctly inferred as the progenitor in *mares parent colts* by $(P\times_S F, p_1, p_2)$ and $p_1 : (mare, colt) \mapsto colt$, but *steers* is incorrectly inferred as the progenitor in *bulls parent steers* since $(F\times_S P, p'_1, p'_2)$ and $p'_1 : (steers, bulls) \mapsto steers$. The commutativity property of the adjunction rules out an architecture that mixes different possible pullbacks. As with full systematicity, quasi-systematic capacities are indivisibly linked by a universal arrow, i.e., (p_1, p_2).

This form of pullback is sufficient when the capacity subgroups (one subgroup per species, in this example) are themselves locally, fully systematic. In some situations, this condition may not hold. For example, suppose we introduce *whale* and *calf* as additional progenitor and offspring elements, respectively. By associating whale and calf with mammal, the pullback above would yield (*whale*, *calf*), but also (*whale*, *steer*), and (*whale*, *heifer*) where these elements were also associated with mammal. Clearly, the term calf is being used in two senses that need to be distinguished. One sense pertains just to cattle, and the broader sense includes large mammals, such as elephants and seals as the parents of calves. These subgroups can be distinguished by using another pullback that incorporates this additional structural information.

The pullback in this new situation is indicated in the following diagram:

$$\begin{tikzcd}
& Pr \arrow[d, dashed, "\langle pg{,}os\rangle"] \arrow[ddl, bend right, "pg"'] \arrow[ddr, bend left, "os"] & \\
& P \times S \times F \arrow[dl, "p_{1,2}"'] \arrow[dr, "p_{2,3}"] & \\
P \times S \arrow[dr, "p_2"'] & & S \times F \arrow[dl, "p_1"] \\
& S &
\end{tikzcd} \tag{8}$$

where the constraining object S contains additional, semantic, information distinguishing the senses of *calf*, and $p_{i,j}$ projects out the ith and jth elements of a triple. The two senses of *calf* are captured by pairing *bovine* with *calf* for one sense and *cetacea* with *calf* for the other in $S \times F$, and *bull* and *cow* with *bovine*, and *whale* with *cetacea* in $P \times S$. Since *bull*, *cow*, *steer* and *heifer*, etc. are not paired with *cetacea*, instances such as *whales parent steers* are not contained within the collection of quasi-systematic capacities.

Language

Our first example is subject-verb agreement: for English speakers, agreement between the subject and verb means that the capacity for *the dogs chase the cats* and *the dog chases the cats* implies the capacity for *the cats chase the dogs*, but not *the cats chases the dogs*, nor *the cat chase the dogs*, etc. The present example is confined to third-person agreement, though the explanation extends to first- and second-person. Subject-verb agreement is enforced by a pullback indicated in the following diagram:

$$\begin{tikzcd}
N \times_S V \arrow[r, "p_2"] \arrow[d, "p_1"'] & V \arrow[d, "s_V"] \\
N \arrow[r, "s_N"'] & \{+3\text{s}, -3\text{s}\}
\end{tikzcd} \tag{9}$$

where N is the set of nouns, V the set of verbs, $S = \{+3\text{s}, -3\text{s}\}$ is the set of attributes, and s_N and s_V are the morphisms mapping nouns and verbs to their singularity attribute (respectively), indicated as $+3$s (-3s) meaning is (not) third-person singular. Hence, quasi-systematicity for this domain is explained by an adjunction involving this pullback.

Our second linguistic example involves the difference between verbs *drench* and *throw*: English speakers say *I drenched the flowers with water*, but not *I drenched water onto the flowers*, whereas they say *I threw water onto the flowers*, but not *I threw the flowers with water* (Johnson, 2004). Whether or not the verb requires a preposition such as *onto*, or *over* is considered to depend on whether or not the meaning of the verb specifies how the water got onto the flowers (Johnson, 2004). Other verbs that require *onto* include: *dripped*, *throw*, *poured*, and *tossed*. Verbs that require no preposition include: *dampened*, *drenched*, and *wet*. The pullback for this situation is similar to the previous one, and indicated in the following diagram:

$$\begin{tikzcd}
V \times_A P \arrow[r, "p_2"] \arrow[d, "p_1"'] & P \arrow[d, "a_P"] \\
V \arrow[r, "a_V"'] & \{+, -\}
\end{tikzcd} \tag{10}$$

where V is the set of verbs, P the set of prepositions {onto, over, ε}, where ε indicates no preposition, $A = \{+, -\}$ is the set of attributes, and a_V and a_P are the morphisms mapping verbs and prepositions to their preposition attribute (respectively), indicated as $+$ ($-$) meaning does (not) require a preposition.

The explanations for quasi-systematicity for these linguistic examples follows the explanation for quasi-systematicity given for the natural relations examples, since they are all just instances of a pullback.

Discussion

A fundamental question for cognitive science concerns the nature of human cognitive architecture, i.e., what are the basic processes and modes of composition that together make up human cognitive behaviour. In so far as cognition is systematic, our category theory approach formally characterizes basic cognitive processes as functors, and mode of compositionality as adjunctions. Thus, functors and adjunctions constitute basic building blocks of human cognitive architecture.

There is common ground between our category theory explanation and the classical compositionality one. Both theories assume complex representations and processes that are built out of simpler ones, and some category theory constructions generalize classical ones (Phillips & Wilson, 2010). So, a classical theory of systematicity may be entirely compatible with our category-theory-level one.

Nonetheless, the quintessential difference between the two theories is the adjunction, which accounts for systematicity without having to stipulate a specific correspondence between processes for constructing representations and processes for accessing components of those constructions. Alan Turing is credited with providing the key advance concerning the foundations of cognitive science, overcoming the problems with associativism by suggesting that cognitive processes are instead (syntactic) computations (Fodor, 2008). Turing's (classical) solution works well for computational systems, because the correspondence between the processes for constructing compositional representations and the processes for accessing their constituents is systematicity maintained by the designer of the system. However, a theory of cognitive systems demands an explanation for such correspondences just in terms of the system and its interaction with the world, not some third party. Our explanation is of this latter sort, where the correspondence is enforced by the commutativity property of the adjunction.

This conception of adjunction as a building block of cognition is unique to our theory, and goes significantly beyond the widespread use of isomorphism (cf. analogy models) in cognitive science generally. A contrast of adjunction versus isomorphism highlights our shift in perspective: a reconception of cognitive architecture in terms of the relationships between structure-sensitive processes, instead of the representations that those processes transform (see also Phillips & Wilson, 2010). Other approaches to cognition, including classical ones typically treat representation in terms of an isomorphism between the representations and the entities those representations are intended to depict. From the category theory perspective, isomorphic domains modelled as categories are the same apart from a change of labels. An adjunction is more general, and potentially more useful, because two domains (involving quite different sorts of processes) that are not isomorphic, may still be systematically related by an adjunction, thereby affording an explanation that is not limited to cases whose domains are only superficially dissimilar.

Given the generally abstract nature of category theory, one may wonder whether our category-theoretic approach is to be regarded as a formal description (or, specification) of cognitive architecture, or a causal explanation. A full discussion of this point is beyond the scope of the current paper. We note, though, that category theory is generally regarded as a constructive theory is the sense that not only does one specify what are the relationships between (mathematical) structures, but how one structure is derived from another via the specified morphisms. The close relationship between category theory and computation has long been exploited by computer scientists. Functors are often used to model higher-order functions (i.e., functions that take, or return other functions). Thus, our category theory approach is well within the tradition of computational theories of mind, though our form of computation is distinctly categorical.

If adjunction is one of the basic components of human cognition, then what is its corresponding neural realization? An adjunction involves a reciprocal relationship between two functors, though the functors may not be inverses of each other. One possible approach to investigating neural correspondences, then, is with the reciprocal relationships between brain regions. For a category theory approach to neural networks, see Healy et al. (2009); and for modeling hippocampal place cells, see Gomez (2010). Further work is needed to establish the theoretical relationship between our category theory approach and neural mechanisms. We leave such theoretical and empirical possibilities as future research.

Acknowledgment. We thank the reviewers for their thoughtful comments. This work was supported by a JSPS Grant-in-aid (22300092) to the first author.

References

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

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

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. New York, NY: Oxford University Press.

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

Goguen, J. A. (1972). Realization is universal. *Theory of Computing Systems*, *6*(4), 359–374.

Gomez, J. (2010). *A new foundation for representation in cognitive and brain science: Category theory and the hippocampus*. Unpublished doctoral dissertation, Escuela Tecnica Superior de Ingenierus Industriales, Spain.

Healy, M. J., Olinger, R. D., Young, R. J., Taylor, S. E., Caudell, T. P., & Larson, K. W. (2009). Applying category theory to improve the performance of a neural architecture. *Neurocomputing*, *72*(13–17), 3158–3173.

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

Lawvere, F. W., & Schanuel, S. H. (1997). *Conceptual mathematics: A first introduction to categories*. Cambridge, UK: Cambridge University Press.

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

Magnan, F., & Reyes, G. E. (1995). Category theory as a conceptual tool in the study of cognition. In J. Macnamara & G. E. Reyes (Eds.), *The logical foundations of cognition* (pp. 57–90). New York: Oxford University Press.

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

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.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.