Title

Universal cognition in the context of resources and goals

Permalink

https://escholarship.org/uc/item/1g15f7zk

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Author

Phillips, Steven

Publication Date

2024

Peer reviewed

Universal cognition in the context of resources and goals

Steven Phillips (steven.phillips@aist.go.jp)
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, 305-8566, Japan

Abstract

The classical (symbol system) theory of cognition is supposed to explain systematicity—the coexistence of cognitive abilities. However, the classical theory does not explain why cognitive systems should be symbolic, nor why cognition sometimes fails to be systematic, so the symbol system assumption is seen by some as *ad hoc*: motivated only to fit the data. A mathematical theory is presented as a framework towards addressing these questions in terms of the available cognitive resources and the intended goals. A cognitive system is supposed to be resource-dependent and goal-driven. Accordingly, systematicity, or lack thereof follows from a universal construction principle (in a category theory sense) in this context—systems of symbols arise (or, fail to arise) as the "best" possible mapping given the available resources and the intended goal.

Keywords: systematicity, language of thought, symbol system, category theory, category, universal construction

Introduction

The classical (symbol system) view of cognition is motivated by a need to explain *systematicity* properties (among other things): why certain cognitive abilities coexist, e.g., having the ability to understand the expression *John loves Mary* and infer that *John* is the lover in this relationship coexists with having the ability to understand the expression *Mary loves John* and infer that *Mary* is the lover in that relationship (Fodor & Pylyshyn, 1988). Such situations are supported by systems of symbols so that the syntactic relations between constituent symbols are consistent with the semantic relations between corresponding constituent entities, i.e. a *language of thought* (LoT, Fodor, 1975). Structurally related abilities coexist when they are realized by the same structure-consistent processes. Causal relations among corresponding physical states are supposed to underwrite inferential relations among the symbolic representations—LoT is realized in the form of a *physical symbol system* (Newell, 1980), like a programming language by a computer, hence LoT is a representational/computational theory of mind (Wilson, 1999).

Although LoT was not intended to explain all forms of cognition, recent proponents argue that LoT-like properties in other (non-linguistic) forms of cognition show that LoT is still the "best game in town" for explanatory coverage (Quilty-Dunn, Porot, & Mandelbaum, 2023). However, classical theory does not explain why cognition should be symbolic, nor why cognition sometimes fails to be systematic, which some see as *ad hoc* (Aizawa, 2003)—classical theory assumes only "canonical" (symbolic) constructions that support evidential systematicity properties and none others (McLaughlin, 2009). Yet, such assumptions are characteristically *ad hoc* in being unconnected to the core principles of the theory and only motivated to fit the data (Aizawa, 2003), which was the same kind of problem raised against neural network theory (Fodor & Pylyshyn, 1988; Fodor & McLaughlin, 1990).

An alternative, *category theory* (Awodey, 2010; Leinster, 2014) approach says that systematicity is a consequence of *universal construction* (Phillips & Wilson, 2010). Category theory affords a (meta)mathematical framework for theories and models of cognition: a *category* (system) consists of a collection of entities called *objects* (states), and a collection of directed relations between objects called *arrows* (processes or relations between states), also called *morphisms* or *maps*, and an operation that combines arrows to form arrows called *composition* (of processes, etc.) that together satisfy certain laws. A category provides the context in which certain objects and arrows are *universal*, in a formal sense that all objects and arrows in that context are composed from the universal ones (Leinster, 2014). Universal constructions act like global optima in the given contexts, which affords a procedure for determining such constructions (Phillips & Wilson, 2016), in contrast to the ad hoc account of simply assuming them.

This approach does not explain systematic failures, however, beyond not obtaining a universal construction (Phillips & Wilson, 2016). In an experiment specifically designed to test the universal construction principle, some participants consistently failed to induce the common structure in a series of cue-target learning tasks (Phillips, Takeda, & Sugimoto, 2016). Post-test reports indicated that some participants did not notice any pattern across the task instances, in which case they did not show above chance prediction of targets for novel cues, in contrast to participants who did observe a common pattern. Another possibility is that participants did not deem it necessary to induce the common structure, i.e. responding on a task-by-task basis may have simply been their intention. We provide a category theory account of cognitive resources and goals to account for these possibilities, building upon the universal construction idea. Accordingly, (lack of) systematicity follows from a universal construction principle (in a category theory sense) in this context—systems of symbols arise (or, fail to arise) as the "best" possible mapping given the available resources and the intended goal.

Our departure from the universal construction explanation

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

for systematicity (Phillips & Wilson, 2010) is to couch the context in which a construction is universal in terms of cognitive resources and goals of the system. So, systematicity may not be present due to lack of resources or different goals. Lack of resources is a familiar way to characterize cognitive capacity in developmental and cognitive psychology, e.g. in the form of counting ability (Carey, 2009), working memory capacity (Cowan, 2001), or the complexity of relations that can be processed (Halford, Wilson, & Phillips, 1998). The basic theory is presented in the next section, with applications to cognition in the section that follows. Some general remarks on this approach are given in the last section.

Basic theory: universal constructions

The basic theory is found in many textbooks (see, e.g., Awodey, 2010; Leinster, 2014), including introductions for cognitive scientists (Phillips, 2022), scientists, generally (Spivak, 2014), and a more general audience (Lawvere & Schanuel, 2009). Yet, the multifaceted nature of category theory (Mac Lane, 1992) can obscure a useful interpretation for the application at hand. For this reason, the basic theory is presented here to help ground the interpretation in terms of universal constructions, cognitive resources and goals.

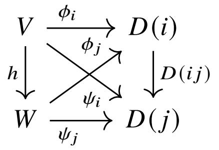
Remark 1. Some basic concepts are assumed. Compare:

- category to (directed) graph, as consisting of:
  - *objects* (cf. vertices),
  - arrows (cf. edges/paths)—directed from a *domain* object (cf. source) to a *codomain* object (cf. target)—with an *identity* (cf. zero-length path/loop) for each object and
  - a *composition operation* that combines pairs of arrows to form arrows (cf. concatenation of edges) satisfying:
    - *unity* (cf. concatenating loops) and
    - *associativity* (cf. order of path concatenations),
- functor to graph homomorphism, and
- natural transformation to homomorphism between graph homomorphisms.

Example 2 (sets)**.** **Set** is the category of sets (objects) and functions (arrows). The composition operation is function composition and identity arrows are identity functions.

Example 3 (diagram)**.** A *diagram* is a functor $D : J \to \mathbf{C}$ that acts like an index picking out a (sub)collection of objects and arrows in $\mathbf{C}$ where J is called the *shape* of D: e.g., diagram $(A, B) : 2 \to \mathbf{Set}$ references the (pair of) sets A and B.

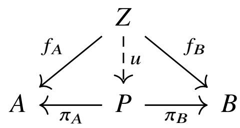
Example 4 (cone)**.** A *cone* to a diagram $D : J \to \mathbf{C}$ is a pair (V, ϕ) consisting of an object V in $\mathbf{C}$, called a *vertex*, and a family of arrows $\phi = \{\phi_j : V \to D_j\}_{j \in J}$, called the *legs* of the cone. The image of D is called the *base* of the cone. A *cone homomorphism* is an arrow $h : V \to W$ making diagram


(1)

commute for each object i and j and arrow ij in J. For instance, replace $D(i)$ and $D(j)$ with A and B, respectively, for a homomorphism of cones to a pair of objects. There is no arrow from A to B when $A \neq B$ in this situation.

Remark 5. A cone acts like a view (or, perspective), V, on a collection of objects/arrows in a field of attention, D, so a cone homomorphism acts like a change of perspective, $h : V \to W$.

Definition 6 (product)**.** In a category $\mathbf{C}$, a *product* of objects A and B is an object P (denoted $A \times B$) and an arrow $\pi = (\pi_A, \pi_B)$ such that for every object Z and arrow $f = (f_A, f_B)$ there exists a unique arrow u making the following diagram commute:


(2)

Example 7 (Cartesian product)**.** In **Set**, the product of A and B is *Cartesian product*: the set $A \times B = \{(a, b) | a \in A, b \in B\}$ and projections $\pi_A : (a, b) \mapsto a$ and $\pi_B : (a, b) \mapsto b$. The unique arrow is $\langle f_A, f_B \rangle : z \mapsto (f_A(z), f_B(z))$.

Remark 8. A product acts like an "extreme" view (limit) on the field of attention: e.g., the closest one can get to a pair of objects while still keeping both objects within the visual field.

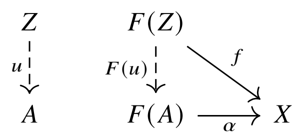
Definition 9 (limit)**.** A limit to a diagram $D : J \to \mathbf{C}$ is a cone (L, ϵ) such that for every cone (V, ϕ) there exists a unique cone homomorphism $u : V \to L$ such that $\phi = \epsilon \circ u$.

Remark 10. See diagram 1—replace h with u, etc.

Example 11 (product—limit)**.** A product is a limit (universal cone) to a pair-shaped diagram, $D : 2 \to \mathbf{C}$.

Remark 12. Products (and limits, generally) are instances of *universal constructions*, i.e. constructions given by a unique-existence condition, e.g., a limit is a universal cone.

Definition 13 (universal morphism)**.** A *universal morphism* from a functor $F : \mathbf{A} \to \mathbf{C}$ to an object X is a pair (A, α) consisting of an object A and an arrow $\alpha : F(A) \to X$ such that for every object Z and arrow $f : F(Z) \to X$ there *exists* a *unique* arrow $u : Z \to A$ making the following diagram commute:


(3)

α is called the *mediating* arrow.

Example 14 (bounds)**.** A partially ordered set $(P, \leq)$ is a category whose objects are the elements $p \in P$ and arrows are given by the partial order, i.e. there is an arrow $p \to p'$ whenever $p \leq p'$. The universal morphism from the inclusion $P \subseteq Q$ to an element $q \in Q$ is the *greatest lower bound* to q, i.e. the element $\overline{p} \in P$ such that $\overline{p} \leq q$ and $p \leq \overline{p}$ for all $p \in P$ (see diagram 3—replace X with q and A with $\overline{p}$, etc.).

Remark 15. $\overline{p}$ is the closest one can get to goal q with the resources made available by (the inclusion of) P.

Example 16 (limits)**.** A limit to a diagram $D : J \to \mathbf{C}$ is a universal morphism from the diagonal functor $\Delta : \mathbf{C} \to \mathbf{C}^J$ to D. For instance, a product of A and B is indicated by

$$\begin{array}{ccc} Z & & \Delta(Z) & & \\ {\scriptstyle \langle f_A, f_B\rangle}\downarrow & & {\scriptstyle \Delta\langle f_A, f_B\rangle}\downarrow & \searrow^{(f_A,f_B)} & \\ A \times B & & \Delta(A\times B) & \xrightarrow[(\pi_A,\pi_B)]{} & (A,B) \end{array} \tag{4}$$

where $\Delta : \mathbf{C} \mapsto \mathbf{C}^2$ acts like duplicate, i.e. $\Delta : A \mapsto (A, A)$.

Definition 17 (comma category)**.** A *comma category* $(F \downarrow G)$ is a category constructed from a pair of functors $F : \mathbf{A} \to \mathbf{C}$ and $G : \mathbf{B} \to \mathbf{C}$ with the same codomain that has for:

- objects the triples (A, B, γ) consisting of the objects A in $\mathbf{A}$ and B in $\mathbf{B}$ and the arrows $\gamma : F(A) \to G(B)$ in $\mathbf{C}$ and
- arrows the pairs (α, β) consisting of the arrows $\alpha : A \to A'$ in $\mathbf{A}$ and the arrows $\beta : B \to B'$ in $\mathbf{B}$ such that the diagrams

$$\begin{array}{ccc} F(A) & \xrightarrow{\gamma} & G(B) \\ {\scriptstyle F(\alpha)}\downarrow & & \downarrow{\scriptstyle G(\beta)} \\ F(A') & \xrightarrow[\gamma']{} & G(B') \end{array} \tag{5}$$

  commute, i.e. $G(\beta) \circ \gamma = \gamma' \circ F(\alpha)$.

Composition is pasting of squares; the identities are $(1_A, 1_B)$.

Example 18 (cones)**.** The category of cones to a J-shaped diagram D in $\mathbf{C}$ is the comma category $(\Delta \downarrow D)$.

Definition 19 (terminal object)**.** In a category $\mathbf{C}$, a *terminal object* is an object, denoted 1, such that for every object Z there exists a unique arrow $! : Z \to 1$, i.e. the imperative.

Example 20 (limit)**.** The limit to D is the terminal object in $(\Delta \downarrow D)$. See diagram 5—replace $G(\beta)$ with the identity 1_D, so the commutative square becomes a triangle (cf. diagram 4).

Example 21 (universal morphism)**.** The universal morphism from F to (object) X is terminal in $(F \downarrow X)$, where X denotes a *constant functor* sending all domain objects to the object X.

Remark 22. Comma categories, $(F \downarrow X)$, formalize cognition as a collection of resource-sensitive, goal-directed processes:

- each object specifies the resources $F(A)$ for obtaining the goal X via a path as the arrow $\gamma : F(A) \to X$,
- each arrow $\alpha : A \to A'$ specifies an alternative path via a change in resources as the composition $\gamma = \gamma' \circ F(\alpha)$ and
- the terminal in $(F \downarrow X)$ is the resource-limited path.

Example 23 (subgoal)**.** The universal morphism from the diagonal functor $\Delta : \mathbf{C} \to \mathbf{C}^2$ to subgoal $(A, 1)$ is indicated by

$$\begin{array}{ccccc} Z & & (Z,Z) & & \\ {\scriptstyle f}\downarrow & & {\scriptstyle (f,f)}\downarrow & \searrow^{(f,!)} & \\ A & & (A,A) & \xrightarrow[(1_A,!)]{} & (A,1) \end{array} \tag{6}$$

i.e. the object A and the arrow $(1_A, !)$ where 1 is the terminal, seen as a subgoal to (A, B) by ignoring the second target, B.

Remark 24. Categorical constructions generally have a *dual* form that is obtained by "arrow reversal" of the *primal* form: e.g., the dual of terminal is *initial*, i.e. an object, denoted 0, such that for every object Z there is a unique arrow $¡ : 0 \to Z$. The dual to limit is *colimit*: a universal *cocone*—the initial object in the category of cocones to a diagram D, i.e. $(D \downarrow \Delta)$.

Example 25 (delimited resource)**.** The universal morphism from functor $\langle 1_{\mathbf{C}}, 0\rangle : A \mapsto (A, 0)$ to pair (A, B) is given by

$$\begin{array}{ccccc} Z & & (Z,0) & & \\ {\scriptstyle f}\downarrow & & {\scriptstyle (f,1_0)}\downarrow & \searrow^{(f,¡)} & \\ A & & (A,0) & \xrightarrow[(1_A,¡)]{} & (A,B) \end{array} \tag{7}$$

where 0 is the initial object and $¡ : 0 \to B$ with $\langle 1_A, 0\rangle$ seen as delimiting (constraining) resources to a single instance of A.

Remark 26. This situation extends to other limits: e.g., the universal morphism from $\langle 1_{\mathbf{C}}, 1_{\mathbf{C}}, 0\rangle$ to (A, B, C) is the pair $(A \times B, \pi¡)$, where $\pi¡ = (\pi_A, \pi_B, ¡)$, i.e. a two-object resource for a three-object visual field (cf. remarks 5 and 8).

Remark 27. Reshaping may circumvent capacity limits: e.g., triple (A, B, C) as "pair" $((A, B), C)$. The universal morphism from $\langle \Delta, 1_{\mathbf{C}}\rangle : A \mapsto ((A, A), A)$ to $((A, B), C)$ consists of the object $(A \times B) \times C$ and the mediating arrow $((\pi_A, \pi_B), \pi_C)$.

Definition 28 (pullback)**.** In a category $\mathbf{C}$, a *pullback* of a pair of arrows $\phi_A : A \to C$ and $\phi_B : B \to C$ is an object P (denoted $A \times_C B$) and an arrow $\pi = (\pi_A, \pi_B)$ such that for every object Z and arrow $f = (f_A, f_B)$ there exists a unique u such that

$$\begin{array}{ccccc} & & Z & & \\ & {\scriptstyle f_A}\swarrow & \downarrow{\scriptstyle u} & \searrow{\scriptstyle f_B} & \\ A & \xleftarrow[\pi_A]{} & P & \xrightarrow[\pi_B]{} & B \\ & {\scriptstyle \phi_A}\searrow & & \swarrow{\scriptstyle \phi_B} & \\ & & C & & \end{array} \tag{8}$$

Example 29 (Constrained product)**.** In **Set**, the pullback of $\phi_A : A \to C$ and $\phi_B : B \to C$ is the pair $(a, b) \in A \times B$ such that $\phi_A(a) = \phi_B(b)$, i.e. the set $A \times_C B = \{(a, b) \in A \times B | \phi_A(a) = \phi_B(b)\}$, with projections π and $u = \langle f_A, f_B\rangle$ (see example 7). If C is the terminal object, hence ϕ_A and ϕ_B are the imperatives $!_A$ and $!_B$, then the pullback is the product of A and B.

Remark 30. A pullback is the *equalizer* of the pair of arrows $(\phi'_A, \phi'_B) : P \rightrightarrows C$, where $\phi'_A = \phi_A \circ \pi_A$, etc. (see diagram 8). As a limit, $J = (\cdot \rightrightarrows \cdot)$, an equalizer in **Set** acts like solutions to simultaneous equations. Finite limits arise from products and equalizers; or pullbacks and terminals (Leinster, 2014).

Cognition in context: resources and goals

Universal constructions are given with respect to a category (context) of some kind. The available cognitive resources and the intended goal are supposed to determine a context in which a construction may (or may not) be universal. This notion of context is formalized as a comma category $(F \downarrow G)$ where the functors F and G are the resources and goals, respectively. Symbol(-like propertie)s follow as universal constructions in this context and failures as constructions in the contexts of limited resources or differing goals. Three cases are studied: (1) compositionality of symbols, (2) complexity of relational processes and (3) intentionality of representations. Cognitive representations are cast in terms of *presheaves*, i.e. functors on a *topological space*, presented first.

Definition 31 (topological space)**.** A *topological space* is a pair (X,T) consisting of a set X and a set T (called the *topology* of X) of subsets U of X called the *open sets* of X such that the empty set and X are open sets, (finite) intersections and (arbitrary) unions of opens sets in T are open sets in T. The space or the topology is also simply denoted X.

Example 32 (discrete)**.** The power set $\mathcal{P}(X) = \{U | U \subseteq X\}$ is an instance of a *discrete topology* on X, i.e. every subset of X is an open set. For instance, suppose the set of colour-shape visual feature dimensions $X_{vis} = \{C, S\}$. The discrete topology on X_{vis} is the set $\{\emptyset, \{C\}, \{S\}, \{C,S\}\}$.

Definition 33 (presheaf)**.** A *presheaf* is a set-valued functor on a topological space, $\mathcal{F} : X^{\mathrm{op}} \to \mathbf{Set}$, sending each open set U of X to the set $\mathcal{F}(U)$ and each inclusion $V \subseteq U$ in X (i.e. $U \subseteq V$ in X^{op}) to the arrow $res_{V,U} : \mathcal{F}(U) \to \mathcal{F}(V)$, called a *restriction map*. The elements $s \in \mathcal{F}(U)$ are called the *sections* of U; an element of $\mathcal{F}(X)$ is called a *global section*.

Example 34 (table)**.** A presheaf on a discrete space acts like a relational table (Abramsky & Brandenburger, 2011): e.g., the table of colour-shape pairs $CS = \{(\bullet, \square), (\bullet, \triangle), (\bullet, \square)\}$ is identified with a presheaf on X_{vis} whose points are the attribute names and global sections are the rows. The restriction maps are the projections: e.g., $\pi_C : (\bullet, \square) \mapsto (\bullet)$.

Definition 35 (sheaf)**.** A *sheaf* is a "complete" presheaf, i.e. (conceptually) the sheaf can be reconstructed by pasting the sections on the open sets where they agree on overlaps.

Example 36 (complete)**.** The table/presheaf CS (example 34) completes to table/sheaf $CS^+ = \{(\bullet, \square), (\bullet, \triangle), (\bullet, \square), (\bullet, \triangle)\}$ by adding the row/global section $(\bullet, \triangle)$. As a relational table, the sheaf is constructed by joining the one-column tables for colour and shape along the empty column.

Remark 37. The sheaf condition is given by the equalizer of

$$\mathcal{F}(U) \longrightarrow \prod_i \mathcal{F}(U_i) \rightrightarrows \prod_{i,j} \mathcal{F}(U_i \cap U_j) \qquad (9)$$

(or, pullback) where i, j index the open sets (cf. diagram 8).

Remark 38. A *presheaf morphism* $\phi : \mathcal{F} \to \mathcal{G}$ is a natural transformation, i.e. as tables, a restriction of a mapped row is the same as the mapping of row restriction. The collections of presheaves and sheaves on a topological space X and their morphisms constitute categories, written $\mathbf{Psh}(X)$ and $\mathbf{Sh}(X)$.

Example 39 (sheafification)**.** The completion of a presheaf $\mathcal{F}$ to the "nearest" sheaf, denoted $\mathcal{F}^+$, is a universal construction, called *sheafification*, indicated by commutative diagram

$$\begin{array}{ccc} \mathcal{F} & \xrightarrow{sh} & \mathcal{F}^+ \\ & \searrow^{\phi} & \downarrow u \\ & & \mathcal{G} \end{array} \qquad \begin{array}{c} \mathcal{F}^+ \\ \downarrow u \\ \mathcal{G} \end{array} \qquad (10)$$

The completion of the colour-shape table (example 36) is an instance, i.e. replace $\mathcal{F}$ and $\mathcal{F}^+$ with CS and CS^+, respectively.

Remark 40. Sheafification is a dual form, i.e. a universal morphism $(\mathcal{F}^+, sh)$ from $\mathcal{F}$ to $\mathcal{I} : \mathbf{Sh}(X) \subseteq \mathbf{Psh}(X)$, which is initial in $(\mathcal{F} \downarrow \mathcal{I})$; equivalently, terminal in $(\mathcal{F} \downarrow \mathcal{I})^{\mathrm{op}}$.

Example 41 (subobject)**.** Presheaf CS acts like a *subtable* of CS^+: every row of CS is a row of CS^+. The category theory analogical abstraction of subset is *subobject*—a cornerstone of *topos theory* (Leinster, 2011; Goldblatt, 2006; Mac Lane & Moerdijk, 1992), linking geometry, algebra and logic. A *subfunctor* is a subobject in a category of functors, hence CS is a *subpresheaf* of CS^+, also written $CS \subseteq CS^+$. Subobjects constitute a partially order set, so CS^+ is the closest sheaf not smaller than CS—cf. the *ceiling function* sending each real number x to the smallest integer not less than x, e.g. $\lceil 2.4 \rceil = 3$: replace $\mathcal{F}$ and $\mathcal{F}^+$ with 2.4 and 3 (diagram 10), respectively.

Remarks 42. The sheafification construction belies two sides of universality, i.e. the unique-existence condition.

1. *Compatible presheaf* (existence): there exists at least one row for every pair of restrictions that agree on their overlaps.

2. *Separated presheaf* (uniqueness): there is at most one row that restricts to the given columns.

A sheaf is a compatible *and* separated presheaf. The second condition corresponds to identifying equivalent rows as the same row (or, put simply, as removing redundant rows)—cf. the *floor function* sending each real x to the largest integer not greater than x, e.g., $\lfloor 2.4 \rfloor = 2$: replace A and X with 2 and 2.4 (diagram 3), respectively. Conceptually, compatibility corresponds to filling out missing information and separability to removing unneeded details.

Canonical compositional symbols

Renewed interest in LoT stems in part from the appearance of symbol-like properties of non-linguistic cognitive abilities, e.g., visual cognition (Quilty-Dunn et al., 2023). Three symbol-like properties are: (1) *abstract conceptual content*, where perceptual instances are stored along with their abstracted concepts, (2) *discreteness*, where the representation of one constituent is not affected by the representation of another constituent, e.g., the constituent *red* in *red square* is the same as the constituent *red* in *red triangle*, and (3) *role-filler independence*, as in the representations of *John loves Mary* and *Mary loves John*, where *John* is the same filler in the *subject* or *object* roles, and *subject* is the same role when filled by *John* or *Mary*. These properties follow as universal constructions in resource/goal-appropriate contexts.

Example 43 (equivalence classes)**.** An *equivalence relation* (i.e. a reflexive, symmetric and transitive relation) $\sim$ on a set A determines an equivalence class $[a]_\sim = \{a' \in A | a' \sim a\}$. The smallest set of equivalence classes induced by $\sim$ is a universal construction (colimit), called *coequalizer*—dual to equalizer (see remark 30)—of the projections $(\sim) \rightrightarrows A$. For instance, a surjective (onto) function $f : X \to Y$ induces the equivalence relation $(\sim_f) = \{(x, x') | f(x) = f(x')\}$, hence the equivalence classes $X/{\sim_f}$ indicated by commutative diagram

$$\begin{array}{ccccc} (\sim_f) & \overset{\pi_1}{\underset{\pi_2}{\rightrightarrows}} & X & \xrightarrow{\kappa} & X/{\sim_f} \\ & & & \searrow^{\psi} & \Big\downarrow{\scriptstyle u} \\ & & & & Z \end{array} \tag{11}$$

i.e. the universal morphism $(X/{\sim_f}, \kappa)$ from $\pi = (\pi_1, \pi_2)$ to the diagonal functor $\Delta : \mathbf{Set} \to \mathbf{Set}^{\rightrightarrows}$, where κ assigns elements to their equivalence classes, i.e. the colimit (universal cocone) to the diagram $\pi : (\cdot \rightrightarrows \cdot) \to \mathbf{Set}$, which is initial in $(\pi \downarrow \Delta)$. For comparison, replace $\mathcal{F}$ and $\mathcal{F}^+$ (diagram 10, remark 40) with π and $X/{\sim_f}$, respectively.

Remark 44. Equivalence classes are basic to the construction of separated presheaves, hence sheaves (see remark 42)—the data on a point $x \in X$, called the *stalk* of $\mathcal{F}$ at x, denoted $\mathcal{F}_x$, is the set of equivalence classes, called *germs*, obtained by taking the colimit over open sets U containing x, i.e. two sections are *germ-equivalent* if they restrict to the same datum as U approaches $\{x\}$, so removing unneeded details.

Example 45 (symbols)**.** Suppose a character recognition map, $alph : \mathtt{A} \mapsto \mathrm{A}, \mathcal{A} \mapsto \mathrm{A}, \mathscr{A} \mapsto \mathrm{A}, \mathtt{B} \mapsto \mathrm{B}, \mathcal{B} \mapsto \mathrm{B}, \mathscr{B} \mapsto \mathrm{B}, \ldots,$ sending letters to their alphabetic characters. The coequalizer for the equivalence relation $\sim_{alph}$ (i.e. f in example 43) sends each letter to its character class. Each class can be replaced by a representative member (by a map u in diagram 11), which can be regarded as a symbol for its class.

Remark 46. Colimit $(X/{\sim_f}, \kappa)$ affords abstract conceptual content (Quilty-Dunn et al., 2023), whereby the perceptual instances are stored with their abstract concept. The unique-existence of a map u from a symbol to the concept it represents is guaranteed by the universal mapping property for colimits.

Definition 47 (section)**.** In a category $\mathbf{C}$, a *section* of an arrow $f : A \to B$ is an arrow $s : B \to A$ such that $f \circ s = 1_B$.

Example 48 (compositional symbols)**.** Compositionality of symbols follows from the fact that every function onto a set (space) X, i.e. $f : Y \to X$, induces a *sheaf of sections*, denoted $\Gamma(Y/X)$, that sends each open set U of X to the sections of f restricted to U, i.e. $\Gamma(Y/X) : U \mapsto \{s : U \to Y | f \circ s = 1_X\}$. So, the classification of visual features to their feature dimensions, e.g., $f_{dim} : \bullet \mapsto C, \bullet \mapsto C, \square \mapsto S, \triangle \mapsto S$, induces the sheaf of sections that is CS^+ (example 36).

Remark 49. A sheaf $\mathcal{F}$ on a space X is reconstructed from its stalks, $\mathcal{F}_x$. If X is discrete, then the sheaf is obtained from the product of its stalks, $\Pi_{x \in X} \mathcal{F}_x$, which embodies the discreteness property of symbol systems (Quilty-Dunn et al., 2023), i.e. each symbol is composable with other symbols without changing their identity. (NB. Every set can be given a discrete topology. Moreover, the discrete topology is terminal in the category of topologies on a set X ordered by inclusion.)

Example 50 (visual feature binding)**.** Other authors (Phillips, Takeda, & Singh, 2012) observed that visual feature binding can be formalized as a pullback of feature-location maps, e.g., colour and shape. The partial information encoded by these maps corresponds to a presheaf on a space whose points encode feature dimensions, e.g. $\{C, S, L\}$ for colour, shape and location, and open sets encode binding structure, i.e. $\{L\}, \{C, L\}$ and $\{S, L\}$, which says that colour and shape bind more closely to location. Complete information is recovered by sheafification yielding the binding of colour to shape as constrained by location (Phillips, 2020). For our purposes, the sheaf condition (diagram 9) involves a pullback that is the terminal in the comma category where the resource functor is the diagonal with shape $\vee = (\cdot \to \cdot \leftarrow \cdot)$, i.e. $\Delta : \mathbf{Set} \to \mathbf{Set}^{\vee}$, and goal functor is the diagram pointing to the location maps, e.g., $\pi_{C,S} = (\pi_C, \pi_S)$, where $\pi_C : (c, l) \mapsto l$ and $\pi_S : (s, l) \mapsto l$, i.e. $(\Delta \downarrow \pi_{C,S})$. The sheaf is terminal with respect to resource and goal functors iterated over all the open sets.

Example 51 (relations)**.** A relation R between sets A and B is a subset of their Cartesian product, $R \subseteq A \times B$, i.e. $(a, b) \in R$ if a is R-related to b, also written aRb. The roles of fillers a and b are implicitly encoded by their position in the (ordered) pair. For instance, *John loves Mary* corresponds to the pair $(John, Mary) \in Loves \subseteq Person \times Person$. Relations represented as sheaves explicitly encode roles as points of the topological space: e.g. the *loves* relation is given as a sheaf on $\{Subject, Object\}$ with the discrete topology consisting of the sections given as the pairs $(John, Mary)$ and $(Mary, John)$.

Remark 52. The sheaf condition ensures the kind of role-filler independence whereby, e.g., *John* can fill both subject and object roles and the subject (or, object) role can be filled by either *John* or *Mary* (Quilty-Dunn et al., 2023), as the sheaf is constructed from the (Cartesian) product of the set containing *John* and *Mary* with itself. In contrast, a presheaf need not represent all possible combinations.

Cognitive complexity of relations

The capacity to process relational information is a significant factor impacting cognitive ability (Halford et al., 1998): e.g., tasks involving (ternary) relations between three variables are more difficult than tasks involving (binary) relations between two variables. For instance, transitive inference—integrating two binary relations ($a \leq b$ and $b \leq c$) to construct a ternary relation ($a \leq b \leq c$) from which they make the inference (i.e. $a \leq c$)—is difficult for young children (around the age of five years) who are limited to at most binary relations (Halford, 1984; Andrews & Halford, 1998). Moreover, children who succeed vs. fail at transitive inference typically succeed vs. fail at other tasks where the relational complexity is ternary vs. binary, respectively, even though the goal of each task is understood (Andrews & Halford, 1998). A capacity to process binary and ternary relations corresponds to a capacity to compute binary and ternary products (Phillips, Wilson, & Halford, 2009). And so, systematic failures arise as a universal construction in the resource-limited comma category, where the resource functor pertains to (binary) products and the goal functor pertains to triples (remark 26).

Example 53 (cognitive complexity)**.** The sheaf theory view of relations (example 51) also affords a universal, resource-limited account of complexity, sketched here, in terms of presheaves on *simplicial complexes*—built from 0-simplexes (points), 1-simplexes (lines), 2-simplexes (triangles), etc. (Ghrist, 2014). The presheaf sends *(co)boundary maps*—connecting points to lines, lines to triangles and so on—to table projections. The premises are represented as a presheaf on a complex with three points (A, B and C), three lines (AB, BC and AC) and one triangle (ABC), where the table attached to AB (and BC) encodes the binary relations, e.g., as two-coloured block towers, *red atop green*, *green atop blue*, etc. (Andrews & Halford, 1998). Inference is obtained by sheafification: the empty table attached to ABC is completed by the pullback of the two projections onto the common boundary, i.e. $\pi_{B,AB}$ and $\pi_{B,BC}$, obtaining the sheaf with the table on ABC encoding ordered triples, e.g., *red atop green atop blue*, and projection onto AC corresponds to the inferred relations, *red atop blue*. The space without the triangle is a subcomplex with three points and three lines. Sheafification in this case is just the identity map, corresponding to the resource-limited systematic failure of younger children.

Remark 54. The pullback corresponds to joining the premise table with itself at the common values (Phillips et al., 2009). Sheafification is also seen as a from of generalization from the training trials as a presheaf morphism to the testing trials as a sheaf morphism and failure to generalize as the situation where the training presheaf is already a sheaf (Phillips, 2018a).

Intentional representations

Systematic failure to exhibit an intended ability may also be attributed to differing goals, as illustrated earlier (example 23). For instance, participants trained on a series of cue-target maps failed to induce the underlying product structure when the learning tasks involved a smaller number of cue pairs even though they learned the training trials to criterion (Phillips et al., 2016). In other words, they met the within-task learning (sub)goal, but not the larger goal of inducing the common task schema, which afforded target prediction on novel trials.

Remark 55. Presheaf morphisms in $\mathbf{Psh}(1) \cong \mathbf{Set}$ correspond to cue-target associations. $\mathbf{Psh}(1)$ is a subcategory of $\mathbf{Psh}(2)$, the category of presheaves on the two-point space affording product (relational) maps. So, systematic failure follows as a universal construction relative to the within-task subgoal.

Example 56 (associative learning)**.** Systematic learning of cue-target associations follows as a limit to a diagram that picks out the training set (cf. example 16, setting $J = 1$). The limit is any set isomorphic to the training set and the mediating arrow is the corresponding bijection. This apparently trivial example of a universal construction belies the universal nature of associative learning.

General remarks: universal cognition

The explanatory standard for systematicity is to say why this property (necessarily, not just possibly) follows from the core principles and assumptions of the theory (Aizawa, 2003; Fodor & Pylyshyn, 1988). Classical theory is problematic, because symbol systems may not support some systematicity properties, so assuming just those (canonical) systems that support systematicity is ad hoc (Aizawa, 2003). The category theory approach derives (absence of) systematicity from universal constructions given the available resources and intended goals. Incorporating resources and goals is closely related to making competence-performance distinctions (see Firestone, 2020). Yet, our point is that some failures are also systematic, demanding the same explanatory standard, that also follow from universal constructions. Goals and resources are not ad hoc assumptions for the categorical explanation, as these assumptions can be assessed independently of the main task, e.g., using self-test reports on participant's strategies—goals (Phillips et al., 2016) and pre-test trials to assess a capacity for binary relations—resources (Andrews & Halford, 1998). These assumptions are motivated by more than just the need to fit data—cognitive systems are naturally physical systems embedded in real environments, so are resource-sensitive.

Universality is supposed to explain systematic failures, yet duality (remark 24), in the form of *adjoint functors* (Leinster, 2014), is also supposed to reconcile opposing properties of cognition (Phillips, 2018b)—sheafification (example 39) is an adjoint functor. LoT, in particular, supposes a (physical) symbol system, but a theory based on symbols as discrete representational states (Quilty-Dunn et al., 2023) eventually must be reconciled with nondiscrete, continuous aspects of cognition. Certain functors to/from (pre)sheaf categories are adjoints and this universal-duality (adjointness) principle is supposed as a formal framework for reconciling symbolic vs. nonsymbolic cognition, including the symbol-like properties of abstract content, discreteness and role-filler independence mentioned earlier (Phillips, 2024).

Categories with subobjects (example 41) generalize the usual set-theoretic approaches based on discrete points to set-like constructions with continuity and variability (see, e.g., Lawvere & Rosebrugh, 2003) that yield deeper links between apparently disparate domains (see Rosiak, 2019, 2022, for an in-depth, philosophical analysis and discussion). Likewise, here, topos theory affords a more general notion of LoT, hence the slogan, *LoT is a topos* (Phillips, 2024). A next step for a universal theory of cognition is to marry a category/topos theory notion of LoT(s) with an empirically informed theory of resources and goals. This step and further exploration of the ideas briefly presented here are topics for future work.

Acknowledgements

This work was supported by a Japanese Society for the Promotion of Science Grant-in-aid (23K11797).

References

Abramsky, S., & Brandenburger, A. (2011). The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, *13*, 113036. doi: 10.1088/1367-2630/13/11/113036

Aizawa, K. (2003). *The systematicity arguments*. New York, NY: Kluwer Academic. doi: 10.1007/978-1-4615-0275-3

Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development*, *13*(4), 479–513. doi: 10.1016/S0885-2014(98)90004-1

Awodey, S. (2010). *Category theory* (2nd ed.). New York, NY: Oxford University Press. doi: 10.5555/2060081

Carey, S. (2009). *The origins of concepts*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195367638.001.0001

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. doi: 10.1017/S0140525X01003922

Firestone, C. (2020). Performance vs. competence in human-machine comparisons. *Proceedings of the National Academy of Sciences*, *117*(43), 26562–26571.

Fodor, J. A. (1975). *The language of thought*. New York, NY: Crowell.

Fodor, J. A., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, *35*(2), 183–204. doi: 10.1016/0010-0277(90)90014-b

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. doi: 10.1016/0010-0277(88)90031-5

Ghrist, R. (2014). *Elementary applied topology*. Createspace.

Goldblatt, R. (2006). *Topoi: The categorial analysis of logic* (Revised ed.). New York, NY: Dover Publications.

Halford, G. S. (1984). Can young children integrate premises in transitivity and serial order tasks? *Cognitive Psychology*, *16*(1), 65–93. doi: 10.1016/0010-0285(84)90004-5

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–831. doi: 10.1017/S0140525X98001769

Lawvere, F. W., & Rosebrugh, R. (2003). *Sets for mathematics*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511755460

Lawvere, F. W., & Schanuel, S. H. (2009). *Conceptual mathematics: A first introduction to categories*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511804199

Leinster, T. (2011, June). An informal introduction to topos theory. *arXiv preprint arXiv:1012.5647v3*. doi: 10.48550/arXiv.1012.5647

Leinster, T. (2014). *Basic category theory* (Vol. 143). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9781107360068

Mac Lane, S. (1992). The protean character of mathematics. In J. Echieverria, A. Ibarra, & T. Mormann (Eds.), *The Space of Mathematics* (pp. 3–13). Berlin: de Gruyter.

Mac Lane, S., & Moerdijk, I. (1992). *Sheaves in geometry and logic: A first introduction to topos theory*. New York, NY: Springer. doi: 10.1007/978-1-4612-0927-0

McLaughlin, B. P. (2009). Systematicity redux. *Synthese*, *170*(2), 251–274. doi: 10.1007/s11229-009-9582-0

Newell, A. (1980). Physical symbol systems. *Cognitive science*, *4*(2), 135–183. doi: 10.1016/S0364-0213(80)80015-2

Phillips, S. (2018a). Going beyond the data as the patching (sheaving) of local knowledge. *Frontiers in Psychology*, *9*, 1926. doi: 10.3389/fpsyg.2018.01926

Phillips, S. (2018b, July). What underlies dual-process cognition? Adjoint and representable functors. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual meeting of the cognitive science society* (pp. 2250–2255). Austin, TX: Cognitive Science Society.

Phillips, S. (2020). Sheaving—a universal construction for semantic compositionality. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20190303. doi: 10.1098/rstb.2019.0303

Phillips, S. (2022). What is category theory to cognitive science? Compositional representation and comparison. *Frontiers in Psychology*, *13*, 1048975. doi: 10.3389/fpsyg.2022.1048975

Phillips, S. (2024). A category theory perspective on the Language of Thought: LoT is universal. *Frontiers in Psychology*, *15*, 1361580. doi: 10.3389/fpsyg.2024.1361580

Phillips, S., Takeda, Y., & Singh, A. (2012). Visual feature integration indicated by phase-locked frontal-parietal EEG signals. *PLoS One*, *7*(3), e32502. doi: 10.1371/journal.pone.0032502

Phillips, S., Takeda, Y., & Sugimoto, F. (2016). Why are there failures of systematicity? The empirical costs and benefits of inducing universal constructions. *Frontiers in Psychology*, *7*, 1310. doi: 10.3389/fpsyg.2016.01310

Phillips, S., & Wilson, W. H. (2010). Categorial compositionality: A category theory explanation for the systematicity of human cognition. *PLoS Computational Biology*, *6*(7), e1000858. doi: 10.1371/journal.pcbi.1000858

Phillips, S., & Wilson, W. H. (2016). Systematicity and a categorical theory of cognitive architecture: Universal construction in context. *Frontiers in Psychology*, *7*, 1139. doi: 10.3389/fpsyg.2016.01139

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. doi: 10.1371/journal.pcbi.1000599

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The re-emergence of the Language of Thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*, *46*, e261. doi: 10.1017/S0140525X22002849

Rosiak, D. (2019). *Towards a classification of continuity and on the emergence of generality*. Unpublished doctoral dissertation, DePaul University, Chicago,IL.

Rosiak, D. (2022). *Sheaf theory through examples*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/12581.001.0001

Spivak, D. I. (2014). *Category theory for the sciences*. Cambridge, MA: MIT Press. doi: 10.5555/2628001

Wilson, R. A. (1999). Philosophy. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. xv–xxxvii). Cambridge, MA: MIT Press.