What Do Transitive Inference and Class Inclusion Have in Common? Categorical (Co)Products and Cognitive Development

Steven Phillips[1]*, William H. Wilson[2], Graeme S. Halford[3]

1 Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan, **2** School of Computer Science and Engineering, The University of New South Wales, Sydney, New South Wales, Australia, **3** School of Psychology, Griffith University, Brisbane, Queensland, Australia

Abstract

Transitive inference, class inclusion and a variety of other inferential abilities have strikingly similar developmental profiles— all are acquired around the age of five. Yet, little is known about the reasons for this correspondence. Category theory was invented as a formal means of establishing commonalities between various mathematical structures. We use category theory to show that transitive inference and class inclusion involve *dual* mathematical structures, called product and coproduct. Other inferential tasks with similar developmental profiles, including matrix completion, cardinality, dimensional changed card sorting, balance-scale (weight-distance integration), and Theory of Mind also involve these structures. By contrast, (co)products are not involved in the behaviours exhibited by younger children on these tasks, or simplified versions that are within their ability. These results point to a fundamental cognitive principle under development during childhood that is the capacity to compute (co)products in the categorical sense.

Introduction

Children acquire various reasoning skills over remarkably similar periods of development. Transitive Inference and Class Inclusion are two behaviours among a suite of inferential abilities that have strikingly similar developmental profiles—all are acquired around the age of five years [1]. For example, older children can infer that if *John is taller than Mary*, and *Mary is taller than Sue*, then *John is taller than Sue*. This form of reasoning is called Transitive Inference. Older children also understand that a grocery store will contain more fruit than apples. That is, the number of items belonging to the superclass is greater than the number of items in any one of its subclasses. This form of reasoning is called Class Inclusion. These two types of inference appear to have little in common. Transitive Inference typically involves physical relationships between objects, while Class Inclusion involves abstract relative sizes of object classes. Nonetheless, explicit tests of these and other inferences for a range of age groups revealed that success was attained from about the median age of five years [1].

Since Piaget, decades of research have revealed important clues regarding the development of inference, yet little is known about the reasons underlying these correspondences (see [2] for reviews). A common theme in two recent proposals is the computing of relational information [3,4]. In regard to *Relational Complexity* theory [3], the correspondence between commonly acquired cognitive behaviours is based on the maximum arity of relations that must be processed (e.g., tasks acquired after age five involve ternary relations, i.e., relations between three items). In regard to

Cognitive Complexity and Control theory [4], the correspondence is based on the common depth of relation hierarchies. Although a relational approach to cognitive behaviour has a formal basis in relational algebra [5], certain assumptions must be made about the units of analysis. For tasks as diverse in procedure and content as Transitive Inference and Class Inclusion, it is difficult to see how the analysis of one task leads naturally to the other. For Relational Complexity theory, Transitive Inference is considered to involve the integration of *two* binary relations between task elements into an ordered triple, or ternary relation; whereas Class Inclusion is regarded as the integration of *three* binary relations between three sets of elements (one complement and two containments) into a ternary relation [2,3]. For Cognitive Complexity and Control theory, Transitive Inference involves relations over *items*; whereas Class Inclusion involves relations over *sets of items*.

This theoretical difficulty is symptomatic of the general problem in cognitive science where the basic components of cognition are unknown. In the absence of such detailed knowledge, cognitive modelers have been forced to assume a particular representational format (e.g., symbolic [6], or subsymbolic [7]). This approach, however, does not lend itself to the current problem, because the elements of Transitive Inference and Class Inclusion tasks (i.e., objects and classes of objects) do not share a common basis. Understandably, then, these sorts of behaviours have tended to be studied in detailed isolation, narrowing the scope for identifying general principles.

Category theory was born out of a desire to establish formal commonalities between various mathematical structures [8,9], and has since been applied to the analysis of computational structures in computer science (see [10–12]). The seminal insight was a shift

Author Summary

Children acquire various reasoning skills during a remarkably similar period of development. Yet, the reasons for these similarities are a mystery. Two examples are Transitive Inference and Class Inclusion, which develop around five years of age. Older children understand that if John is taller than Mary, and Mary is taller than Sue, then John is also taller than Sue. This form of reasoning is called transitive inference. Older children also understand that there are more fruits than apples. This inference is called class inclusion. We explain why these and a variety of other abilities show the same development using a branch of mathematics called category theory. Category theory reveals that they have related underlying structure. So, despite their apparent superficial differences these reasoning abilities have similar profiles of development because they involve related sorts of processes.

from objects as the primary focus of analysis to their transformations. Contrast, for instance, sets defined in terms of (the properties of) the objects they contain—Set Theory—against sets defined in terms of the morphisms that map to or from them—Category Theory [13]. This insight motivates our categorical approach to the analysis of inference, and our way around the current impasse. In cognitive science, several authors have used category theory for a conceptual analysis of space and time [14–16], though we know of only one other application that has modeled empirical data [17]. Since our application of category theory to cognitive behaviour is novel, we first introduce the basic category theory constructs needed for our subsequent analysis of Transitive Inference, Class Inclusion, and other paradigms. The analysis begins with a brief introduction of the sort of data our approach is intended to explain, which primarily concerns contrasts between younger and older children relative to age five, and correlations across paradigms. Finally, we extend our categorical approach to more complex levels of inference. Our main point is that, despite the apparent lack of resemblance, all these tasks are formally connected via the categorical (co)product, to be defined below. The significance of this result is that it opens the door to an entirely new (empirical) approach to identifying general principles, particularly in regard to the development of inferential abilities, that are less likely to be revealed by standard modeling methods.

Methods

In this section, we provide the basic category theory definitions and constructs used in our subsequent analysis of various inferential abilities. Detailed introductions to category theory are found in [12,18,19]. Category theory is abstract in the sense that its entities may not refer to particular concrete objects, such as system states, or task stimuli, or even other mathematical objects. This sort of abstractness is a strength of the theory, permitting one to see the formal connection between otherwise disparate fields. Nonetheless, this abstractness may also be a source of bewilderment to those unfamiliar with this approach. Hence, throughout this section, we provide specific examples of category theory concepts for didactic purposes in some cases, and as a prelude to our analysis in others.

Category

A *category* **C** consists of:

- a class |**C**| of objects $(A, B, \ldots)$;

- a set $\mathbf{C}(A, B)$ of morphisms (also called arrows, or maps), from A to B, where for each morphism $f : A \to B$ indicates that A is the *domain*, or *source* and B is the *codomain*, or *target* of f (i.e., $A = \mathrm{dom}(f)$ and $B = \mathrm{cod}(f)$);

- a morphism $1_A : A \to A$ called the *identity* for each object A; and

- a composition operation, denoted "$\circ$", of morphisms $f : A \to B$ and $g : B \to C$, written $g \circ f : A \to C$, satisfying 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 $h : C \to D$.

One immediately recognizable example is the category **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 "$\circ$". Another, less obvious, example is the category of Euclidean spaces, **Euc**, which has R^n spaces as objects, where n is a natural number; $m \times n$ matrices for morphisms $f : R^m \to R^n$, where the identity matrix is the identity morphism; and matrix multiplication is the composition operation. From a cognitive perspective, an object may be a cognitive state, set of states, or some other entity employing symbolic, or numerical representations, and a morphism may be some cognitive process transforming one state to another. At present, we do not prejudge the cognitive nature of objects and morphisms for the reasons already mentioned.

Categories exist for a diverse range of structures, with objects more complex than sets of elements, and structure-preserving morphisms more complex than associations. For example, the following morphism

$$(\Re, +) \xrightarrow{exp} (\Re, \times) \tag{1}$$

maps from object $(\Re, +)$ to object $(\Re, \times)$, where each object consists of the set of real numbers with additional internal structure (i.e., a rule—respectively, addition and multiplication—for combining two numbers into another number). This morphism maps real numbers x to e^x, where "+" in the domain corresponds to "$\times$" in the codomain. Structure is preserved by exp, because the transformation of the result of applying the rule to the numbers is the same as the result of applying the corresponding rule to the transformed numbers. In this case, $e^{x+y} = e^x \times e^y$, for all $x, y \in \Re$. This morphism and its (co)domain objects are members of the category of semigroups, which has semigroups for objects and semigroup homomorphisms for arrows. A semigroup is just a set S with an associative binary operation $*$, and a semigroup homomorphism, $f : (S, *) \to (T, \#)$, preserves an object's internal structure as illustrated: that is, $f(s_1 * s_2) = f(s_1) \# f(s_2)$. Hence, $ıeıxıp$ is a semigroup homomorphism. But, not every function is a morphism in this category. For example, inc (i.e., increment by 1)

$$(\Re, +) \xrightarrow{inc} (\Re, +) \tag{2}$$

is not a semigroup homomorphism, because $1 + (x + y) \neq (1 + x) + (1 + y)$. These examples illustrate that although category theory is abstract, it is not arbitrary. So, statements derived from the theory are, in principle, testable and falsifiable.

Dual

We need to introduce the notion of the *dual* $\mathbf{C^{op}}$ of a category **C**. $\mathbf{C^{op}}$ is essentially **C** with all arrows reversed. That is, the set of

objects in $\mathbf{C}^{\mathbf{op}}$ is the same as in $\mathbf{C}$; there is a one-to-one correspondence between arrows in $\mathbf{C}$ and $\mathbf{C}^{\mathbf{op}}$ such that the arrow $f : A \rightarrow B$ in $\mathbf{C}$ corresponds to the arrow $\overline{f} : B \rightarrow A$ in $\mathbf{C}^{\mathbf{op}}$; and the composition $\overline{f} \circ \overline{g} = \overline{g \circ f} : C \rightarrow A$ is defined in $\mathbf{C}^{\mathbf{op}}$ exactly when $g \circ f : A \rightarrow C$ is defined in $\mathbf{C}$. A definition or a proposition and proof about $\mathbf{C}$ gives rise to its dual in $\mathbf{C}^{\mathbf{op}}$ by taking the original definition/proposition/proof and reversing all the arrows. Any valid argument about the arrows in $\mathbf{C}$ is valid for the dual argument in $\mathbf{C}^{\mathbf{op}}$, so proving or defining something in $\mathbf{C}$ gives you something for free in $\mathbf{C}^{\mathbf{op}}$. Obviously, reversing an arrow twice will return the original arrow, so $\mathbf{C}^{\mathbf{op}^{\mathbf{op}}} = \mathbf{C}$.

Some examples of duals involve certain types of morphisms, called epimorphisms, monomorphisms and isomorphisms. A morphism $f : A \rightarrow B$ is an *epimorphism*, if for any pair of morphisms $i, j : B \rightarrow D$, $i \circ f = j \circ f$ implies $i = j$. That is, f is an epimorphism if whenever the following diagram *commutes*, $i = j$,

$$A \xrightarrow{f} B \underset{j}{\overset{i}{\rightrightarrows}} D \tag{3}$$

(Commutative diagrams afford proof by *arrow chasing*. A diagram is said to be commutative if the compositions of the morphisms on any two paths through the diagram, from a common start object to a common finish object, are equal, except when both paths are of length 1. In Diagram 3, the start object is A, the finish object is D, and the two paths are f, i and f, j.) For **Set** and **Euc**, a morphism f is an epimorphism if and only if it is onto (i.e., informally, there are no elements in the codomain that are unreachable from elements in the domain via f). A morphism $f : A \rightarrow B$ is a *monomorphism*, if for any pair of morphisms $g, h : C \rightarrow A$, $f \circ g = f \circ h$ implies $g = h$,

$$C \underset{h}{\overset{g}{\rightrightarrows}} A \xrightarrow{f} B \tag{4}$$

For **Set** and **Euc**, a morphism f is a monomorphism if and only if it is one-to-one. By reversing arrows, we see that the definition of an epimorphism in a category $\mathbf{C}$ is the definition of a monomorphism in the category $\mathbf{C}^{\mathbf{op}}$ (i.e., the definitions are dual). A morphism $f : A \rightarrow B$ is an *isomorphism* if there exists a morphism $g : B \rightarrow A$, such that $g \circ f = 1_A$ and $f \circ g = 1_B$. An isomorphism in $\mathbf{C}$ is also an isomorphism in $\mathbf{C}^{\mathbf{op}}$. A more formal treatment of duality can be found in Text S1.

Products

Cognitive behavior generally involves some means of integrating information. A general notion of integration is the categorical product. In any category $\mathbf{C}$, a *product* of two objects A and B is an object P together with two morphisms $p_1 : P \rightarrow A$ and $p_2 : P \rightarrow B$, such that for any pair of morphisms $z_1 : Z \rightarrow A$ and $z_2 : Z \rightarrow B$, there is a unique morphism $u : Z \rightarrow P$, such that the following diagram commutes

$$\tag{5}$$

where a broken arrow indicates that there exists exactly one morphism making the diagram commute. The morphisms used in the definition of a product (p_1 and p_2 above) 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 $p'_1 : P' \rightarrow A$ and $p'_2 : P' \rightarrow B$ there is one and only one isomorphism between P and P' that makes a diagram like the one above commute. This means that P is not unique, only unique with respect to another product object via isomorphism (a point to which we will return shortly), which is why the definition refers to *a* product, not *the* product. An essential characteristic of a product object is that the constituents A and B are retrievable via the projection morphisms. P is also written $A \times B$, and 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

$$\tag{6}$$

In **Set**, P is, up to isomorphism, the Cartesian product $A \times B$, $p_1 : A \times B \rightarrow A$, $p_2 : A \times B \rightarrow B$, 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 product function $\langle z_1, z_2 \rangle : Z \rightarrow A \times B$, 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 $f : a \mapsto b$.) For example, suppose $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$, then $P = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$, and $p_1(a_1, b_1) = a_1$, $p_2(a_1, b_1) = b_1$, and so on. Suppose $Z = \{1\}$, $z_1 : 1 \mapsto a_1$, and $z_2 : 1 \mapsto b_1$, then the only morphism u making this example commute is $u = \langle z_1, z_2 \rangle : 1 \mapsto (a_1, b_1)$.

One can think of tasks involving stimuli that vary along two task-relevant dimensions as examples involving categorical products. For example, classification tasks where the rule is based on, say, stimulus colour and size involves a product, with the set of task stimuli as the product object and the determination of colour and size features as the projection morphisms. Conservation tasks, for example, predicting whether the amount of liquid in one container is the same as another where the containers vary in, say, height and width also involve products. In this case, the product object is a set of volumes and the projection maps recover the associated heights and widths. We will see further examples of tasks involving products in the next section.

For our purposes, the categorical product (P, p_1, p_2) is a statement about a cognitive (sub)system, whereas the triple (Z, z_1, z_2) is a constraint on what constitutes a valid product rather than a specific claim about cognition. Notice that (Z, z_1, z_2) is not necessarily a product in its own right, as the example above illustrated, since the one element set $Z = \{1\}$ is not isomorphic to the Cartesian product containing four elements. Notice, further, that although (P, p_1, p_2) does pertain to the cognitive system it is not a commitment to a particular representation and process. To illustrate, the product object in the previous example could just as easily be defined as the set $P' = \{x_1, x_2, x_3, x_4\}$, without explicitly identifying the components a_i and b_j, so long as the projections p_1 and p_2 recover those components appropriately. From the categorical perspective, these two isomorphic alternatives are the "same", relieving us of any prior commitment to, say, classical [20] or functional [21] compositionality, which has been a contentious issue when framing theories of cognition [22].

Coproducts (sums)

A related notion of information integration is the categorical coproduct. In any category $\mathbf{C}$, a *coproduct* (or, *sum*) of two objects A

and B is an object Q together with two morphisms $q_1 : A{\to}Q$ and $q_2 : B{\to}Q$, such that for any pair of morphisms $z_1 : A{\to}Z$ and $z_2 : B{\to}Z$, there is a unique morphism $u : Q{\to}Z$, such that the following diagram commutes

$$A \xrightarrow{q_1} Q \xleftarrow{q_2} B \qquad\qquad (7)$$

(with z_1, u, z_2 mapping to Z)

The morphisms used in the definition of a coproduct (q_1 and q_2 above) are sometimes called injection morphisms. A coproduct object Q is also unique up to a unique isomorphism. Q is also written $A+B$, and since u is uniquely determined by z_1 and z_2, u is often written as $[z_1,z_2]$, and so the coproduct diagram becomes

$$A \xrightarrow{q_1} A+B \xleftarrow{q_2} B \qquad\qquad (8)$$

(with z_1, $[z_1,z_2]$, z_2 mapping to Z)

In **Set**, Q is the *disjoint union* $A \uplus B$, $q_1 : A{\to}A \uplus B$, and $q_2 : B{\to}A \uplus B$. Suppose, for example, $A=\{a,b\}$ and $B=\{x,y\}$, then $Q=\{(1,a),(1,b),(2,x),(2,y)\}$. Basically, every element in A and B is augmented with a label that identifies the set from which it came. Unlike set union, which removes duplicates, all information is maintained.

If we reverse all the arrows in the definition of a coproduct we get a product. A product in a category **C** is a coproduct in $\mathbf{C^{op}}$. Coproducts are dual to products. The duality between product and coproduct is shown formally in Text S1.

One way to think about coproducts in terms of cognitive tasks is to regard the label as the context or condition under which a stimulus is associated with a particular action. Experimental paradigms designed to assess cognitive flexibility, such as the Wisconsin Card Sorting Task, are examples. For instance, in one context, say, a reward schedule based on colour, a red triangle may require one type of response, but for a reward schedule based on shape, the red triangle requires a different type of response. In this case, the coproduct object is the disjoint union of the stimulus set with itself with colour and shape as labels, and the response is determined by a map from the coproduct object to a set of actions.

Pullbacks and pushouts

More generally, information integration is often subject to satisfying some constraint. Hence, product and coproduct are instances of more general constructs known as pullbacks and pushouts, respectively. A *pullback* of morphisms $f : A{\to}C$ and $g : B{\to}C$ is an object P and a pair of morphisms $p_1 : P{\to}A$ and $p_2 : P{\to}B$ satisfying $f{\circ}p_1=g{\circ}p_2$, such that for any pair of morphisms $z_1 : Z{\to}A$ and $z_2 : Z{\to}B$ such that $f{\circ}z_1=g{\circ}z_2$, there is a unique morphism $u : Z{\to}P$, such that the following diagram commutes:

$$\qquad\qquad (9)$$

(diagram with Z, P, A, B, C and morphisms z_1, z_2, u, p_1, p_2, f, g)

In such a pullback, P may also be denoted by $A \times {}_C B$. The constraint is contained in the requirement that the square in Diagram 9 should commute.

Intersection is an example of pullback in **Set**, where $P=A{\bigcap}B$, $C=A{\bigcup}B$, and p_1, p_2, f and g are inclusions. More generally, a pullback is a constrained product, restricted to satisfy the constraints imposed by f and g, so that $f{\circ}p_1=g{\circ}p_2$, with P as the set of solutions.

Pushout is dual to pullback. A *pushout* of morphisms $f : C{\to}A$ and $g : C{\to}B$ is an object Q and a pair of morphisms $q_1 : A{\to}Q$ and $q_2 : B{\to}Q$ satisfying $q_1 f=q_2 g$, such that for any pair of morphisms $z_1 : A{\to}Z$ and $z_2 : B{\to}Z$ such that $z_1 f=z_2 g$, there is a unique morphism $u : Q{\to}Z$, such that the following diagram commutes:

$$\qquad\qquad (10)$$

(diagram with C, B, A, Q, Z and morphisms g, f, q_1, q_2, z_1, z_2, u)

In such a pushout, Q may also be denoted by $A + {}_C B$.

Given the duality, union is an example of pushout in **Set**, where $Q=A{\bigcup}B$, $C=A{\bigcap}B$, and all morphisms are inclusions. In this case, the pushout is also a pullback. A more general pushout in **Set** involves a form of disjoint union such that elements $a_i=f(c_k)$ and $b_j=g(c_k)$ are *identified* (i.e., "glued" together) in the pushout object [9,23]. For example, suppose $C=\{c\}$, $f : c{\mapsto}a_1$ and $g : c{\mapsto}b_2$, and $A=\{(a_1,a_2)\}$ and $B=\{(b_2,b_3)\}$, then $A \uplus_C B=\{(1,a_1),(1,a_2),(2,b_3)\}$ or, equivalently, $\{(1,a_1),(2,b_2),(2,b_3)\}$. Here $(1,a_2)=(2,b_2)$, because they have been identified as described above. (These elements are actually equivalence classes whose members are identified by *coequalizers* [24], but we use this form for convenience.) In general, Q is the integration of components providing no more and no less information than necessary to satisfy the requirement that $q_1{\circ}f=q_2{\circ}g$.

For our purposes, the commutative squares in the pullback and pushout diagrams pertain to statements about cognitive (sub)systems, and (Z,z_1,z_2) constrains what constitutes a valid pullback/pushout construction. Aside from definitions, then, we no longer refer to Z and associated morphisms z_1, z_2, and u, so they are omitted from subsequent diagrams.

Initial and terminal objects and their (co)products

An *initial* object in a category **C** is an object 0, such that for every object A there is exactly one morphism $f : 0{\to}A$ in **C**. A *terminal* object is an object 1, such that for every object A there exists a unique morphism $g : A{\to}1$ in **C**. In **Set**, the only initial object is the empty set $\varnothing$, and any one-element set, e.g., $\{a\}$, is a terminal object. (In **Set**, the initial object, 0, has 0 members, while the terminal object, 1, has 1 member. In some other categories, e.g. **Euc**, the initial object is also the terminal object, and is then called a null object. Some categories, such as a discrete category with no non-identity arrows, lack an initial object, or a terminal object, or both.) Multiple initial objects in a category are not distinguished because they are isomorphic, and the same also applies to terminal objects [18,24]. In a category with initial and terminal objects, products and coproducts, a product of an object A with a terminal object is isomorphic to A, $A \times 1 \cong A$; and a coproduct of A with an initial object is

isomorphic to A, $A + 0 \cong A$ [18,24]. For example, the Cartesian product $\{a,b,c\} \times \{1\} = \{(a,1),(b,1),(c,1)\} \cong \{a,b,c\}$; and the disjoint union $\{a,b,c\} \uplus \varnothing = \{(1,a),(1,b),(1,c)\} \cong \{a,b,c\}$. The following diagram

$$
\begin{array}{ccc}
A \times B & \xrightarrow{\ p_2\ } & B \\
{\scriptstyle p_1} \downarrow & & \downarrow {\scriptstyle g} \\
A & \xrightarrow{\ f\ } & 1
\end{array}
\qquad (11)
$$

involving a terminal object (1) is always a pullback, and when $B = 1$, p_1 is an isomorphism. The following diagram

$$
\begin{array}{ccc}
0 & \xrightarrow{\ g\ } & B \\
{\scriptstyle f} \downarrow & & \downarrow {\scriptstyle q_2} \\
A & \xrightarrow{\ q_1\ } & A + B
\end{array}
\qquad (12)
$$

involving an initial object (0) is always a pushout, and when $B = 0$, q_1 is an isomorphism. For subsequent pullback/pushout diagrams, we omit references to specific morphisms with an initial object as its domain, or a terminal object as its codomain, since their existence is guaranteed by definition. Usually, the initial or terminal object is also omitted in these cases, but we choose to show it for conformity with the other diagrams.

These "special" cases are important for determining whether a system that apparently involves a (co)product is in fact isomorphic to one that does not. We will see an example of this situation in the next section. In these situations, we say that task difficulty is related to the simpler, non-(co)product form.

Notice that we could have explained all this just in terms of the particular product, coproduct, pullback, pushout, initial and terminal object that prevail in **Set**. Presenting in the more general case of products, etc., in an arbitrary category, makes it clearer that these are not constructions specific to **Set**, but instances of a wider phenomenon.

Results

In this section, we apply category theory concepts to the analysis of results from several studies that have provided empirical evidence of within group similarities and between group differences in behavioural performance across multiple tasks. The objective is to identify a formal basis for an equivalence class of tasks that accounts for these similarities and differences. Our seed paradigms are Transitive Inference and Class Inclusion, which were tested on age groups ranging from three to eight years [1]. The main finding was that significant above chance performances on Transitive Inference and Class Inclusion were observed around five years. By contrast, younger children (three- and four-year-olds) showed significant above chance performance only on simpler versions of these two tasks. The thrust of our analysis is to show that these tasks are formally connected by the dual relationship between product (Transitive Inference) and coproduct (Class Inclusion). By contrast, the simpler versions of these tasks do not involve a (co)product, or involve a (co)product in the trivial sense (e.g., $A \times 1 \cong A$). Then, in the remainder of this section, we extend this analysis to other paradigms, including: Matrix Completion, Cardinality, Card Sorting, Balance-scale, and Theory of Mind. In each case, the more difficult version of the task, where significant above chance performance was observed in the five-year-olds and older children, involves a product or coproduct. By contrast, performance in the younger age groups

does not, or involves a (co)product with an initial or terminal object, which reduces via isomorphism to a single map. Thus, the core characteristic that distinguishes performance by younger (three-, or four-year-olds) versus older (five-year-olds and above) children is computing the categorical (co)product and its encompassing pullback (pushout).

The data of primary concern here are the correlations in achievement across paradigms and the significant differences between age groups within paradigms. Age five is regarded as a "nominal" timepoint in that some children exhibit success at a younger or older age. For example, 11% of the three- and four-year-olds, and 71% of six-year-olds succeeded on Transitive Inference, and respectively 15% and 67% succeeded on Class Inclusion [1] (Table 16). Correlations are directly testable using within-participant across-paradigm studies. For this reason, our primary data sources come from three studies showing significant correlations between Transitive Inference, Class Inclusion, and Cardinality [1]; Transitive Inference, Class Inclusion, Cardinality and Theory of Mind [25]; and Transitive Inference, Class Inclusion, and Balance-scale[26]. For example, the correlations in performance between Transitive Inference and Class Inclusion (0.54), Transitive Inference and Cardinality (0.54), and Class Inclusion and Cardinality (0.47) were all significant ($p < .01$) [1] (Table 6). However, to illustrate the diverse applicability of our category theory approach, we also include two other contrast studies revealing significant differences between the same age groups (i.e., younger versus older than age five) albeit with different individuals for Matrix Completion (see [2]), and Dimensional Change Card Sorting [27].

Transitive Inference

A transitive inference has the general form that given aRb and bRc, then one can infer aRc, where R is some binary relation that has the transitivity property. A Transitive Inference task, as typically administered to children, involves presenting participants with a series of premise pairs followed by a series of test pairs to assess inferential capability. The premise series usually consists of four pairs, AB, BC, CD, and DE, and testing is done on non-adjacent pair, BD (not in the premise series). AC and DE are not considered as evidence of transitive inference, because a consistent response is obtainable by noting that A or E was paired with only one other stimulus.

Transitivity is a property of relations, so a transitive inference is just a particular operation in relational algebra. In relational algebra, an *equijoin* of two relations is the set of tuples that have the same values on the specified attributes. For example, suppose $R = \{(A,B),(B,C)\}$ and $S = \{(C,D),(D,E)\}$, then the equijoin along the second and first attributes of R and S (respectively) is $\{(B,C,D)\}$. (Only tuples with the same values at the specified attributes are joined, and the redundant attribute removed. For further details of relational operators, see for example [28].) A transitive inference, then, involves a product of premise relations, indicated in the following example diagram

$$
(b,c) \xleftarrow{\ \pi_{[1,2]}\ } (b,c,d) \xrightarrow{\ \pi_{[2,3]}\ } (c,d)
\qquad (13)
$$

where π_i is the *project* operator in relational algebra, returning the values of each relation instance at the attributes listed by i. The transitive inference bRd involves the constraint that premises involving b and d share a common element c in the second and first positions, respectively (i.e., bRc and cRd). This constraint is captured by the following diagram (which relates to a pullback)

$$(b,c,d) \overset{\pi_{[2,3]}}{\longmapsto} (c,d)$$

$$\Big\downarrow \pi_{[1,2]} \qquad \pi_{[1]} \Big\uparrow \qquad\qquad (14)$$

$$(b,c) \overset{\pi_{[2]}}{\longmapsto} c$$

where the joins of other premise pairs (e.g., AB and BC based on the common element B) are omitted for clarity.

To contrast younger versus older children's performance, children were presented with difficult and simple versions of this paradigm [1]. In the difficult version of the task, children were presented with two non-adjacent blocks BD and asked which block would be higher in a tower constructed from the premise pairs (mini-towers). From Diagram 14 we see that this inference involves a map from the product object. That is, given b as one block, then the other block will be higher if it corresponds to d. In the simpler version of the task, children were given one of the mini-towers and a sequence of adjacent blocks and asked to build the complete tower (e.g., BC, followed by D, A, E). Each step only requires a map from one of the premise objects to determine where the next block should be placed. Thus, it does not require computing the product. Significantly, while younger and older children were successful on the simpler version of the task, the older children but not the younger ones were generally successful on the difficult version [1].

Class Inclusion

In a Class Inclusion task, participants are given examples of a superclass, and two complementary subclasses and asked about their relative sizes. For example, given the superclass, *fruit*, and subclasses *apples* and *non-apples*, participants are asked, *Are there more apples than fruit?* We show that class inclusion involves a coproduct. Coproduct is the dual of product, hence there is a duality between Class Inclusion and Transitive Inference.

Class inclusion is a property of sets, so a class inclusion inference involves a particular set operation—disjoint union. As we have seen, the disjoint union of two objects in the category of sets is the coproduct. Suppose, for example, the set of apple referents, or indices $\{i_1,\ldots,i_a\}$ and non-apple indices $\{j_1,\ldots,j_n\}$. The coproduct is

$$\{i_1,\ldots,i_a\} \overset{q_A}{\longrightarrow} \begin{matrix}\{(A,i_1),\ldots,(A,i_a),\\(N,j_1),\ldots,(N,j_n)\}\end{matrix} \overset{q_N}{\longleftarrow} \{j_1,\ldots,j_m\} \quad (15)$$

where q_A and q_N are the apple and non-apple injection maps, respectively. The inference is obtained by observing the cardinality of each set. Typically, a Class Inclusion task involves complementary subsets, so their intersection is empty. This arrangement is captured in the following pushout diagram

$$\varnothing \longrightarrow \{j_1,\ldots,j_n\}$$

$$\Big\downarrow \qquad\qquad q_N \Big\downarrow \qquad\qquad (16)$$

$$\{i_1,\ldots,i_a\} \overset{q_A}{\longrightarrow} \begin{matrix}\{(A,i_1),\ldots,(A,i_a),\\(N,j_1),\ldots,(N,j_n)\}\end{matrix}$$

In a variation of Class Inclusion where $A \cap B = C$, elements common to subclasses A and B would be identified by $A \uplus_C B$, so they would not be counted twice in the superclass, i.e., $|A \uplus_C B| = |A \cup B|$.

The same groups of children who were tested on Transitive Inference were also tested on Class Inclusion [1]. Three questions

were posed to children who performed a version of Class Inclusion consisting of blue triangles and circles, so that the two subclasses were triangles and circles and the superclass was blue shapes. They were: (1) *Are there more triangles than circles?* (2) *Are there more blue things or more triangles?* (3) *Are there more circles or more blue things?* The older children were successful on all three questions, whereas the younger children were generally successful on the first question only [1]. Questions 2 and 3 involve maps from one of the component objects and the coproduct object to their cardinalities. By contrast, Question 1 involves maps from the component objects only, so the coproduct object is not involved.

There is a subtle difference between the diagrams for Transitive Inference and Class Inclusion. Transitive Inference involves a *constrained* product, while Class Inclusion involves an *unconstrained* coproduct. The bottom-right object in Diagram 14 is not a terminal object (other constraining elements were omitted), whereas the top-left object in Diagram 16 is the initial object. This difference has implications for pullback/pushout diagrams containing (co)products of more than two objects, which we address in the next section when we consider a more complex version of Class Inclusion. The other paradigms considered in the remainder of this section involve only unconstrained (co)products.

Other paradigms

Transitive Inference and Class Inclusion are both difficult for children below about the age of five years. Our analysis indicates that underlying this common difficulty is a lack of capacity to compute categorical (co)products. In the remainder of this section, we analyze other tasks used to compare performance within and contrast performance between groups of younger and older children.

Matrix Completion. In a modified version of Matrix Completion [2], children are presented with a grid of figures that vary along rows and columns in either one or two feature dimensions (e.g., colour and shape). The task is to infer a missing figure that matches features with the other figures in the corresponding row and column. For example, if the first, second and third rows contain circles, triangles and squares; and the first, second and third columns contain red, green and blue figures (respectively), then the missing figure in the second row and column is a green triangle. This inference is obtained from a product of the object containing the shapes and the object containing colours. The pullback is indicated by the following diagram

$$S \times C \overset{\pi_c}{\longrightarrow} C$$

$$\Big\downarrow \pi_s \qquad\qquad \Big\downarrow \qquad\qquad (17)$$

$$S \longrightarrow 1$$

where $S \times C$ is the Cartesian product of the set of shapes S and colours C; π_s and π_c are the shape and colour projections; and 1 is terminal (i.e., there are no constraints on the product). The grid identifying shape locations is also construed as a product, $S \overset{p_X}{\leftarrow} X \times Y \overset{p_Y}{\rightarrow} C$, where X and Y are the row and column positions in the grid and p_X and p_Y are the respective index to shape and colour maps. We regard this product as input to the cognitive system—thus $X \times Y$ takes the roll of Z in the definition of product. The defining property of a product guarantees the existence of $\langle z_1, z_2 \rangle : X \times Y \to S \times C$, and $X \times Y \cong S \times C$ (see Products), so the correct coloured shape can be inferred from the missing figure location. Older children have demonstrated successful performance on this task. However, younger children

were generally shown to be capable of only simpler forms of this task where one of the feature dimensions was constant (e.g., all red shapes, or all triangles, see [2]). In the simpler forms of this task, one of the component objects contains only one element. For example, if all shapes were coloured red, then the colour object is the set {red}. As indicated in the previous section, a single element set is a terminal object in the category **Set**. The product object as indicated in the following diagram

$$S \times \{\text{red}\} \xrightarrow{\pi_c} \{\text{red}\}$$
$$\left\downarrow \pi_s \right. \qquad\qquad \left\downarrow \right. \qquad\qquad (18)$$
$$S \longrightarrow 1$$

is isomorphic to the shape object, S, and therefore not required for the inference. The correct figure is inferred by a direct map to the shape object (i.e., p_X). Since all terminal objects are isomorphic, the same result is obtained regardless of the particular feature held constant, as we would expect.

Cardinality. Participants are presented with rows of items (e.g., four ducks, five frogs, or seven balls), and are asked three types of questions: (1) *How many x are there?* where x identifies the type of item (e.g., ducks); (2) *Can you show me y by drawing a circle around y?* where y is the number of items counted; and (3) *If you counted from the other end, how many would there be?* The third question directly tests the understanding that counting does not depend on order. Thus, a counting strategy that simply increments a counter in a particular order would fail when the order was reversed. A counting strategy that does not rely on item order must keep track of two types of items: those that have been counted; and those that have not. These two types are complementary subsets of the total set of items to be counted. Hence, Cardinality involves the same pushout structure as Class Inclusion, as indicated by the following diagram

$$\varnothing \longrightarrow \{j_1, \ldots, j_u\}$$
$$\left\downarrow \qquad\qquad \left\downarrow q_U \right. \qquad (19)$$
$$\{i_1, \ldots, i_c\} \xrightarrow{q_C} \begin{array}{l}\{(C, i_1), \ldots, (C, i_c),\\ (U, j_1), \ldots, (U, j_u)\}\end{array}$$

where C and U refer to the counted and uncounted subsets, respectively. In general, four-year-olds were successful on the first two questions, but only the older groups (five years and above) were successful on the third [1]. Questions 1 and 2 do not require anything more than in-order counting of a set of elements already identified by a single type or location. Hence, counting in these two situations does not involve constructing a coproduct of subtypes. Although this task arguably has an added component where elements are transferred from the uncounted to the counted set, it is not part of the coproduct process.

Dimensional change card sorting. In this task, participants are presented cards identifiable by coloured shapes on the visible side. Two target cards are placed on a table. Children play two sorting games by placing additional (sort) cards under one of the target cards based on the same colour (colour game), or same shape (shape game). Suppose, for example, the target cards were labeled *red triangle* (▼) and *green circle* (◯). In the context of playing the colour game, a child should place a *red circle* (●) card under the *red triangle* target card, and a *green triangle* (▽) card under the *green circle* target card. In the context of the shape game, a *red circle* card should be placed under the *green circle* target, and a *green triangle* card under the *red triangle* target. In general, only the five- and six-year-olds performed this task above chance level, but not the three- and

four-year-olds [27]. By contrast, three- and four-year-olds were only successful on a simpler form of the task where one of the two feature dimensions was constant. For example, in the shape game, the additional cards were either red triangle or red circle cards. The more difficult version involves a product between dimension (colour, shape) and the two sort cards (●,▽) as indicated by the following diagram

$$\begin{array}{l}\{(C, \bullet), (C, \triangledown),\\ (S, \bullet), (S, \triangledown)\}\end{array} \xrightarrow{f} \{\bullet, \triangledown\}$$
$$\left\downarrow d \qquad\qquad\qquad \left\downarrow \right. \qquad (20)$$
$$\{C, S\} \longrightarrow 1$$

where d and f are the dimension and figure projections (respectively). The inference is obtained by a map from the product object to the object containing the two target cards (not shown). In the simpler version of this task, the sort cards differ on only one dimension, so the games involve different sets of sort cards. For the colour game, the sort cards are ● and ◯; and for the shape game, ● and ▼. In this case, the dimension and sort card objects do not form a *correct* product—because the sort card object now has three elements, the product contains pairs $(C, ▼)$ and $(S, ◯)$ that were not part of the task. Rather, in each game, the target is inferred from just one feature dimension, either colour or shape but not both. Hence, the simpler version does not involve a product.

Balance-scale: weight-distance integration. In a modified form of the balance-scale task, called weight-distance integration, participants were shown a one-arm balance and asked to predict the degree of tilt given a weight placed at a distance from the pivot [26]. The arm was horizontal when no weight was added, and pointed to one of nine pictured animals when tilted. Thus, a tilt angle prediction was indicated by the expected animal pointed to after releasing the arm. A spring provided the appropriate balance mechanism. There were three levels of weight and three levels of distance. Five-, six- and seven-year-olds predicted tilt by integrating both distance and weight information, either additively or multiplicatively. Four-year-olds predicted tilt without integration (i.e., by weight, or distance only). Three-year-olds used neither strategy and were generally non-systematic in their responses [26]. Since product and coproduct are generalizations of multiplication and addition, then the integrative strategies are naturally captured by pullbacks and pushouts, respectively. The pullback is indicated by the diagram

$$W \times D \xrightarrow{\pi_d} D$$
$$\left\downarrow \pi_w \qquad\qquad \left\downarrow \right. \qquad (21)$$
$$W \longrightarrow 1$$

where $W \times D$ is the Cartesian product of the set of weights W and distances D; π_w and π_d are the weight and distance projections; and 1 is terminal (i.e., there are no constraints on the product). Tilt prediction is a map $t_\times : W \times D \to A$ from the product object to the set of animals indicating the degree of tilt (not shown). The additive strategy is captured by the following pushout diagram

$$\varnothing \longrightarrow D$$
$$\left\downarrow \qquad\qquad \left\downarrow q_d \right. \qquad (22)$$
$$W \xrightarrow{q_w} W + D$$

where $W + D$ is the disjoint union of weight and distance sets; and q_w and q_d are the corresponding injections. Tilt prediction is a map $t_+ : W + D \rightarrow A$ (not shown). The simpler strategy used by the four-year-olds just involves a map from either the weight or distance object to the predicted tilt position.

Theory of mind. For this paradigm two sorts of tasks were employed: appearance-reality; and false-belief [25]. As an example of an appearance-reality task, children are shown milk poured into a glass wrapped in a red filter. The colour of the milk is visible before and after being poured into the glass. After filling the glass they are asked an appearance question, *When you look at this milk right now, does it look red or does it look white?*; and a reality question *What color is the milk really and truly? Is it white or is it red?* As an example of a false-belief task, children are shown drawings of a boy and a puppy and told the story, *Sam wants to find his puppy. Sam's puppy is really in the kitchen. Sam thinks his puppy is in the bathroom.* They are then asked a belief question, *Where will Sam look first for his puppy, the bathroom or the kitchen?*; and a reality question, *Where is the puppy really, the bathroom or the kitchen?* Four variations of appearance-reality and four variations of false-belief tasks were tested on three age groups: three-, four- and five-year-olds (16 participants in each group). Only one of the three-year-olds, seven of the four-year-olds and twelve of the five-year-olds passed the combined tasks, where pooled responses over the eight tasks were significantly above chance [25]. Both appearance-reality and false-belief involve a pushout. The following diagram indicates the pushout for the milk task example

$$
\begin{array}{ccc}
\varnothing & \longrightarrow & \{\text{white}\} \\
\downarrow & & \downarrow{\scriptstyle q_r} \\
\{\text{red}\} & \xrightarrow{q_a} & \{(A, \text{red}), \\
& & (R, \text{white})\}
\end{array}
\tag{23}
$$

where q_a and q_r are the injection maps, identifying colour in the context of appearance and reality (respectively). The puppy task example is similarly indicated by the following pushout diagram

$$
\begin{array}{ccc}
\varnothing & \longrightarrow & \{\text{kitchen}\} \\
\downarrow & & \downarrow{\scriptstyle q_r} \\
\{\text{bathroom}\} & \xrightarrow{q_b} & \{(B, \text{bathroom}), \\
& & (R, \text{kitchen})\}
\end{array}
\tag{24}
$$

where q_b and q_r are the injection maps, identifying location in the context of belief and reality (respectively). In both cases, the inference is obtained by a map from the coproduct object to colour or location. By contrast, without the coproduct object one cannot determine the correct context for colour or location. Hence, the expected number of correct responses is the same as chance, corresponding to the performance of the younger children.

Summary

The distinguishing characteristic at the heart of the behavioural difference between younger (less than five years old) versus older (more than five years old) children is the categorical (co)product. In the case of Transitive Inference, Matrix Completion, and Card Sorting, this difference was realised by task design (e.g., one versus two relevant feature dimensions). In the case of Class Inclusion and Cardinality, this difference was realized by questions probing,

for example, one versus two feature dimensions. And, in the case of Balance-scale and Theory of Mind, this difference was realized by alternative task strategies as inferred from the types of response errors. In each paradigm, the more difficult situation observed in the older children required access to a (co)product. By contrast, the less difficult situation observed in younger and older children involved directly accessing the component objects without computing or accessing a (co)product. These correspondences have been confirmed directly with the same participants performing multiple paradigms that included: Transitive Inference, Class Inclusion, and Cardinality [1]; Transitive Inference, Class Inclusion, Cardinality and Theory of Mind [25]; and Transitive Inference, Class Inclusion, and Balance-scale [26].

Beyond early development: Finite (co)products

So far, our analysis has been confined to early development around the age of five, where the capacity to compute (co)products was identified as crucial. The more interesting statistic for our purposes is the correlation across paradigms, rather than a specific age of attainment. That is, for example, whether or not a four(six)-year-old who succeeds (fails) at Transitive Inference also succeeds (fails) at Class Inclusion. However, the simpler versions of these tasks often form baselines that are within the capacity of all children. In these situations, *floor* effects may attenuate the ability to detect significant correlations. A methodological solution is to contrast tasks at "higher levels" of complexity at which neither level constituents a baseline (i.e., within the capacity of all participants). Hence, in this section, we extend our analysis to more complex tasks.

A number of studies point to higher complexity levels, at least in adult cognition. For example, adults were tested on their ability to identify the number of interactions underlying fictitious data sets reported as bar graphs [29]. A two-way interaction, for instance, was identifiable by observing that the change in bar height between conditions a_1 and a_2 under condition b_1 differed under condition b_2. The maximum number of interactions that adults could effectively recognize was about four [29]. Adults have also been tested on Raven's Progressive Matrices, which is closely related to Matrix Completion, where the number of feature dimensions was increased to three. Functional magnetic resonance imaging revealed significant differences in activity for regions in the prefrontal cortex when figures varied along three versus two feature dimensions [30]. These sorts of tasks have been characterized in terms of the arity of relations processed (e.g., binary, ternary, quaternary) [3], or the number of related task "variables" [31]. Our main purpose in this section is to show how our category theory approach incorporates higher levels of complexity. Unlike the studies examined in the previous section, there have not been multi-paradigm within-participant comparison/between-participant contrast studies for these more complex tasks. So, we proceed by extending the analysis to more complex versions of the seven paradigms considered above.

In category theory, the (co)product extends naturally to any finite number of objects. Moreover, the degenerate case where the number of objects is one corresponds to the *no (co)product* cases in the previous section. First, we provide the basic definitions before showing how the existing paradigms are extendable to more complex cases.

Finite (co)products. In any category **C**, a *finite product* of n objects $A_1, \ldots, A_n$ is an object P together with n morphisms $p_i : P \rightarrow A_i$, such that for any n-tuple of morphisms $z_i : Z \rightarrow A_i$, there is a unique morphism $u : Z \rightarrow P$, such that the following diagram commutes
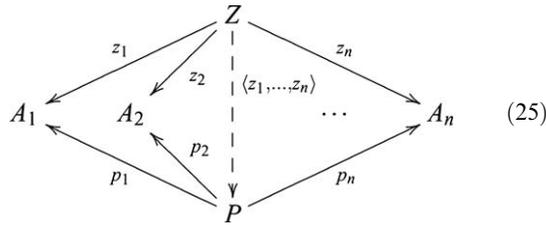
where P is denoted by $A_1 \times \ldots \times A_n$, and u by $\langle z_1, \ldots, z_n \rangle$.

The finite coproduct is defined similarly. In any category $\mathbf{C}$, a *finite coproduct* of n objects $A_1, \ldots, A_n$ is an object Q together with n morphisms $q_i : A_i \to Q$, such that for any n-tuple of morphisms $z_i : A_i \to Z$, there is a unique morphism $u : Q \to Z$, such that the following diagram commutes
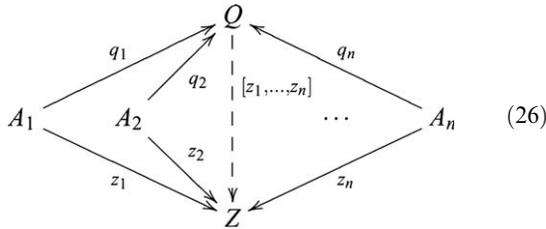

where Q is denoted by $A_1 + \ldots + A_n$, and u by $[z_1, \ldots, z_n]$. Naturally, finite product and finite coproduct are dual. That is, a coproduct of n objects in $\mathbf{C}$ is a product of n objects in $\mathbf{C}^{\mathbf{op}}$.

There are four ways of constructing a product of three objects, and the product objects, though not equal, are isomorphic, i.e., $A \times B \times C \cong (A \times B) \times C \cong A \times (B \times C) \cong (A \times B) \times_B (B \times C)$. The construction $A \times (B \times C)$ is indicated in the following diagram
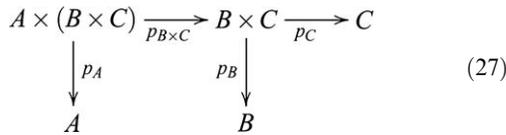

The diagram for the construction $(A \times B) \times C$ is similar except that the product objects are arranged vertically. However, the construction of $(A \times B) \times_B (B \times C)$ also involves a pullback, as indicated in the following diagram
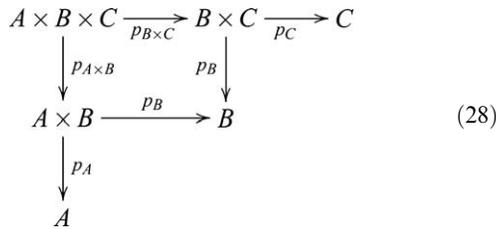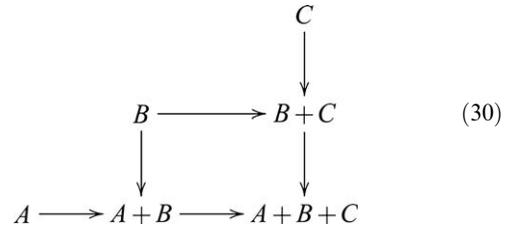

where the ternary product object is denoted $A \times B \times C$, since the four ternary product objects are isomorphic.

There are also four ways of constructing a coproduct of three objects, and the coproduct objects are also isomorphic, i.e., $A + B + C \cong (A + B) + C \cong A + (B + C) \cong (A + B) +_B (B + C)$. These constructions are dual to the respective products. The construction of $(A + B) + C$ is indicated by the diagram
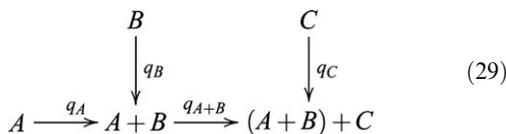

and $(A + B) +_B (B + C)$, which contains a pushout, is indicated by the diagram

where the ternary coproduct is denoted $A + B + C$.

Extensions to paradigms. All seven paradigms analyzed in the previous section can be extended in terms of (co)products of more than two objects. Only ternary (co)products are considered here, but extensions to more objects are also possible. We focus on Transitive Inference and Class Inclusion, and sketch extensions to the other paradigms. Transitive Inference can be extended to include an additional premise EF, and an additional nonadjacent test pair BE that requires two equijoins, for example, BC and CD to infer BD, and BD and DE to infer BE. In category theory terms, this inference involves three pullbacks, indicated by the diagram
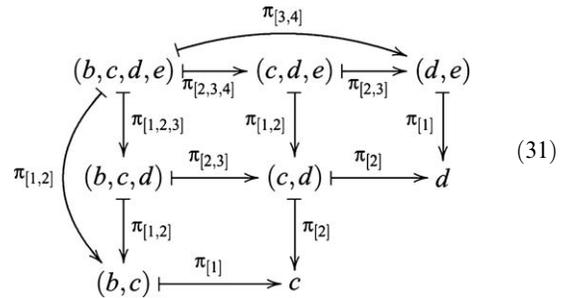
omitting some composition morphisms for clarity. By the *pullback lemma* [24], if two adjacent commutative squares are pullbacks, then the composed *commutative rectangle* is a pullback. This diagram contains two commutative rectangles, indicated in part by the composition arrows $\pi_{[3,4]}$ and $\pi_{[1,2]}$.

Class Inclusion can be extended dually by supposing an additional subclass (e.g., squares). For example, participants are presented with small blue triangles (T), small red circles (C) and large red squares (S). They are asked: (1) *Are there more triangles and circles than red shapes?* (2) *Are there more triangles and circles than shapes?* (3) *Are there more circles and squares than shapes?* Question 1 involves binary coproducts, whereas Questions 2 and 3 involve ternary coproducts. For comparison with Question 1 of the original Class Inclusion task, the question *Are there more triangles or circles?* involves neither binary nor ternary coproduct, or what might be called a unary coproduct. The ternary coproduct is indicated in the diagram
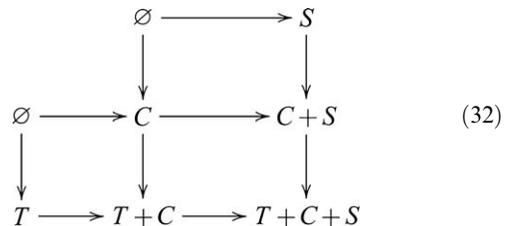
Notice that although this diagram involves a ternary coproduct, which is dual to a ternary product, the diagram itself is not dual to Diagram 31 for extended Transitive Inference. The reason is that

the two initial objects in the extended Class Inclusion diagram are the same (and so too are the two morphisms with C as their codomain). Since the diagrams do not have the same number of objects (i.e., eight versus seven), they cannot be isomorphic. This difference, which we touched on earlier, arises because the coproduct is unconstrained.

An alternative version of extended Class Inclusion that uses constrained coproducts involves subclasses containing common elements. For example, suppose that instead we have a collection of small and large rectangular bars of various colours and orientations. Within this collection, three subclasses are relevant: small (S), red (R) and vertical (V) bars, which include small red (SR) and red vertical (RV) bars. The following corresponding questions are asked: (1) *Are there more small bars and red bars than red bars and vertical bars?* (2) *Are there more small bars and red bars than bars?* (3) *Are there more red bars and vertical bars than bars?* The constrained coproducts are indicated in the following diagram.

$$
\begin{array}{ccc}
RV & \longrightarrow & V \\
\downarrow & & \downarrow \\
SR \longrightarrow R & \longrightarrow & R+V \\
\downarrow & \downarrow & \downarrow \\
S \longrightarrow S+R & \longrightarrow & S+R+V
\end{array}
\qquad (33)
$$

where SR and RV are not the same object. Hence, this diagram is dual to the extended Transitive Inference diagram.

Matrix Completion, Dimensional Change Card Sorting, and Balance-scale involve similar extensions to ternary products, though the latter two are redesigned to accommodate all three levels of products (i.e., unary, binary, and ternary) within the one paradigm. For Matrix Completion, the figures vary along a third feature dimension, such as size. In this case, the task involves a ternary product of colour, shape and size, i.e., $C \times Sh \times Sz$. Dimensional Change Card Sorting can be modified to include a third *switch* rule indicating whether the cards are to be sorted using the same (no switch) or opposite (switch) colour or shape feature. In this case, the task involves a ternary product of rule, dimension and shape, i.e., $R \times D \times S$. The Balance-scale task can be modified so that tilt angle also depends on spring strength. This case involves a product of the spring (S), weight and distance, i.e., $S \times W \times D$.

Theory of Mind and Cardinality involve more substantial changes, so we address these two tasks separately. Theory of Mind can be extended by including an additional transformation condition that involves mixing powered chocolate, which changes the colour of milk to brown. In this case, there are two binary coproducts for separately combining the reality and filtered glass contexts, and reality and mixing contexts, and one ternary coproduct for combining all three contexts, as indicated by the following diagram

$$
\begin{array}{ccc}
\varnothing & \longrightarrow & \{\text{brown}\} \\
& & \downarrow \\
& \{\text{white}\} \longrightarrow & \{(R,\text{white}), \\
& & (M,\text{brown})\} \\
& & \downarrow \\
\{\text{red}\} \longrightarrow \{(F,\text{red}), & \longrightarrow & \{(F,\text{red}), \\
(R,\text{white})\} & & (R,\text{white}), \\
& & (M,\text{brown})\}
\end{array}
\qquad (34)
$$

where R is the reality context, F is the filter context, and M is the mixing context, and prefixing for the ternary coproduct elements is omitted for clarity. This diagram, like Diagram 32, involves unconstrained coproducts. It is isomorphic to the unconstrained version of extended Class Inclusion (Diagram 32), but not the constrained version (Diagram 33), or extended Transitive Inference (Diagram 31). Unlike Class Inclusion, though, there does not appear to be a constrained coproduct version of Theory of Mind in this form, since there is only one object of interest (e.g., milk) in the task.

Cardinality in its current form, though, does not appear to have an extension to ternary coproducts. A possible alternative form, similar to extended Class Inclusion, requires participants to count various combinations subclasses/superclasses (e.g., triangles, circle and squares). The diagram for this case involves a ternary coproduct like the one for unconstrained or constrained Class Inclusion (see Diagram 32 and 33), where the objects are sets of indices and the coproduct is disjoint union (see Diagram 16).

Segmenting and chunking. For some tasks, there may exist alternative task strategies for achieving the same goals without exceeding capacity limits. In the context of Relational Complexity Theory, two general strategies were identified as segmenting and chunking [3]. *Segmenting* refers to serializing a process by computing intermediate results. In fact, an example of segmenting appeared in the previous section as the simplified version of Transitive Inference, where complete premise towers were constructed from a sequence of mini-towers, circumventing the binary product. The key to this strategy was supplied by the experimenter in the form of adjacent blocks. *Chunking* refers to recoding information to temporarily reduce (relational) details. Labeling is a common strategy. For example, the relationships between two hydrogen and one oxygen atom are chunked as water. A key to this strategy is learning. Here, we show the category theory equivalent for the extended versions of Transitive Inference and Class Inclusion.

From the definitions, we saw four ways of constructing ternary (co)product objects. Although these objects are isomorphic, Diagram 27 and 28 for constructing $A \times (B \times C)$ and $(A \times B) \times_B (B \times C)$ are not, since the first refers to five objects, but the second refers to six. This difference suggests an alternative diagram for computing Transitive Inference BE from premises BC, CD, and DE. Indeed, in the following diagram

$$
\begin{array}{ccccccc}
& & \pi_{[1,3]} & & & & \\
(c,d,e) \xmapsto{\pi_{[2,3]}} (d,e) & & & (b,c,e) \xmapsto{\pi_{[2,3]}} (c,e) & & \\
\downarrow \pi_{[1,2]} \quad \downarrow \pi_{[1]} & & & \downarrow \pi_{[1,2]} \quad \downarrow \pi_{[1]} & & \\
(c,d) \xmapsto{\pi_{[2]}} d & & & (b,c) \xmapsto{\pi_{[2]}} c & &
\end{array}
\qquad (35)
$$

the inference BE is derived by mapping out elements b and e from the product object $\{(b,c,e)\}$ in the right pullback with the morphism $\pi'_{[1,3]}$ (not shown). Notice that the two pullbacks in this diagram involve only binary products. The crucial step involves the projection morphism $\pi_{[1,3]}$, which removes the middle term d, as it is no longer needed to obtain the inference via a second pullback. The key to this strategy is being able to segment the inference into a sequence of two steps.

A related situation also arises for Class Inclusion, which is shown in the following diagram

$$
\begin{array}{ccccccc}
\varnothing \longrightarrow C & & \varnothing \longrightarrow S \\
\downarrow \quad \downarrow & & \downarrow \quad \downarrow \\
T \longrightarrow T+C \xrightarrow{f} TC \longrightarrow TC+S
\end{array}
\qquad (36)
$$

where the morphism $f : T + C \rightarrow TC$ is not inclusion, but some recoding (chunking) of triangles and circles into instances of a common object type, TC. For example, a possible key to accessing this type of strategy is to recode on the basis of a common feature by asking the question *Are there more small shapes than shapes?* These two examples illustrate in category theory terms how the complexity of a task may be influenced by strategy.

Discussion

Using category theory constructs, we have revealed a formal connection between Transitive Inference and Class Inclusion. Transitive Inference involves a categorical product of premise relations. Class Inclusion involves a coproduct between two complementary subclasses. In category theory, product and coproduct are dual. Thus, the formal connection between Transitive Inference and Class Inclusion is that they involve the "same" (isomorphic) processes in the categorical sense. This connection extends to other tasks establishing an equivalence class of inferential abilities formally based on the need to compute (co)products. In the simpler, one-dimensional version of Matrix Completion, the apparent product is isomorphic to a structure that does not involve a (co)product. Note that children are not required to first compute the (co)product to realize that it's reducible: they use a (co)product-free strategy which works, because of the simpler nature of the task. These results point to a fundamental principle under development during childhood that is the capacity to compute (co)products.

The implication that computing (co)products is fundamental to cognitive development raises two general questions: (1) Is the connection between these inferential abilities real, or just a coincidence? (2) If the connection is real, what does computing a (co)product mean in terms of possible neurocognitive processes? To the first question, as with any theory, one cannot rule out the possibility of being discounted by new data. The best one can hope for is to account for a wide variety of cases that are within the intended scope of the theory. In this regard, the empirical evidence now available and the variety of cases analyzed, both positive and negative conditions in each of seven paradigms, gives us cause for confidence that the connection is indeed real.

There are several caveats, however, in regard to establishing correspondences between paradigms and age groups. First, as already mentioned, an important consideration is the correlation across paradigms, not a specific age of achievement. Second, task knowledge and familiarity with materials will obviously be modulating factors. Third, in some cases, there may exist alternative task strategies that circumvent a particular level of complexity, as shown in the extended versions of Transitive Inference and Class Inclusion. These sorts of considerations have been discussed elsewhere in the context of Relational Complexity theory [32]. Hence, while we argue that categorical (co)product captures an important aspect of cognitive development, it is not intended to be the only consideration, nor is it necessarily incompatible with other approaches.

Category theory offers a potentially powerful approach to theorizing about cognition by not having to presuppose an, as yet, unknown internal structure for cognitive states representing task elements. Notice that the definition of a functor, and therefore duality (see Text S1) does not make reference to the elements within an object (i.e., an object's *internal* structure). The definitions refer only to the morphisms, which constrain the relationships between objects (i.e., their *external* structure). So, one is not required to make an *a priori* commitment to, say, symbolic or subsymbolic computational processes. In this sense, category theory complements more detailed (e.g., symbolic, or connectionist) approaches to cognitive modeling. In particular, we started with the difficulty that Transitive Inference and Class Inclusion poses for *Relational Complexity* and *Complexity and Cognitive Control* theories. From our categorical perspective, we now see that the Relational Complexity explanation of Transitive Inference (i.e., integration of two binary relations into a ternary relation) is a special case of a categorical product. The commutative diagrams for ternary (co)products also show how one may incorporate a levels of hierarchy explanation, as may be employed by Complexity and Cognitive Control theory, where a ternary (co)product may be computed from two binary (co)products.

While the abstractness afforded by category theory is generally seen as a strength, it leaves open the question of what exactly is being computed in these situations. To the second question, then, we look to neuroscience. One of the major attractions of category theory for mathematicians and computer scientists is that it offers abstraction (hence, generalization) *with* precision. Cognitive neuroscience research has implicated the prefrontal cortex as important for processing relational information [30,33–35]. For example, patients with damage to prefrontal cortex were significantly worse on Transitive Inference and Class Inclusion tasks than normals and patients with anterior temporal cortical damage [35]. Adults, but not children (8–12 years old) showed sustained activity in rostrolateral prefrontal cortex during the more difficult two-relation than one-relation condition of a Raven's Progressive Matrices task [33]. The general suggestion has been that regions within the prefrontal cortex are responsible, in some informal sense, for the integration and maintenance of relational information [30]. Our category theory approach makes more precise claims in formal terms of pullbacks and pushouts.

Research on the neural basis of reasoning has focussed on localizing functionality to specific cortical regions, particularly within the prefrontal cortex. Yet, the commutative diagrams clearly show the importance of transformations between objects. One intriguing possibility is that the morphisms correspond to functional connectivity realized in part by long-distance cortical connections. An area where the neural basis of cognitive function has been studied in detail is visual attention (see [36]). Conjunctive visual search involves finding a target item among a display of non-targets, where the target is uniquely identified by a conjunction of features, such as colour and orientation. In categorical terms, conjunctive search involves a product of (e.g., colour and orientation) feature maps. Each feature map is a set of location-feature relationships, and their conjunction is the product of those maps constrained by location (i.e., a pullback). It is well-known that conjunctive search is more difficult (steeper search slope) than feature search (see [37]). Interestingly, a visual search study on monkeys using implanted electrodes revealed greater frontal-parietal neural synchrony in the lower gamma band (22–34 Hz) for conjunctive than feature search [38]. A corresponding significant increase in phase synchrony between frontal and parietal scalp electrodes in the same frequency band was also reported in humans [39]. Whether the product underlying conjunctive search relates to the products identified here remains to be determined. What this example illustrates is a further benefit of a categorical approach, where the methods of one field are shown to have novel applications in another—in this example, phase synchrony as an indicator of complexity.

A recurring theme in our analysis of these tasks is the integration (either multiplicatively, or additively) of multiple sources of information. Regions within the prefrontal cortex are often assigned this role, both anatomically and functionally (see [40] for a review). A general theory of intelligence proposes that

maturation of the prefrontal cortex in coordination with other cortical regions is a key factor [41]. Hence, maturation of cortical connectivity is a possible biological basis for the observed correspondences in the development of inference, though we do not regard maturation as the only factor, as already discussed.

More generally, we have used category theory to propose new experiments that directly test comparisons and contrasts for all levels. The basis for determining whether tasks belong to the same level is isomorphism, either between objects or the diagrams (categories) to which they belong. In regard to the latter, we identified a subtle difference between diagrams containing constrained versus unconstrained (co)products. This difference speaks to the potential power of category theory in that it affords a finer grained analysis within the major levels defined by (co)product arity (i.e., unary, binary, ternary, etc). Although further work is needed to ascertain the empirical implications of these differences both within and across higher levels, the examples provided show how this work may proceed.

There are two main types of predictions for these extended paradigms that follow naturally from the arities of computed (co)oroducts. They are: (1) tasks involving (co)products of arity n will yield significantly lower performance than tasks involving (co)products are arity $m < n$ for participants within the same age group, excluding of course floor effects, where performance on neither task is above chance; and (2) tasks at the same arity will yield significance performance correlations. A corollary to these predictions is that older participants will generally outperform younger participants on a task at a given arity.

One may wonder whether other category theory-based models could account for the same developmental data. **Set** has been the categorical basis for our analysis. A natural alternative for modeling relations is the category **Rel**, which has sets, $A, B, \ldots$, for objects; and for morphisms from A to B, relations $R \subseteq A \times B$, instead of functions, where the identity morphism on A is the equality relation, $\{(a,a) | a \in A\}$; and composition defined so that for $R : A \to B$ and $S : B \to C$, $S \circ R = \{(a,c) | \exists b \in B : (a,b) \in R \wedge (b,c) \in S\}$. Composition in this category is essentially an equijoin. Thus, Transitive Inference on a set A is represented more succinctly by a diagram consisting of a morphism, $R : A \to A$ and its composition with itself, $R \circ R$, being the inference. However, Class Inclusion does not lend itself to a more succinct representation in **Rel**, so the analogy between the two cannot be captured in **Rel**. (The inclusion relations $A \subset C$ and $B \subset C$ capture the data for Class Inclusion, but there is no valid composition operation in this context.) This example reinforces our earlier point that category theory, while abstract, is not an arbitrary fit-for-all formulation.

These two categorical bases for Transitive Inference (**Set** and **Rel**) raise another point in regard to a notion of cognitive flexibility mentioned earlier. The **Set** version of Transitive Inference indicates that to make the inference one must consider the constraining item c (in Diagram 14) from two perspectives conjointly: in the case of the blocks task, as the block that is both higher than b and lower than d. By contrast, in the **Rel** version, the inference relies on just one morphism or perspective, applied twice. This difference also has implications for developmental and comparative psychology in that simply demonstrating transitivity in infants and non-humans is not sufficient evidence of a cognitive capacity that is in some way equivalent to older children and adult humans. Our categorical (co)product formulation says that if they have the capacity for Transitive Inference in the same cognitively flexible manner, then they should also have the capacity for other inferences involving (co)products such as Class Inclusion, assuming a means of administering the test that is appropriate for the cohort.

Category theory affords a view of the forest despite the trees. It helps reveal unseen connections between (cognitive) structures. And, in doing so, the methods and results from one field become applicable to another. That was the original motivation for having a science of cognition.

Supporting Information

Text S1 Duality
Found at: doi:10.1371/journal.pcbi.1000599.s001 (0.09 MB PDF)

Acknowledgment

Author Contributions

Analyzed the data: SP WHW GSH. Contributed reagents/materials/analysis tools: SP WHW. Wrote the paper: SP WHW GSH. Conceived the original idea and designed the study: SP.

References

1. Andrews G, Halford GS (2002) A cognitive complexity metric applied to cognitive development. Cogn Psychol 45: 153–219.
2. Halford GS (1993) Children's understanding: The development of mental models. HillsdaleNJ: Lawrence Erlbaum.
3. Halford GS, Wilson WH, Phillips S (1998) Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. Behav Brain Sci 21: 803–831.
4. Zelazo PD, Frye D (1998) Cognitive complexity and control: II. The development of executive function in childhood. Curr Dir Psychol Sci 7: 121–126.
5. Phillips S, Halford GS, Wilson WH (1995) The processing of associations versus the processing of relations and symbols: A systematic comparison. In: Moore JD, Lehman JF, eds. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society. pp 688–691.
6. Newell A (1990) Unified theories of cognition. CambridgeMA: Harvard University Press.
7. McClelland JL, Rumelhart DE, eds (1986) Parallel Distributed processing: Explorations in the microstructure of cogntion, volume 2 of *Computational models of cognition and perception*. CambridgeMA: MIT Press.
8. Eilenberg S, Mac Lane S (1945) General theory of natural equivalences. Trans Amer Math Soc 58: 231–294.
9. Mac Lane S (2000) Categories for the Working Mathematician. Graduate Texts in Mathematics. New York, NY: Springer, 2nd edition.
10. Arbib MA, Manes EG (1975) Arrows, structures, and functors: The categorical imperative. London, UK: Academic Press.
11. Goguen J (1991) A categorical manifesto. Math Structures Comput Sci 1: 49–67.
12. Pierce BC (1991) Basic category theory for computer scientists. Foundations of Computing. Cambridge, UK: MIT Press.
13. Lawvere FW, Rosebrugh R (2003) Sets for Mathematics. CambridgeMassachusetts: Cambridge University Press.
14. Kasturirangan R (2004) Mapping spatial relations. Ph.D. thesis, Massachusetts Institute of Technology.
15. Goguen J (2005) Mathematical models of cognitive space and time. In: Andler D, Okada M, eds. Proceedings of the Symposium on Reasoning and Cognition. Japan: Keio University.
16. Goguen J (2005) What is a concept? In: Dau F, Mugnier ML, Stumme G, eds. Conceptual Structures: Common Semantics for Sharing Knowledge. Berlin: Springer. pp 52–77.
17. Halford GS, Wilson WH (1980) A category theory approach to cognitive development. Cogn Psychol 12: 356–411.
18. Awodey S (2006) Category theory. Oxford Logic Guides. New York, NY: Oxford University Press.
19. Lawvere FW, Schanuel SH (1997) Conceptual mathematics: A first introduction to categories. Foundations of Computing. Cambridge, UK: Cambridge University Press.
20. Fodor JA, Pylyshyn ZW (1988) Connectionism and cognitive architecture: A critical analysis. Cognition 28: 3–71.
21. van Gelder T (1990) Compositionality: A connectionist variation on a classical theme. Cogn Sci 14: 355–384.
22. Aizawa K (2003) The systematicity arguments. Studies in Mind and Brain. New York: Kluwer Academic.

23. Barr M, Wells C (1990) Category theory for computing science. Prentice Hall International Series in Computer Science. New York: Prentice Hall, first edition.
24. Goldblatt R (2006) Topoi: The categorial analysis of logic. Dover Publications, revised edition.
25. Andrews G, Halford GS, Bunch KM, Bowden D, Jones T (2003) Theory of mind and relational complexity. Child Dev 74: 1476–1499.
26. Andrews G, Halford GS, Murphy K, Knox K (2009) Integration of weight and distance information in young children: The role of relational complexity. Cogn Dev 24: 49–60.
27. Halford GS, Bunch K, McCredden JE (2007) Problem decomposability as a factor in complexity of the dimensional change card sort task. Cogn Dev 22: 384–391.
28. Halpin TA (1995) Conceptual schema and relational database design. Sydney: Prentice Hall Australia, 2nd edition.
29. Halford GS, Baker R, McCredden JE, Bain JD (2005) How many variables can humans process? Psychol Sci 16: 70–76.
30. Christoff K, Prabhakaran V, Dorfman J, Kroger KJ, Zhao Z, et al. (2001) Rostral prefrontal cortex involvement in relational processing during reasoning. NeuroImage 14: 1136–1149.
31. Halford GS, Phillips S, Wilson WH, McCredden J, Andrews G, et al. (2007) Relational processing is fundamental to the central executive and is limited to four variables. In: Osaka N, Logie R, D'Esposito M, eds. The Cognitive Neuroscience of working memory. New York: Oxford University Press, chapter 16. pp 261–280.
32. Halford GS, Wilson WH, Phillips S (1998) Authors' response: Relational complexity metric is effective when assessments are based on actual cognitive processes. Behav Brain Sci 21: 848–864.
33. Crone EA, Wendelken C, van Leijenhorst L, Honomichl RD, Christoff K, et al. (2009) Neurocognitive development of relational reasoning. Dev Sci 12: 55–66.
34. Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, et al. (2002) Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parameteric study of relational complexity. Cereb Cortex 12: 477–485.
35. Waltz JA, Knowlton BJ, Holyoak KJ, Boone KB, Mishkin FS, et al. (1999) A system for relational reasoning in human prefrontal cortex. Psychol Sci 10: 119–125.
36. Chalupa LM, Werner JS, eds (2004) The Visual Neurosciences, volume 2. Cambridge, MA: MIT Press.
37. Wolfe JM (2003) Moving towards solutions to some enduring controversies in visual search. Trends Cogn Sci 7: 70–76.
38. Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315: 1860–1862.
39. Phillips S, Takeda Y (2009) Greater frontal-parietal synchrony at lower frequencies for inefficient than efficient visual search in human EEG. Int J Psychophysiol 73: 350–354.
40. Ramnani N, Owen AM (2004) Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. Nat Rev Neurosci 5: 184–194.
41. Jung RE, Haler RJ (2007) The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. Behav Brain Sci 30: 135–187.