University of New Mexico
UNM Digital Repository

Electrical & Computer Engineering Technical Reports

Engineering Publications

7-1-2008

A Model of Human Categorization and Similarity Based Upon Category Theory

Michael Healy

Thomas Caudell

Timothy Goldsmith

Recommended Citation

Healy, Michael; Thomas Caudell; and Timothy Goldsmith. "A Model of Human Categorization and Similarity Based Upon Category Theory." (2008). https://digitalrepository.unm.edu/ece_rpts/28

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SCHOOL OF ENGINEERING
UNIVERSITY OF NEW MEXICO

A Model of Human Categorization and Similarity Based Upon Category Theory

Michael John Healy
Department of Electrical and Computer Engineering
e-mail: `mjhealy@ece.unm.edu`

Thomas Preston Caudell
Department of Electrical and Computer Engineering and
Department of Computer Science
e-mail: `tpc@ece.unm.edu`

Timothy E. Goldsmith
Department of Psychology
e-mail: `gold@ece.unm.edu`

University of New Mexico
Albuquerque, New Mexico 87131, USA

UNM Technical Report: EECE-TR-08-0010

Report Date: June 19, 2008

Abstract

Categorization and the judgement of similarity are fundamental in cognition. We propose that these and other activities are based upon an underlying structure of knowledge, or concept representation, in the brain. Further, we propose that this structure can be represented mathematically in a declarative form via category theory, the mathematical theory of structure. We test the resulting mathematical model in an experiment in which human subjects provide judgements of similarity for pairs of line drawings using a numerical scale to represent degrees of similarity. The resulting numerical similarities are compared with those derived from the category-theoretic model by comparing diagrams. The diagrams represent distributed concept structures underlying the line drawings. To compare with a more conventional analysis technique, we also compare the human judgements with those provided by a two-dimensional feature space model equipped with a distance metric for the line drawings. The results are equally favorable for both models. Because of this and the putative explanatory power of the category-theoretic model, we propose that this model is worthy of further exploration as a mathematical model for cognitive science.

Keywords

categorization, (mathematical) category, cognition, colimit, concept, diagram, feature, morphism, similarity, theory

1 Introduction

Categorization is a most fundamental cognitive activity, serving as the basis for much of our thinking and action. Equally fundamental is the judgment of similarity. The perceived similarity between objects determines the likelihood that the objects will be placed in the same category. Whether category structure determines how similarity is perceived or whether perceived similarity dictates the existence of categories is more a question of metaphysics than of empirical investigation. In either case, both are critical to intelligent behavior.

For the last several decades psychologists have attempted to identify formal models that capture human category and similarity judgments. In these models objects and events are often represented as either points in multidimensional space or vectors of features, and then some process for operating on these representations (e.g., distance or set-theoretic measures) is used to determine the similarity of pairs of objects. Various models (e.g., prototype and exemplar) of how the measures of similarity map the objects into categories have been studied. The literature is replete with studies investigating empirical support for these types of categorization and similarity models. However, no single model has emerged as a clear representation of how humans make such judgments.

The present study tests a new model for human category and similarity judgements. The new model introduces an apparently novel element for studies of categorization and other phenomena. The novel element is a mathematically precise formalization of structure, the ways in which entities are related. The model is a theory of concept analysis and categorization based upon the notion that the exemplars of a category are instances of a concept and therefore share a structure specified by the concept, the category itself has a structure consisting of relations on exemplars, relations on categories form a structure based upon abstraction and specialization, and, finally, these three kinds of structure are interrelated. Because the new model is theory-based, we shall refer to it as a theoretical model. The test reported here is an experiment to test a theory.

There is at least partial support in the literature for the notion that structure is an important determinant of category membership. Tversky's [18] feature contrast function allows for a variety of similarity measures based upon set intersections and set differences among subsets of exemplar features. Relative to a fixed set containing all features under consideration, the intersections and differences are part of a structure known as a Boolean lattice, with the subset relation serving as the lattice relation. See Chapter 1 of [3] for a discussion of Boolean lattices and similar structures. Malt and Johnson [13] probe a type of category structure based upon a separation of features into two different types, physical features and functional features. Through experiments with similarity of exemplars, they explore the question of whether the functional type provides core feature sets for concepts. Type distinctions do more than separate features into different classes; they form a structure based upon the subtype relation (or, regarding types and classes as sets for simplicity, this structure is, like the first, based upon the subset relation). For example, physical features have subtypes consisting of perceptual (colors, textures), enumerative (number of legs on a chair), dimensional (size, weight), material (aluminum, wood, fabric) and so forth. Murphy and Wisnieski [15] investigate the effectiveness of basic versus superordinate concepts in categorization, where basic concepts are those that are apparently most readily called upon by subjects performing in experiments. They find evidence that, while subjects are faster at categorizing isolated exemplars based upon basic concepts ("chair", "table"), the basic concept advantage is greatly reduced when calling upon superordinates ("furniture") to categorize objects in scenes ("the furniture in the room" versus "the conference-room chairs") . The superordinates are thought to include information about multiple as opposed to single objects and also relations among objects within the scene context, thereby providing more information to constrain category membership. Two levels of structure, between-concept (basic versus superordinate) and within-concept (feature sets for objects in isolation versus multi-object properties and relations), are involved here. Goldsmith and colleagues [4, 1, 8] have performed experiments with several algorithms for human concept analysis, most notably the Pathfinder algorithm [17]. Pathfinder generates a connected graph that depicts local concept relationships. It can be applied to data garnered from human judgements of concept similarities as well as graphs generated based upon a formal analysis of concept similarities. A result of the analysis based on these experiments is a method for the evaluation of a person's conceptual structure in terms of the level of competence or coherence it represents. Here, coherence measures on a 0-1 scale the internal consistency of a set of ratings by examining the extent to which sets of ratings satisfy a generalized triangle inequality law applied to a concept-to-concept distance measure. Again, structure arises naturally in this work as graphs or networks based upon concept similarities expressed as a conceptual distance measure. Medin [14] explores the evolving notion of “concept”, including categories formed from similarity relations, probabilistic statements, and other representations. The current trend is to regard concepts as theories, where a theory is a description of a domain of discourse consisting of an inferentially closed system of statements. Because they allow for inferences to be drawn from expressions of existing knowledge, theories provide not only for describing exemplars and/or prototypes but for reasoning about them as well. As Medin indicates, theories express significantly more constraints upon category membership than do other notions of “concept”. Yet with all the advantages of theories as opposed to feature-matching, the notion of similarity need not be discarded. Through the categorization process, similarity plays an epistemological role in providing access to exemplar properties and relations, the most basic content of concepts-as-theories. Not only do theories constrain category membership, but similarity constrains the search for theories that explain category membership. There are different kinds of similarity relations, and theories are related by an abstraction/specialization hierarchy. Both relations form structures.

Medin’s analysis serves as a point of departure for the new model of categorization and the testing of it presented here. When seen as a theory in the same sense as a scientific theory, a concept can be seen as having an internal structure. The structure is determined by the properties (feature specifications) and relations of the objects that make up the subject matter of the theory (see Figure 1), together with the statements that combine properties and relations to assert propositions, the laws of the theory. A scientific theory can be stated in a mixture of natural language, equations, inequalities, statistical statements, and pictorial schematics, with the proviso that any language used in the theory must be made as precise and unambiguous as possible. We shall refer to all these forms of statement as declarative or symbolic, and to the theory as a declarative or symbolic concept. Inferences can be drawn based upon a set of basic statements, which are regarded as axioms of the theory. This together with its validity as a description of its domain of discourse gives a scientific theory its predictive capability. We shall extend this notion of a theory to all concepts, on the grounds that a concept is a description whose instances are meant to be accessible to scientific study. We shall not argue the point here as to whether concepts held by humans including beliefs, social predispositions, self-awareness and familiarity with other individuals, which typically have introspective and emotional content, are amenable to such an analytical requirement. Let us speculate that in such cases it is at least true that a symbolic statement can be substituted for the actual concept without undue harm to the representation of cognitive properties.

In the present effort, the concepts involved are very simple, for they express line drawings; in fact, they are theories about geometric shapes. The shapes of interest in the experiment to be described are moderately complex, radially symmetric shapes formed from simple geometric shapes. Each shape, whether simple or complex, is described by a theory, representing its concept. For our purposes, exemplars of that shape are line drawings of the shape being described. However, there are three caveats. First, because of the theoretical model, which is based upon the mathematical discipline of category theory, some of the shapes are abstract and, hence, may be difficult to picture: They may be regarded as “amorphous shape place-holders”. The meaning of this will become clear in the discussion of later sections. Second, for the same reason, the category associated with each concept has many exemplars other than the simple shape described. Mathematically speaking, this is because in formal logic and model theory, with which we are concerned, any instance (object, event, situation) that satisfies the axioms of a theory is regarded as an instance of it; formally, the instance is called *a model of the theory*. Here, we interpret exemplars of a concept as models of the corresponding theory. So, for example, for the theory of chairs as physical objects, stating that a chair has a seat, a back, and four legs (where it is assumed that these objects are suitably defined) expresses the properties any model (exemplar) must have; however, there are many kinds of chairs, obtained by embellishing the simple chair-model with fabric covering, cushions, arm rests, fine wood carving and stains, and any of a variety of other treatments. Similarly, any line drawing that embellishes one of our line drawings (with more lines, or colors or textures) is a model of our drawing’s theory, which just describes the latter. Finally, model theory is a mathematical discipline, and models are purely mathematical constructs (for example, geometries). By regarding as models objects with physical and other properties which are not mathematically motivated, we are adopting the mathematics as a formalization of a conceived reality. This use of terminology exactly corresponds to the fact that we are regarding as theories objects ordinarily known as concepts, even though the theories are in actuality a formalization of the concepts they represent. Consequently,

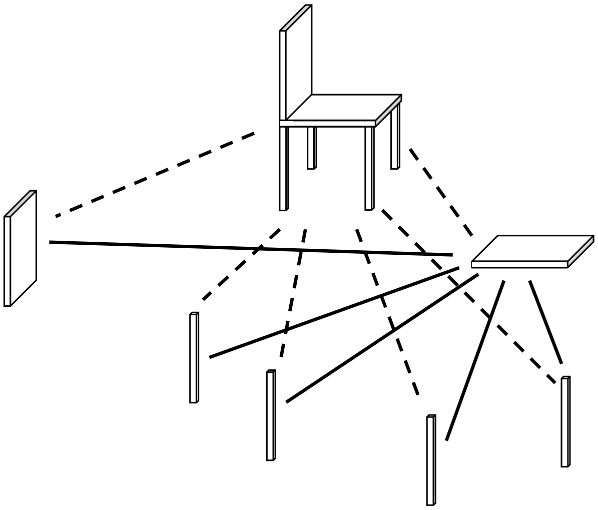

Figure 1: **A theory representing the concept *chair* in terms of physical features must describe the features and state how the features are related to make the whole. Here, the only features shown are the major shapes common to all chairs: four legs, a seat, and a seat-back. The join relation on legs, seat, and back is illustrated by solid lines. The part-whole relation for legs-chair, seat-chair and back-chair is illustrated by dashed lines.**

it is well to keep in mind that the true theories and models are mathematical entities which *represent* concepts and exemplars *to some degree of approximation.* As long as we keep this in mind, our identification of mathematical representation with reality can be justified by the fact that our theoretical model is just one in a long tradition of mathematical models that have been proposed for the sciences.

In the current study, we applied a mathematical model of a new kind to the modeling of human categorization. This new formalization is based in a discipline of mathematics called category theory, alternately known as conceptual mathematics and as the theory of structure. The objective was to test a theory of the semantics of neural networks, including the brain, that is based upon this branch of mathematics. The specific part of category theory we apply is known as categorical logic. Categorical logic includes categorical model theory, but often we mention them both for emphasis.

To date, there have been limited connections between category theory and human cognition. One exception is the work of Macnamara and Reyes [12], who used category theory to elucidate linguistic structure and the logical foundations of human cognition. Their general hypothesis is that the basic processes underlying human cognition will map onto the universal properties of category theory. We propose to investigate the more specific claim that category theory can offer a valid model of human categorization. Specifically, we are interested in examining how well certain quantities within category theory match category-relevant human judgments for a set of visual stimuli. Because similarity is assumed to be foundational to categorization, our initial empirical work focused on modeling human judgments of pairwise similarity of visual stimuli. We hypothesized that similarities defined by a category-theoretic construct known as the colimit would predict human category judgments of similarity on a set of visual stimuli. Underlying this is an assumed correspondence between the formation of complex two-dimensional shapes from components in a category of geometric shape theories, and cognitive function expressed as a category representing the operations carried out by the brain.

Because category theory is novel to many, we present a basic introduction to it in Section 2, where we first motivate its use with a discussion of theories and categories. This section introduces an important categorical construct, the colimit, which operates on a diagram to mathematically derive complex objects by "blending together" object components having specified relationships. For a full account of the theoretical model, see [7] or [5]. In Section 3, we show how geometric shapes can be represented by colimits of diagrams of interrelated components. In Section 4, we present the vocabulary of line drawings used in the experiment and introduce a similarity measure for pairs of drawings. The measure is a calculation based upon the diagrams whose colimits represent the shapes. Section 5 contains a description of the experiment, which compares the similarity judgements made by human subjects for pairs of line drawings, on a scale of 1 (little or no similarity) to 5 (very similar or exact), against similarity numbers derived from the category-theoretic diagrams. Finally, Section 6 presents the results and Section 7 contains a summary discussion and conclusion.

2 Theories and Categories

2.1 Concepts as theories, structure, and the blending of theories

Each concept has an associated category consisting of its instances, or exemplars. Because the exemplars share the properties and relations described by the concept, they themselves reflect its structure. A theory in logic can be used to express the structure with mathematical rigor, although as stated in the Introduction the theory is really a formalization intended to capture the sense of the concept. Consider chairs, purely in terms of their physical shape features. A theory about chairs as shape objects is reflected in the way each chair appears and is put together, with legs and back joined to seat in the typical fashion of a chair. The joining is a structure that combines the physical features of a chair in a systematic way to make the whole chair (again, refer to Figure 1). In the theory, this is expressed through statements that relate descriptions of the physical features to form a description of the whole. The relations and manner in which they work together in the statements of the theory form a structure consisting of relations among the various feature descriptions, where a feature can be either a part or an attribute. Thus, the physical theory of chairs has a structure that corresponds to the structure of a model of the theory. The model is a mathematical object put together from other objects which are models of simpler, more abstract theories; as such, it inherits the mathematical features of those objects. The relationship between the structures of theory and model is a formalization of the relationship between concept and exemplar.

Similar statements can be made about concepts describing aspects of exemplars other than their physical manifestation. For example, the function of a chair can be stated as a theory, in terms of the salient properties of other objects which are involved in the function. Expressed as a visual concept, a chair can be visualized as an object upon which a person can sit, a consequence of the chair providing a platform that serves as a seat with legs and a seat-back. This can be stated as a theory about an abstract platform and a sequence of bipedal figures, one standing and one sitting upon the platform, without mention of the other physical features of a chair. This differs from the physical theory in that it allows for models that include objects for sitting that are not chairs, but can be benches, swings, and other suitable objects. Here, the same observation holds as with the physical-object theory of chairs, and at this point it is useful to recall the discussion of exemplars as models of a theory in Section 1. The theory is reflected in any situation (possible world, model) in which a person is sitting, which can be visualized as a structure formed of relations among objects (the bipedal figures and an abstract platform on which one of them sits). Conversely, a model is reflected in the structure of the theory, which consists of statements about a bipedal figure sitting. Once again, this relationship between the theory and model structures formalizes the relationship between concept and exemplar.

A third theory of chairs can be obtained by combining the physical and function theories. One might expect this combination to yield a more highly constrained theory, combining the constraints upon models imposed by the physical-object theory and the function theory. Without the function theory, chairs exist as physical objects, but what are they good for? To display flowers upon, a temporary place to put items such as periodicals or holiday decorations? The primary function of a chair is not present. On the other hand, as just mentioned, without the physical theory the sitting function can be fulfilled by any of a number of objects, so the theory is not specific to chairs. The combination ought to constrain its models so that their specified function is a place to sit and the thing sat upon is a chair. However, notice that this constraint can only obtain if the two theories are "blended together" properly, in the sense that a place to sit and a chair are the same thing. This is illustrated pictorially in Figure 2. As in the Figure, it will be our practice to illustrate theories pictorially in this way as an alternative to writing them out in symbolic logic. And as will be explained in detail, this "blending" formalizes the intuition about properly combining concepts so that the resulting exemplars appear as they should.

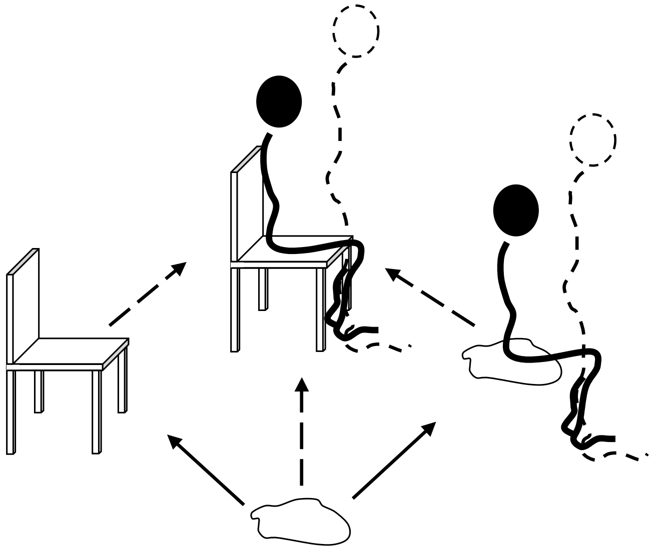

Figure 2: **A theory formalizing the physical-shape concept** `chair` **and another formalizing the functional concept** `place-to-sit` **must be "blended together" to achieve a shape+function theory formalizing the concept** `chair-as-place-to-sit`**. The solid arrows represent the mapping of a theory** `platform` **expressing a flat but otherwise-abstract surface into the other two theories. The concept** `platform` **acts as a "subtheory" on which the blending is to take place. The dashed arrows are the mappings that show how** `platform`**,** `chair`**, and** `place-to-sit` **are expressed in** `chair-as-place-to-sit`**.**

One might ask, Why worry about this "blending"? Isn't it obvious that "chair" and "place to sit" are one and the same thing in the combination of physical and function theories? The intuitive answer is that "blending" expresses the superposition of concepts so that the resulting combination appears without any ambiguity, for there can be different ways to combine two concepts. This intuition is clearly expressed mathematically in the categorical representation which is described presently. To state this in words: If the two theories are merely thrown together, say, by listing their statements, there is no guarantee that "the seat" in the chair-physical-shape theory will be the same object as "a place to sit" in the shape+function theory. Suppose, for example, that some unfamiliar artifact is being described in two ways separately, one being its physical shape and the other its function. Certain features important in properly combining the two descriptions may have been assigned different symbols or names because they play different roles, physical shape in one description and functional part in the other. Would it be clear in all such cases which physical part of the artifact takes on a particular part of the function? To make it clear requires having the knowledge of the appropriate correspondence between physical feature and function, which constitutes a "blending" of the two theories along corresponding parts. An intuitive, if speculative, part of the answer lies in an analysis of the neural mechanisms of the human brain in organizing knowledge about its environment, the impetus for our theoretical model. Imagine that one were to attempt construction of a machine to organize conceptual knowledge about chairs, say, by building a neural network "brain" model with visual sensory input connections. How is the machine to "know how" to organize its neural representations of "chair" and "place to sit" (assuming it can form these representations) in such a way that it can recognize a situation describable as "chair as place to sit"? Since it is to proceed autonomously, the machine must somehow express a "concept-joining" algorithm in a reliable fashion. This algorithm must of necessity include the "concept blending" illustrated in Figure 2.

Within our theoretical model, there is a precise sense in which concepts are "blended". As shown in Figure 2, this requires that the two theories have parts (in this case, descriptions of the seat of the chair and of a platform for sitting, respectively) that can be matched, feature for feature. These theory parts amount to two ways of stating the same thing, perhaps with different symbols and in different contexts. This implies that there is a third theory that is more abstract than either of the theories to be "blended" and can be mapped via symbol substitutions to these two theories by mapping directly onto their matching parts. In effect, the abstract theory is transported into each of the two theories, possibly transformed but still stating essentially the same thing. The two solid arrows in Figure 2 represent the two separate mappings of the more abstract theory `platform` into the two more specialized theories `chair` and `place-to-sit`. The parts of the two specialized theories corresponding to the abstract theory are called the *images* of the abstract theory under the two mappings. The dashed arrows represent the mappings of all three theories into the even more specialized theory `chair-as-place-to-sit`. The dashed arrows are a consequence of the "blending" and show where the images under these mappings of the three theories reside in the more specialized theory. In category theory, the "blended" theory and the three dashed arrows form a conical structure called a *colimit*. This can be calculated by a symbolic algorithm which uses as input the diagram consisting of the three theories `platform`, `chair` and `place-to-sit` and the two arrows `platform` $\longrightarrow$ `chair` and `platform` $\longrightarrow$ `place-to-sit`. This diagram is called the *base diagram of the colimit*, and the diagram formed by adjoining the colimit to its base diagram is called the *defining diagram of the colimit*. Figure 2 shows the entire defining diagram. The next section will make these notions more precise.

2.2 A system of theories forms a category

The discussion of the mapping arrows in the "theory blending" example illustrated in Figure 2 implicitly suggests that many concepts are related in this manner. There are two fundamental requirements of any such arrow; stating that it is a symbol substitution merely suggests its form. Let us consider a mapping m from a theory A (for example, `platform`) to a theory B (for example, `chair`, or, alternatively, `place-to-sit`), expressed symbolically in the mathematical notation $m{:}A \longrightarrow B$. The theory A is called the *domain* of m and B is called the *codomain*. The first requirement is that the symbols forming the *signature* of A—the feature sets, the properties, and the relations—must map to symbols of the same kind in its image within B. For example, the theory *platform* builds on a theory of geometry. Starting with points as undefined quantities, it defines lines, line segments, rays emanating from a point, planes, angles, and polygons as geometric constructs. A polygon is a region in a plane having a boundary which is a very special kind of simple closed curve—a polygonal one[1]. Further, the theory is extended to enable the description of an abstract shape, a region bounded by a simple (having no self-crossings) but otherwise unspecified closed curve. In `chair`, the abstract shape has as image the region appearing as a `face` of a `box` forming the `seat`, so the region is bounded by a rectangle. The symbol substitution in m can substitute relatively complex symbol-strings for simple ones, as long as the kind of a signature entity is preserved. For example, the rectangle in `chair` is a composite of simpler geometric entities, hence, describing it requires a more complex symbol-string than does the abstract shape in `platform`. The second requirement of a theory mapping is that an axiom of A—a statement accepted as true within the theory—must be transformed by the symbol substitution into either an axiom or a theorem of B—a statement that is true within the latter theory. The mapping m is said to be *truth-preserving*. Because the mapping m preserves the kinds of the entities in the signature of A and is truth-preserving, it is called a `theory morphism`; we also use the term `concept morphism` in our theoretical model. Concept morphisms can be found for many pairs of concepts. To delineate them may require some familiarity with theories stated symbolically, but the basic notions required are quite intuitive.

In general, the codomain of a theory morphism is more complex than its domain. For example, in `chair` the abstract shape of `platform` has as its image a face of a box, whose shape is that of a rectangle. This suggests that in the case in which B is `chair`, it is certainly more complex than theory A, more specialized and, hence, more constraining on its exemplars. Also, the image of A in B is expressed with more symbols. Theories form a system in which there are many truth-preserving mappings from theories that are more abstract to theories that are more specialized or more complex because they contain added constraints on their exemplars. The theory

[1] In fact, the polygonal region also has the property of *convexity*, making it even more special.

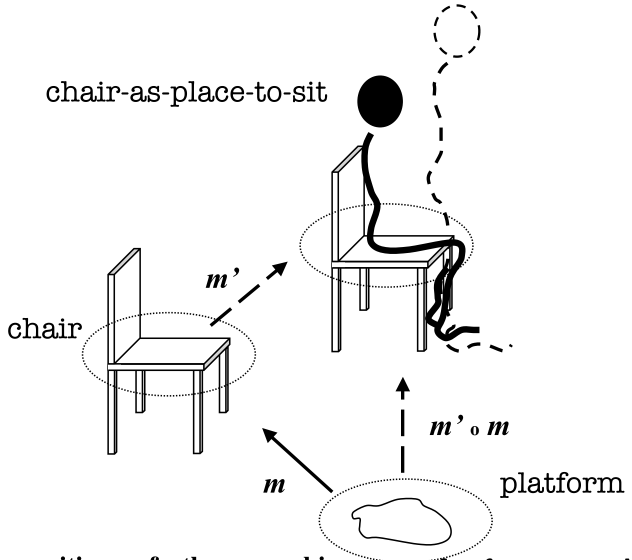

Figure 3: **The composition of the morphisms $m{:}\mathtt{platform} \longrightarrow \mathtt{chair}$ and $m'{:}\mathtt{chair} \longrightarrow \mathtt{chair-as-place-to-sit}$ yields a morphism $m' \circ m{:}\mathtt{platform} \longrightarrow \mathtt{chair-as-place-to-sit}$. Each item in `platform` can be traced via m to an item in `chair`, and then via m' to an item in `chair-as-place-to-sit`. For example, the abstract shape can be traced to its destination via each morphism, and, hence, via the composite morphism $m' \circ m$.**

`platform` in Figure 2 is more abstract in this sense than either of `chair` or `place-to-sit`, and as stated before the two solid arrows represent the two separate mappings of it into the two more specialized theories.

The most important fact about the mappings we have called theory morphisms is that they are morphisms in the sense of category theory. What this means is that the mappings are *composable*. To understand this, consider again Figure 2, and for the time being let us call the mappings theory mappings instead of morphisms. Let m be the mapping $m{:}\mathtt{platform} \longrightarrow \mathtt{chair}$ (one of the solid arrows) and let m' be the mapping $m'{:}\mathtt{chair} \longrightarrow \mathtt{chair-as-place-to-sit}$ (one of the dashed arrows). From these two arrows we may form the composition arrow $m' \circ m{:}\mathtt{platform} \longrightarrow \mathtt{chair-as-place-to-sit}$, whose domain is `platform` (the domain of m) and whose codomain is `chair-as-place-to-sit` (the codomain of m'). The composite $m' \circ m$ is obtained simply by noticing where each item in the signature of `platform` is mapped into `chair` via m, and then noticing where *that* item (the m-image of the `platform` item in `chair`) is mapped into `chair-as-place-to-sit` (the m'-image). For example, the closed curve in `platform` maps to the boundary of the rectangle in `chair` via m, and the latter item maps to the boundary of the rectangle in `chair-as-place-to-sit` via m' (Figure 3). The reverse ordering of the mappings in the symbol for the composite ($m' \circ m$) is a historical convention. The point here is that any system of objects with a compositional relation, such as theories with theory mappings, forms a *category* in the mathematical sense, provided that the composition always exists when the codomain of one relation arrow is the domain of another and, moreover, that the composition operation satisfies two axioms. The relations are then called the *morphisms* or *arrows* of the category. The first axiom is the identity axiom, stating that every object A has an arrow $\mathrm{id}_A{:}A \longrightarrow A$ whose composition with any other arrow $f{:}A \longrightarrow B$ or $g{:}C \longrightarrow A$ yields that arrow: That is, $f \circ \mathrm{id}_A{:}A \longrightarrow B$ is actually just f, stated by the equation $f \circ \mathrm{id}_A = f$, and, similarly, $\mathrm{id}_A \circ g = g$. The second is the associativity axiom, which states that if there are three arrows $f{:}A \longrightarrow B$, $g{:}B \longrightarrow C$ and $h{:}C \longrightarrow D$, then it is immaterial in which order one forms the composition of the three. That is, $h \circ (g \circ f) = (h \circ g) \circ f{:}A \longrightarrow D$, so the composition of any three such morphisms can be written without parentheses, as $h \circ g \circ f{:}A \longrightarrow D$. It is a simple matter to verify that the theory mappings are composable and their compositions satisfy the two axioms. For example, the identity arrow for the object (theory in this case) `platform`, $\mathrm{id}_{\mathtt{platform}}{:}\mathtt{platform} \longrightarrow \mathtt{platform}$, maps the plane to itself, the closed curve to itself, and all other

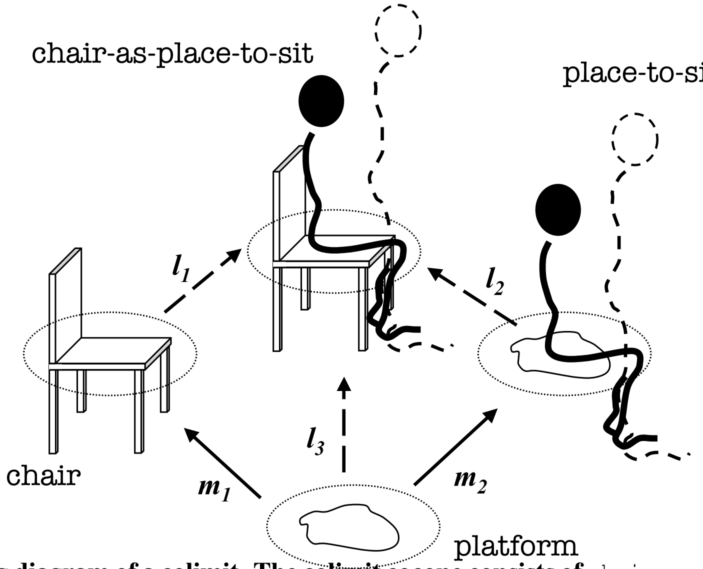

Figure 4: **The defining diagram of a colimit. The colimit cocone consists of** `chair-as-place-to-sit` **and the morphisms ℓ_1, ℓ_2 and ℓ_3. Each triangular array of morphisms commutes, $\ell_1 \circ m_1 = \ell_3$ and $\ell_2 \circ m_2 = \ell_3$.**

parts of the theory to themselves. The mapping is structure-preserving, and, in particular, it is truth-preserving. That the theory mappings are composable subject to the two axioms justifies our calling them morphisms or arrows, for they satisfy the criteria of category theory. The theories are objects of a category of theories, and the theory mappings are the morphisms or arrows of this category.

2.3 Concepts as colimits

Let us revisit the discussion of Figure 2, which introduced the notion of a colimit for the diagram consisting of the three theories `platform`, `chair`, and `place-to-sit` and the two solid arrows. We now know these arrows to be morphisms in a category of theories which includes the three theories as objects. Any collection of objects and morphisms in a category constitutes a diagram. The domain and codomain objects of a morphism are considered part of the morphism, so they are automatically included in any diagram which includes the morphism. A diagram may have many objects and morphisms, objects but no morphisms (a *discrete* diagram), or no objects at all (hence, no morphisms; this describes the *empty* diagram). A fundamental notion in category theory is that of a *commutative* or *commuting* diagram. A particular use of this notion occurs in defining *cocones* and *colimits* (and also cones and limits, which will not be discussed here). In a commutative diagram, any two morphisms having a common domain and codomain with at least one of the two being a composition morphism formed from diagram morphisms, are one and the same morphism.

Figure 4 is a replay of Figure 2 with the two solid arrows now labelled $m_1{:}\mathtt{platform} \longrightarrow \mathtt{chair}$ and $m_2{:}\mathtt{platform} \longrightarrow \mathtt{place-to-sit}$. Let us call the diagram consisting of these three objects and two morphisms D, which can also be written as a set consisting of its objects and morphisms, $\{\mathtt{platform}, \mathtt{chair}, \mathtt{place-to-sit}, m_1,$ since the objects are all either a domain or codomain of some diagram morphism, however, let us simply express D as the set of its morphisms, $\{m_1, m_2\}$. This type of expression will be used for all diagrams. The dashed arrows, whose common codomain is the theory `chair-as-place-to-sit`, are called leg morphisms for a *colimit cocone* for D which has the object `chair-as-place-to-sit` as its apex (all terminology used here is standard with most authors with the exception of "leg morphisms", which seems to us intuitive). These leg morphisms are now labelled

$$\ell_1\colon\ \ \mathtt{chair} \longrightarrow \mathtt{chair-as-place-to-sit},$$

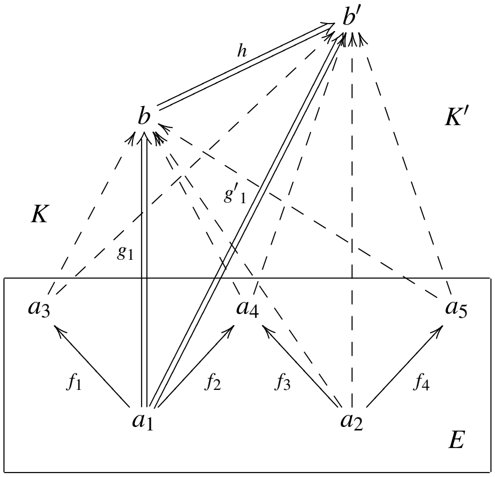

Figure 5: **A cocone morphism $h{:}K \longrightarrow K'$ in the category of cocones for a diagram E in some category C. Here, E has the form shown in the box, with objects a_1, a_2, a_3, a_4, a_5 and morphisms f_1, f_2, f_3, f_4. The cocone morphism is actually a C-morphism $h{:}b \longrightarrow b'$ between the apices of cocones K and K' that forms commutative triangles with the corresponding leg morphisms of the two cocones. A sample cocone morphism triangle is highlighted; the commutativity means that $h \circ g_1 = g'_1$. The cocone leg morphisms other than g_1 (in K) and g'_1 (in K') are not labelled to simplify the illustration.**

$$\begin{aligned} \ell_2 &: \ \mathtt{place-to-sit} \longrightarrow \mathtt{chair-as-place-to-sit}, \quad \text{and} \\ \ell_3 &: \ \mathtt{platform} \longrightarrow \mathtt{chair-as-place-to-sit}. \end{aligned}$$

From the preceding discussion, we know how to interpret the composition morphisms $\ell_1 \circ m_1$ and $\ell_2 \circ m_2$.

A colimit for D is a cocone that has a property called *initiality*. Thus, the definition of a colimit for D has two parts: First, it is a cocone, and second, it is an initial cocone. Likewise, the defining property of a cocone for D has two parts. First, it has an apex, such as `chair-as-place-to-sit`, and a single leg morphism for every object of D whose domain is that object and whose codomain is the apex, for example $\ell_1{:}\mathtt{chair} \longrightarrow \mathtt{chair-as-place-to-sit}$. Second, each triangle formed by a diagram morphism and the two leg morphisms having the diagram morphism's domain and codomain, respectively, as their domains, forms a diagram in its own right and that diagram commutes. For diagram D and the cocone of Figure 4, there are two such triangles, $\{m_1, \ell_1, \ell_3\}$ and $\{m_2, \ell_2, \ell_3\}$, and their commutativity is expressed as $\ell_1 \circ m_1 = \ell_3$ and $\ell_2 \circ m_2 = \ell_3$. The two morphisms $\ell_1 \circ m_1$ and ℓ_3 have the same domain `platform` and the same codomain `chair-as-place-to-sit`, ℓ_3 is a morphism of the triangle $\{m_1, \ell_1, \ell_3\}$, and $\ell_1 \circ m_1$ is a composition of morphisms of $\{m_1, \ell_1, \ell_3\}$. A similar statement can be made about the other triangle, and, hence, stating that they commute is equivalent to stating that the above equations hold. Asserting that the conical array of arrows ℓ_1, ℓ_2, ℓ_3 is a cocone for D is the same as asserting that, together with the morphisms m_1 and m_2 of D, it forms these two commutative triangles.

The second part of the defining property of a colimit for D is that it is a very special cocone called an *initial cocone*. An initial object in any category is an object which, for any other object of the category, is the domain of a unique morphism having the other object as its codomain. An initial cocone is an initial object in a category of cocones. This implies that there are cocone morphisms. Indeed, there is a category associated with any diagram E in any category B whose objects are the cocones for E; if E has no cocones, the category is empty. A cocone morphism $h{:}K \longrightarrow K'$ having a cocone K for E as its domain and a cocone K' for E as its codomain is a morphism $h{:}b \longrightarrow b'$ in category B having the apex b of K as its domain and having the apex b' of K' as its codomain, with the property that h forms a commutative triangle with each pair of leg morphisms of K and K' that share the same domain object in E^2. This is illustrated in Figure 5, where one of the commutative triangles of the cocone morphism is highlighted by double arrows.

Cocones and the intiality property may sound a bit complicated, but in the theory category these notions have an intuitive consequence. The theory category has colimits for all of its diagrams[3]. This fact is associated with a theorem in category theory that in effect specifies how the apex and leg morphisms of a colimit for any diagram can be calculated by an automated process. For the theory category, this process amounts to a symbol-processing algorithm. The mathematical significance of this for concepts can be seen from the commutative triangles of the colimit cocone and from the fact that this cocone has the initiality property. First, notice that the two triangles (m_1, ℓ_1, ℓ_3) and (m_2, ℓ_2, ℓ_3) in the example have a common side, the morphism ℓ_3. Therefore, since $\ell_1 \circ m_1 = \ell_3$ and $\ell_2 \circ m_2 = \ell_3$, then by the law of equality $\ell_1 \circ m_1 = \ell_2 \circ m_2$. From what has been said about theory morphisms, the two morphisms $m_1{:}\mathtt{platform} \longrightarrow \mathtt{chair}$ and $m_2{:}\mathtt{platform} \longrightarrow \mathtt{place-to-sit}$ map the theory `platform` into each of `chair` and `place-to-sit`, but in a separate way in each. That is, although truth is preserved, the abstract shape in `platform` maps to a rectangle in `chair` but to another abstract shape in `place-to-sit`. However, because the two compositions $\ell_1 \circ m_1$ and $\ell_2 \circ m_2$ are one and the same morphism, the two images of the abstract shape of `platform` in `chair` and `place-to-sit` must have the same image in `chair-as-place-to-sit` via ℓ_1 and ℓ_2. The same applies to everything in the theory `platform`, and this formalizes the "blending" of `chair` and `place-to-sit` along their common "subtheory" `platform`. Notice that the image of everything in `chair` and `place-to-sit` that is not an image of something in `platform` will retain its separate identity. For example, theory items specific to the chair's seat back and to the biped sitting postures will retain their separate presence in `chair-as-place-to-sit`.

The foregoing explains the intuition about "blending" or superposition of the images of `platform` in `chair` and `place-to-sit` to form `chair-as-place-to-sit`, but it does not explain the intuition about intiality. For this, note that the cocone pictured in Figure 4 could have been part of a larger cocone having a diagram that includes D. This larger diagram would have included other theories and morphisms; its apex would therefore be a more complex theory having `chair-as-place-to-sit` as a subtheory. It is not difficult to see that there is an infinite number of diagrams that include D as a subdiagram, the number of possible concepts being infinite. However, the colimit cocone being an initial cocone precludes its base diagram being part of a larger cocone, for its initiality means that its apex `chair-as-place-to-sit` can be mapped into the apex of any cocone for the same base diagram (including itself), and in a unique way. This says exactly that `chair-as-place-to-sit` is as advertised a theory that just "blends" `chair` and `place-to-sit` along `platform`, and nothing more.

3 Line Drawings and Colimits

The purpose of the work described in this paper is to test the idea that similarity judgements by humans can be explained in terms of diagrams and colimits of diagrams in the theory category. In the experiment to be described, line drawings are analyzed as models of theories representing concepts about the geometry appearing in the drawings. This is consistent with the mathematical model, in which exemplars of categories are formalized as models of theories that formalize the concepts. In this section, we shall describe the line drawings and delineate their derivation as models of theories which are colimits of diagrams in a category of theories. Given this, a further derivation will result in a similarity measure.

3.1 Diagrams for composite shapes

There are 16 line drawings for the experiment, consisting of complex but symmetrical geometric shapes. These differ in two ways. First, there are four basic shape types, labelled "SAW", "COMPASS", "SPIKES", and

[2] Notice that the same symbol h is used for both the category B morphism and the diagram E cocone category morphism; mathematicians typically overload a symbol in this manner for economy of notation when the meaning is clear.

[3] This property is shared by many important categories.

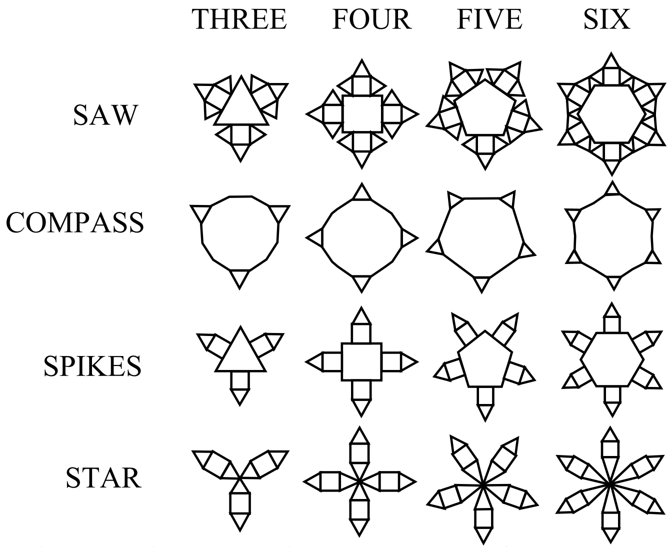

Figure 6: **The line drawings used in the experiment. Rows: The four shape types. Columns: The number of arms per shape.**

"STAR" in Figure 6. Second, each shape type is realized in four shapes, with 3, 4, 5, and 6 arms, respectively, resulting in a total of 16 shapes. The shapes are meant to be reasonably complex so that differences (or, dually, similarities) can vary over a detectable range, and yet simple enough to provide examples that do not introduce confounding effects. The figures are radially symmetric composites of polygons. The shape at the center of each composite is either a point or a polygon, the latter ranging through triangles, squares, pentagons, hexagons, and decagons to dodecahedrons. Each center has from three to six radial arms of identical shape, and these are formed from triangles and squares.

Each composite shape is regarded as an initial model for its theory. All theories share a common subtheory which, as with the theory `platform` which builds on it, consists of a simple Euclidean-like geometry[4] with definitions of certain basic geometric items: points, lines, line segments, planes, angles, and bounded plane regions as before. In this initial investigation, we proceeded without fully accurate, detailed theories[5], letting the experimental results indicate the extent to which accurate theories are necessary to achieve an unambiguous result. As previously noted, our theories are illustrated with pictures as opposed to the symbolic mathematics a full formalism would demand. When it is required for clarity, we add remarks about the missing symbols and their meanings.

Because each of the 16 shapes is a composite of simpler geometric shapes, its theory can be found as a colimit of theories describing the latter shapes. The base diagram for each colimit contains these theories as objects, and its morphisms express the "blending" of the shapes they describe. Notice, for example, that composite shapes of the type SAW (Figure 6) have arms consisting of a square with three adjoined triangles. The adjoining is a part of the "blending" that forms the SAW composite, and is represented as a superposition of each of three edges of the square with each of three triangles. This is a part of the motivation for the diagram of Figure 7. However, notice that while there are three triangles per arm (where there are n arms, with n taking one of the values 3, 4, 5, or 6), the picture depicts only a single theory specific to triangles, there being only one triangle icon in the picture. Notice also that most of the theories appear to depict an abstract shape (a simple closed curve), in the manner of the abstract platform shape in figures 2–4, and this shape has several points on its boundary curve.

[4] Theories of geometry vary depending upon their axioms.

[5] For example, the quantities mentioned such as lines and planes would be defined and axioms postulated in a formal logic theory. We simply assume that this has been done.

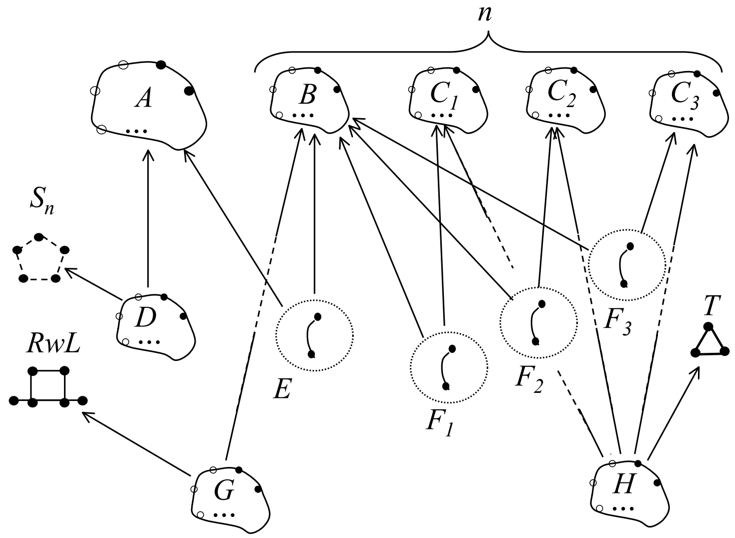

Figure 7: **The base diagram for the derivation of the shape type SAW as a colimit. The n indicates that there are a number n arms, and, hence, the part of the figure under the horizontal brace is to be repeated n times with the exception of theories $G, H,$ the triangle theory, and the arrow having H as its domain and the triangle theory as its codomain.**

These peculiarities are our means of addressing two issues which must be discussed at this point.

First of all, there can be multiple copies of a particular shape in a composite, such as the three triangles in each arm of a SAW. With n arms as indicated in Figure 7, there are $3n$ triangles not counting the single triangle serving as the central mass of the SAW when $n = 3$ (see Figure 8). Because there are n arms, the diagram is understood to have n objects for each of the objects $B, C_1, C_2, C_3, E, F_1, F_2, F_3$ and n morphisms for each of the morphisms $G \longrightarrow B, H \longrightarrow C_1, H \longrightarrow C_2, H \longrightarrow C_3, E \longrightarrow A, E \longrightarrow B, F_1 \longrightarrow B, F_1 \longrightarrow C_1, F_2 \longrightarrow B, F_2 \longrightarrow C_2, F_3 \longrightarrow B,$ and $F_3 \longrightarrow C_3$. In a complete illustration of the diagram, each object would have its own distinct name. For example, there would be n separately-named theories in place of C_1, all having the same structure (that is, content) except with different symbols.

To express the adjoining of each triangle via a diagram, its separate identity must be represented in either a theory or a separate morphism. This is true for all shapes having multiple copies in a composite, such as squares. Suppose, for instance, that multiple triangles were to be represented by separate theories in a diagram. These theories would all share the common geometry subtheory, and also a description of triangles[6]. Each theory specific to a particular triangle in the diagram would also include a triangle constant—a symbolic name representing a particular example of the triangle type. Through the diagram morphisms, the colimit apex theory would also share the common triangle theory, and theories about other shapes as well. Only one subtheory of basic geometry would be present, but descriptions of the distinct types of shapes such as squares and triangles would maintain their separate identities. Of most significance in our discussion, the constants representing the distinct copies of each shape—square or triangle—would also maintain their separate identities. Hence, the colimit theory would contain the basic theory of geometry, descriptions of squares and triangles as constructs formed from the basic geometry, and a constant representing each particular square and triangle. Of course, the same would be true for other shapes present. The resulting colimit theory would then describe the joining of shapes resulting from the "blending" of their theories to form one of the 16 composite shapes, one of the 16 line drawings in our experiment.

[6] The description would explain that a triangle is formed from three line segments meeting at its vertices, or by three intersecting angles (plane regions enclosed by rays emanating from the vertices).

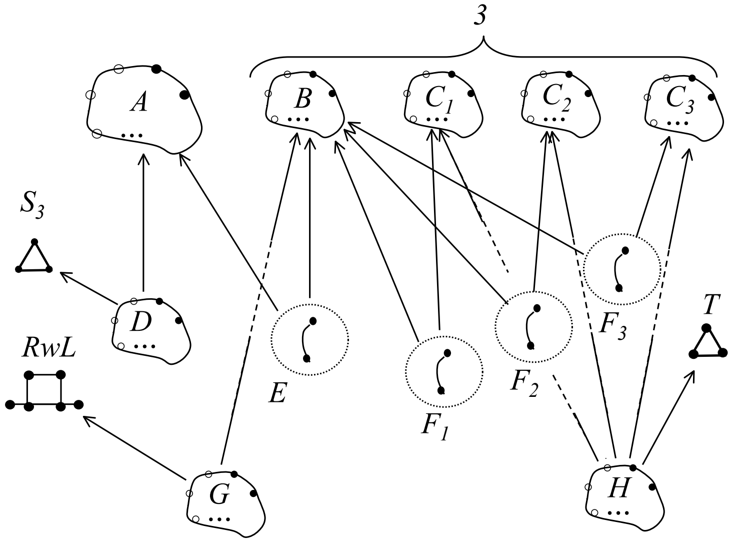

Figure 8: **The base diagram for the derivation of the SAW colimit with 3 arms, $n = 3$ in this case. The center is specified to be a triangle via the morphisms $D \longrightarrow A$ and $D \longrightarrow S_3$.**

A diagram comprising theories for specified points, triangles, squares, etc., and morhisms carefully selected to express the "blending" needed, is one alternative for expressing the construction of the desired composite. Each copy of a shape can be represented in a separate theory by a constant of that shape type, accompanied by a description of how that shape is formed from points, lines, and so forth. This would be one way to address the first issue. However, there is another issue which is of fundamental importance in applying our theoretical model to the derivation of a similarity measure for pairs of composite shapes. To see this, notice that, apparently, any two of the 16 shapes represented in the drawings can be regarded as having some degree of similarity, and in two different ways: They can have the same number of radial arms (a sort of common gestalt), or they can be composites of all or nearly all the same basic shapes (having the same or nearly the same shape features joined in the same or nearly the same manner in each arm). For example, there is a certain similarity between any two of SAW, COMPASS, SPIKES or STAR with three arms, a figure with three radiating arms being the gestalt. On the other hand, two SAWS have a certain similarity since their shape features are the same except for the polygon at the center, which ranges from a triangle (with three arms) to a hexagon (with six arms). The similarity of features is to a large extent addressed by the kind of diagram described in the previous paragraph. However, the gestalt similarity is not addressed by this alternative. Both kinds of similarity must be addressed, because as far as we know either or both of them might be emphasized in the similarity judgements made by a given human subject.

3.2 The two semantic aspects of line drawings

The foregoing explains the use of abstract shapes in the diagrams for our line drawings. Not only the semantics of the features, but also the semantics of the layout of the drawing must be present in order to capture all semantic aspects that can contribute to similarity judgements. The purpose of the abstract-shape theories and their morphisms is to express the second aspect, the layout of the components in a composite shape. This is one of the two aspects of its semantics that can affect similarity judgements. To combine both aspects of similarity—feature semantics and layout semantics—in a single diagram whose colimit is to express the composite shape in a line drawing, the abstract shape theories must be included. However, their inclusion modify the diagram in a fundamental way, and demands their use for the specification of both aspects of the semantics of similarity. Let us examine abstract shape theories and their uses in more detail.

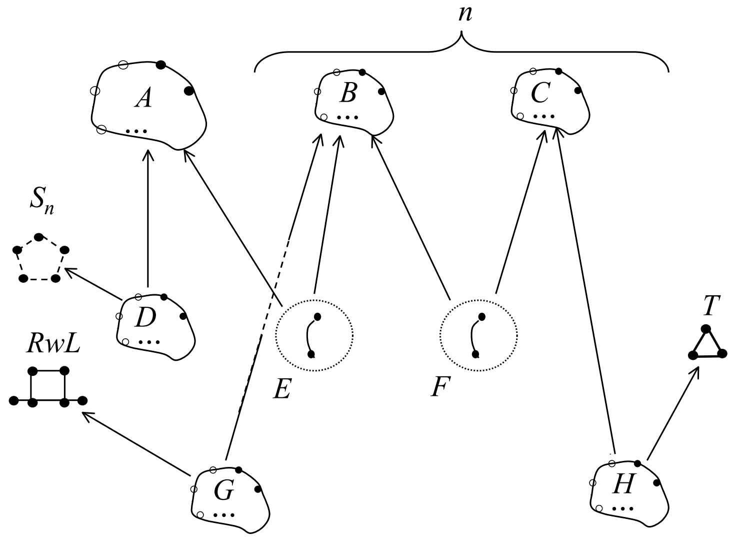

Figure 9: **The base diagram for the derivation of the SPIKES colimit.**

In figures 7 and 8, there are theories labelled $A, B, C_1, C_2, C_3, D, G,$ and H, all copies of a theory specifying an abstract shape. Figure 9 shows the simpler diagram for SPIKES, where the abstract shape theories C_1, C_2, C_3 have been replaced in the diagram by the single abstract shape theory C. There are abstract shape theories also in the diagrams for COMPASS and STAR, but Figure 9 will serve to illustrate their properties and the use made of them. The shape in each abstract shape theory is specified by a constant giving it its own unique name, and its boundary—a closed curve—has several points, also specified by constants. There are more of these point constants than there are vertices on any of the polygonal shapes which constitute the feature semantics of the line drawings. There is no requirement, however, that the point constants in an abstract shape theory all represent distinct points. This allows a morphism having the abstract shape theory as its domain to map several of the point constants to the same point constant in its codomain theory. This is useful because the abstract shapes are used in expressing both the feature and layout semantics, in a single diagram.

A polygonal shape theory describes a specific geometric shape. Consequently, the theory specifies, in particular, that its point constants represent distinct vertices. When the codomain of a morphism whose domain is an abstract shape theory is a polygonal shape theory, its vertices will be the images of the point constants. This is where the ability to map several boundary points in an abstract shape theory to a single point in another shape theory becomes useful. The abstract and polygonal shape theories are constructed so that there will always be fewer vertices specified in the polygonal shape theories than there are point constants in the abstract shape theories. This allows an abstract shape theory to map to a polygonal shape theory, regardless of the number of vertices of the polygon specified by the latter theory. This occurs, for example, in the morphism `platform` $\longrightarrow$ `chair`, where the chair polygon (the seat platform) has only four vertices. The excess over 4 of point constants in the abstract shape specified in `platform` all map to a single vertex of the rectangle specified in `chair`.

As for the abstract-shape morphisms, notice that the morphisms[7] $D \longrightarrow A,\ G \longrightarrow B,$ and $H \longrightarrow C$ in Figure 9 map one abstract shape theory to another. Each of these morphisms maps each specified point on the boundary curve of the shape specified in its domain to the corresponding point on the image boundary curve of the shape specified in its codomain. In fact, the morphisms are one-to-one mappings and amount to simply changing the names of items such as constants[8].

Let us examine the morphisms that express the layout semantics. They specify the layout as a combining

[7] Morphism names are not necessary since there is at most one morphism for any two objects in the diagram.

[8] Were we presenting all the category-theoretic details, we would call these mappings *isomorphisms*.

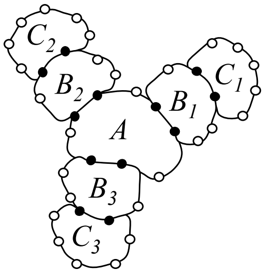

Figure 10: **The abstract SPIKES shape for** $n = 3$

of abstract shapes to form each arm and attach it to the center part of the composite shape. Notice that each of E, F_1, F_2, F_3 is a theory describing an abstract curve segment with specified end points (point constants). To attach the arm to the central mass, a segment of the closed curve specified in E, bounded by two specified points, is to be superimposed upon one of the segments of the closed curve in A, also bounded by two specified points. The same segment specified in E is to be superimposed upon upon one of the segments of the closed curve specified in B. This join operation is specified by the two morphisms $E \longrightarrow A$ and $E \longrightarrow B$, which, when composed with the appropriate colimit leg morphisms in the defining diagram of a colimit, force the two images of the curve segment of E in A and B to have the same image in the colimit theory. This effectively fuses together or joins the abstract shapes specified in A and B along a curve segment. All the joins of abstract shapes needed to express the layout have this form, with a theory specifying a curve segment and its end points mapped to segments on the boundaries of the abstract shapes in two shape theories via two morphisms. Were the abstract-shape joins the only content of the diagram of Figure 9, the colimit theory would specify the composite of abstract shapes shown in Figure 10, which illustrates the "abstract SPIKES" with $n = 3$.

The feature semantics of the composite shape in each line drawing is specified by the other morphisms in its base diagram. In the colimit of the diagram of Figure 9 specifying SPIKES, the morphisms $G \longrightarrow RwL$ and $G \longrightarrow B$ have the effect of specifying that the abstract shape specified in B is to be a square occupying the middle of a line segment. The pair of join morphisms $E \longrightarrow A$ and $E \longrightarrow B$ specifies that the curve segment along which the shapes from A and B are joined is the curve segment from B that matches that same line segment from RwL. Simultaneously, $D \longrightarrow S_n$ and $D \longrightarrow A$ specify that the image of A is a polygon, its type depending upon the number n of arms specified in the actual diagram. Also, the curve segment along which A and B are joined is a face of the polygon. The net result of all the morphisms in the diagram is that in the colimit theory, each arm is specified to have at its base a square attached in the middle portion of a face of the polygon. Similarly, $H \longrightarrow C$ and $H \longrightarrow T$ specify that the outer part of each arm is to be a triangle.

Together, the abstract shape morphisms and the morphisms associating abstract shape with specific shape express the two semantic aspects of a line drawing—its layout and its component shapes, which we have called features of the composite shape. Both contribute to similarity or dissimilarity for pairs of line drawings . How, then, do we use this information to derive a similarity measure for pairs of line drawings?

4 A Diagrammatic Similarity Measure

Category theory provides a mathematically precise framework for expressing concepts, their relationships, their categories of exemplars (which can be seen as mathematical categories of models), and a construction that automates the derivation of concepts as colimits of diagrams of their component concepts. In the previous section, this machinery was applied to the two-pronged semantics of complex objects in line drawings. The complex objects were represented as colimits of diagrams of concepts that express their parts, both as abstract shapes and as specific geometric shapes, and the manner in which the parts are related in forming the complex shape. Based upon the idea that this scheme is a good approximation to the neural structures and operations that represent these concepts in the human brain, the scheme can be applied to similarity judgements made by humans. The basic idea is that similarity judgements for pairs of line drawings reflect the "amount of sameness" in the diagrams from which their complex shapes are derived as colimits. Let us make this idea of sameness more precise.

4.1 Morphisms in common

Our similarity measure is based upon the idea that it is the morphisms more than the objects that determine a category. There are many examples in category theory texts that illustrate this proposition [16, 2, 3, 10, 11], but the reasoning behind it is simple. First, the objects of a given type can be related in different ways. For example, the positive integers $1, 2, 3, 4, \ldots$ are related by the less than or equal relation $\leq$ and also by the divisibility relation $|$, where $n \,|\, m$ exactly when n divides m evenly with no remainder, and otherwise n and m are unrelated. The latter kind of relation is called a *partial order*; in the present case, two positive integers are not always related. The former kind of relation is a total order, since one number is always less than or equal to another. Second, as stated in a previous section, it is almost universally understood in category theory that a morphism includes its domain and codomain objects. Thus, a category depends upon its morphisms and is determined by them. The positive integers yield two such categories, where each of the relations can be described as a system of morphisms and the composition of morphisms in each category is simply the transitive property (if n is related to m and m is related to r, then n is related to r) which holds for both relations. Based upon the fact that the morphisms determine a category, it is reasonable to base the similarity of two diagrams upon the relative number of morphisms they share. We can call this their morphism overlap, or, simply, their overlap. Representing overlap as the intersection of their two sets of morphisms, we can write down the set of morphisms in each diagram and then form the intersection of the two sets. Then, we can count the morphisms in the intersection. There is more to say about this, but to say it let us proceed mathematically through examples.

Consider two of our experimental shapes: SAW and SPIKES. Let us examine two examples: One in which SAW is the shape type for both diagrams and they differ in the number of arms, $n = 3$ in one diagram and $n = 4$ in the other, and one in which one diagram is for SAW with $n = 3$ and the other diagram is for SPIKES with $n = 4$. Now as discussed before, the diagrams used here have two types of morphisms: those involving only abstract shape theories, and those involving a polygonal shape theory.

For the first example, from Figure 7 the intersection of morphism sets includes, first, all the abstract-shape morphisms. There is one for specifying the polygon at the center; this is $D \longrightarrow A$ (the other morphism involves the specific polygon). Then, since the two diagrams have the morphisms for three arms in common, there are three morphisms for specifying the polygonal shape at the base of each of three arms: $G \longrightarrow B_1, G \longrightarrow B_2, G \longrightarrow B_3$, where the subscripts on B indicate the separate arms. Next, there are three for each arm for specifying the three polygons attached to the base of each arm. Running through triples of subscripts for each arm, these are $H \longrightarrow C_1, H \longrightarrow C_2, H \longrightarrow C_3, H \longrightarrow C_4, H \longrightarrow C_5, H \longrightarrow C_6, H \longrightarrow C_7, H \longrightarrow C_8, H \longrightarrow C_9$. Finally, there are the morphisms whose domain is an abstract curve segment; these attach the shapes to form the layout: $E_1 \longrightarrow A, E_1 \longrightarrow B_1, F_1 \longrightarrow B_1, F_1 \longrightarrow C_1, F_2 \longrightarrow B_1, F_2 \longrightarrow C_2, F_3 \longrightarrow B_1$, and $F_3 \longrightarrow C_3, E_2 \longrightarrow A, E_2 \longrightarrow B_2, F_4 \longrightarrow B_2, F_4 \longrightarrow C_4, F_5 \longrightarrow B_2, F_5 \longrightarrow C_5, F_6 \longrightarrow B_2$, and $F_6 \longrightarrow C_6, E_3 \longrightarrow A, E_3 \longrightarrow B_3, F_7 \longrightarrow B_3, F_7 \longrightarrow C_7, F_8 \longrightarrow B_3, F_8 \longrightarrow C_8, F_9 \longrightarrow B_3$, and $F_9 \longrightarrow C_9$. Now, since a generally-applicable procedure will save time in calculation, let us begin concocting formulae for counting the morphisms in an intersection. For the abstract morphisms in the intersection of sets for two diagrams for SAW with different numbers of arms, the formula is as follows: The intersection will include all abstract morphisms for the minimum of the two numbers, which is 3 in this case, $\min(3,4) = 3$. For SAW with 3 and 4 arms, respectively, the number of abstract-shape morphisms is 1 for the central polygon plus, for each arm, 1 for the base polygon, 3 for the polygons attached to the base, and 2 for each pair of shapes that must be attached. Attaching the base to the center and attaching each outer shape to the base makes 8 shape-attaching morphisms, for a total of 12 abstract morphisms per arm. The total number of abstract morphisms in the intersection is $1 + 3 \cdot (1 + 3 + 8) = 1 + 3 \cdot 12$. Generalizing by applying the min operation to 3 and 4, this is $1 + \min(3,4) \cdot 12$; generalizing further, to n_1 versus n_2 arms and letting juxtaposition of numbers represent multiplication, this can be written compactly as $1 + 12\min(n_1, n_2)$.

Now, for the morphisms involving a specific polygonal shape theory, in Figure 7 there is 1 per polygonal shape per position in the composite, for a total of 3 (those whose domains are the abstract shape theories D, G and H). However, in changing the number of arms, a glance at Figure 6 shows that changing the number of arms results in a change in the center polygon, whose theory is labelled S_n in Figure 7 because it depends upon the number of arms. Therefore, the center specific-polygon morphism $D \longrightarrow S_n$ will not appear in the intersection of morphism sets; this means that there are 2 polygon-specific morphisms in the intersection. The final morphism count is

$$\begin{aligned}\text{SAW3SAW4} &= \text{abstract shape} + \text{polygon specific} \\ &= (1 + 12\min(n_1, n_2)) + 2 \\ &= 3 + 12\min(n_1, n_2) \quad .\end{aligned}$$

Since $\min(n_1, n_2) = 3$ in this case, the actual morphism count for the intersection is

$$\text{SAW3SAW4} = 3 + 12 \cdot 3 = 39 \quad .$$

But let us generalize the morphism count formula to the case SAWn_1 SAWn_2 for any pair (n_1, n_2), including the case when the number of arms does not change (comparing a line drawing with itself). In the latter case, we must allow for the inclusion of the morphism $D \longrightarrow S_n$ because the number of arms is the same in both SAW line drawings and, hence, the center polygon is the same. Let $\{n_1 = n_2\}$ denote a function which compares n_1 and n_2 and takes the value 1 if they are equal and 0 if they are not. Then the formula we seek for the general case, comparing two SAW line drawings, is

$$\text{SAW}n_1\text{SAW}n_2 \quad = 3 + 12\min(n_1, n_2) + \{n_1 = n_2\} \quad .$$

In the second example, one diagram is for SAW with $n = 3$ and the other diagram is for SPIKES with $n = 4$. As in the first example, the intersection of morphism sets from the two diagrams will retain morphisms for 3 arms, or in our notation $\min(n_1, n_2)$ where $n_1 = 3$ and $n_2 = 4$. Unlike the first example, the retained 3 arms will also retain only the morphisms for SPIKES arms, since SPIKES has 1 triangle per arm by contrast with SAW, which has 3 triangles per arm. Notice that the diagrams for SAW and SPIKES differ only in this one respect. As it happens, the intersection of morphism sets is that of the first example except that the many morphisms per arm, $H \longrightarrow C_1, H \longrightarrow C_2, H \longrightarrow C_3, H \longrightarrow C_4, H \longrightarrow C_5, H \longrightarrow C_6, \ldots$ and $F_1 \longrightarrow B_1, F_1 \longrightarrow C_1, F_2 \longrightarrow B_1, F_2 \longrightarrow C_2, F_3 \longrightarrow B_1$, and $F_3 \longrightarrow C_3, F_4 \longrightarrow B_2, F_4 \longrightarrow C_4, F_5 \longrightarrow B_2, F_5 \longrightarrow C_5, \ldots$, are eliminated in favor of fewer morphisms per arm. The new intersection of morphism sets is as follows: As before, the abstract $D \longrightarrow A$ will be present, along with the three $G \longrightarrow B_1, G \longrightarrow B_2, G \longrightarrow B_3$, one for each arm. However, now there is only one for each arm for specifying the attached polygons (which turn out to be triangles), $H \longrightarrow C_1, H \longrightarrow C_2, H \longrightarrow C_3,$. Here, we use the indexing notation of the SPIKES diagram; that is, $C_1, C_2,$ and $C_3,$ are no longer associated with one single arm; instead, each is associated with an arm of its own. Finally, and again using the indexing notation of the SPIKES diagram, the morphisms whose domain is an abstract curve segment are $E_1 \longrightarrow A, E_1 \longrightarrow B_1, F_1 \longrightarrow B_1, F_1 \longrightarrow C_1, E_2 \longrightarrow A, E_2 \longrightarrow B_2, F_2 \longrightarrow B_2, F_2 \longrightarrow C_2, E_3 \longrightarrow A, E_3 \longrightarrow B_3, F_3 \longrightarrow B_3,$ and $F_3 \longrightarrow C_3$. As for the polygon-specific morphisms, they are the same as in the previous example: $G \longrightarrow RwL, H \longrightarrow T$ and, if $n_1 = n_2, D \longrightarrow S_n$. Counting the morphisms,

$$\text{SAW3SPIKES4} = 3 + 6\min(n_1, n_2) = 3 + 6 \cdot 3 = 21 \quad .$$

Again generalizing the formula to cover all SAW-SPIKES diagram comparisons,

$$\mathrm{SAW}n_1\mathrm{SPIKES}n_2 = 3 + 6\min(n_1, n_2) + \{n_1 = n_2\} \quad .$$

4.2 Derivation of the similarity measure

We ended each of the two examples in the last section by presenting a general formula. The formulae apply to all cases, of SAW with n_1 arms versus SAW with n_2 arms, and of SAW with n_1 arms versus SPIKES with n_2 arms, respectively. In each case, the calculated number increases with either $\min(n_1, n_2)$ or $\{n_1 = n_2\}$. This has the effect of assigning a roughly equal weight to two alternatives, the two shapes having the same number of arms and both shapes having an increase in the number of arms. Having $\{n_1 = n_2\} = 1$ means that both numbers are the same and the formula gives this one point. Having a larger $\min(n_1, n_2)$ makes it more likely that the numbers are close to being the same, since they are upper-bounded by 6 (there are at most 6 arms in a shape), and the formula gives this a number of points equal to the minimum itself. But the value of the minimum being larger also puts weight on the magnitude of the minimum. Also, noting the SAW versus SPIKE comparisons, the calculated numbers for the SAW versus SAW and SAW versus SPIKES comparisons favor SAW versus SAW because the arms in both diagrams of a pair are more complex, and, hence, more morphisms are retained in the intersection of their morphism sets. These effects would seem intuitively reasonable for a similarity measure except for two considerations: First, assigning roughly equal weight to a having a larger number of arms for both figures as well as having a larger number of arms in common may confound the similarity of the shapes with their complexity in terms of the number of arms. Two figures of higher complexity may or may not be in some sense more similar than two figures of lower complexity; there would seem to be room for argument on this point. In any case, having the minimum of the two numbers of arms approach its upper bound of 6 would seem the most reasonable role for increased complexity to play in a similarity judgement, for then the two numbers must approach equality. The second consideration is perhaps the most important of the two: The calculated numbers are not normalized, and this calls into question their reliability as similarity values. When concocting a measure of similarity, it is well to ask: Similarity with respect to what? For example, note that $\mathrm{SAW3SAW3} = 40$ while $\mathrm{SAW3SAW4} = 39$ and $\mathrm{SAW4SAW4} = 52$. Is it justifiable to have only a small gap between the similarity values for SAW3 with itself and SAW with 3 versus 4 arms, and such a large gap between the similarity values for SAW3 with itself and SAW4 with itself?

Clearly, the effect of sheer size needs to be mitigated in concocting a similarity measure. Even more important is the effect of complexity, for it raises a troublesome issue and has impacts beyond those in the examples just analyzed. The effect of complexity has already appeared in SAW versus SAW as compared with SAW versus SPIKES. What happens when we derive the formulae for the other comparisons, say, for example, SAW versus COMPASS?

The question, Similar with respect to what?, has yet more ramifications. As the literature shows, similarity can have many definitions based upon many sets of criteria. Perhaps a standard is needed that applies across criteria. Normalization seems an intuitively reasonable one and has been applied many times. For all the foregoing reasons, our similarity measure is a normalized version of the formulae suggested by the analysis of the last section. To be specific, we can divide the numbers we have been calculating, the number of morphisms two diagrams have in common, by the sum of the numbers of morphisms in both diagrams. Similarity will then be judged by the number of morphisms two diagrams have in common relative to their total number of morphisms. This is expected to reduce the effects of sheer size and complexity and to emphasize that part of their layout and those features that two shapes have in common.

In the language of finite sets[9], the intersection of sets A and B is denoted $A \cap B$, their union by $A \cup B$, and their symmetric difference (the number of elements in their union exclusive of the number of elements in their intersection) by $A \Delta B$[10]. For our normalized formulae, we desire the cardinality (number of elements)

[9] To be correct, this analysis needs to be framed in terms of category theory rather than set theory. However, the categorical analysis requires more background than can be supplied here. The discussion in terms of sets provides a workable approximation to the truth.

[10] Notation for this quantity varies.

of $A \cup B$, denoted $\mathrm{card}(A \cup B)$. Since $\mathrm{card}(A \Delta B) = \mathrm{card}(A \cup B) - \mathrm{card}(A \cap B)$, this can be calculated as $\mathrm{card}(A \cup B) = \mathrm{card}(A \Delta B) + \mathrm{card}(A \cap B)$. We can accomplish this by first calculating the number $\mathrm{card}(A \cap B)$, which is exactly what our intersection formulae do, and then adding in the number of morphisms from each diagram that were excluded, which is the $\mathrm{card}(A \Delta B)$ term.

The result of this analysis is the following similarity measure for each of the previous examples. For the case $\mathrm{SAW}n_1\,\mathrm{SAW}n_2$, where as usual $|x|$ denotes the absolute value of the number x, the excluded morphisms are just the ones for the missing arms and also the polygon-specific morphisms for the two centers provided the center differs between the two shapes (that is, provided the number of arms differs). The resulting number we shall call $\Delta M(\mathrm{SAW}n_1\mathrm{SAW}n_2)$, where $\Delta M(\mathrm{SAW}n_1\mathrm{SAW}n_2) = 12|n_1 - n_2| + 2(1 - \{n_1 = n_2\})$. Notice that the second term is nonzero just in case the number of arms is NOT the same. The previously-derived intersection formula we shall call $\cap M(\mathrm{SAW}n_1\mathrm{SAW}n_2)$. The similarity measure for this case is then computed as follows:

$$\begin{aligned}
\cap M(\mathrm{SAW}n_1\mathrm{SAW}n_2) &= 3 + 12\min(n_1, n_2) + \{n_1 = n_2\} \quad , \\
\Delta M(\mathrm{SAW}n_1\mathrm{SAW}n_2) &= 12|n_1 - n_2| + 2(1 - \{n_1 = n_2\}) \quad , \\
M(\mathrm{SAW}n_1\mathrm{SAW}n_2) &= \frac{\cap M(\mathrm{SAW}n_1\mathrm{SAW}n_2)}{\Delta M(\mathrm{SAW}n_1\mathrm{SAW}n_2) + \cap M(\mathrm{SAW}n_1\mathrm{SAW}n_2)} \quad .
\end{aligned}$$

Notice that

1. The similarity value $M(\mathrm{SAW}n_1\mathrm{SAW}n_2)$ is normalized, $0 < M(\mathrm{SAW}n_1\mathrm{SAW}n_2) \leq 1$, and
2. $M(\mathrm{SAW}n_1\mathrm{SAW}n_2) = 1$ just in case $n_1 = n_2$, which makes perfect sense.

For the case $\mathrm{SAW}n_1\mathrm{SPIKES}n_2$, the formula for ΔM depends on two new considerations: First, the $\min(n_1, n_2)$ arms that contribute to $\cap M$ have 6 morphisms missing from the SAW arms that match the SPIKES arms to form the intersection. We need to account for the missing SAW morphisms. Second, when $n_1 \neq n_2$, either SAW has more arms ($n_1 > n_2$) or SPIKES has more arms ($n_1 < n_2$). We need to account for the extra arms of either SAW or SPIKES, and to do this in a single, general formula, we need a new term. To this end, let

$$[x] = \begin{cases} x & \mathrm{if} x \geq 0 \quad , \\ 0 & \mathrm{otherwise} \quad . \end{cases}$$

Then, after performing the bookkeeping, we can see that the similarity measure for SAW versus SPIKES has the following form:

$$\begin{aligned}
\cap M(\mathrm{SAW}n_1\mathrm{SPIKES}n_2) &= 3 + 6\min(n_1, n_2) + \{n_1 = n_2\} \quad , \\
\Delta M(\mathrm{SAW}n_1\mathrm{SPIKES}n_2) &= 6\min(n_1, n_2) + 12[n_1 - n_2] + 6[n_2 - n_1] \\
&\quad +2(1 - \{n_1 = n_2\}) \quad , \\
M(\mathrm{SAW}n_1\mathrm{SPIKES}n_2) &= \frac{\cap M(\mathrm{SAW}n_1\mathrm{SPIKES}n_2)}{\Delta M(\mathrm{SAW}n_1\mathrm{SPIKES}n_2) + \cap M(\mathrm{SAW}n_1\mathrm{SPIKES}n_2)} \quad .
\end{aligned}$$

Notice that, again, the similarity measure is normalized; however, because of the $6\min(n_1, n_2)$ term, it is not unity when $n_1 = n_2$. But again the latter property is sensible, because SAW and SPIKES are never alike.

4.3 The similarity formulas

In this section, we list the similarity formulas for all cases. In actuality, we simply list the ΔM and $\cap M$ formulas: The similarity formula M is always

$$M = \frac{\cap M}{\Delta M + \cap M} \quad .$$

SAWn_1SAWn_2:
$\cap M = 3 + 12\min(n_1, n_2) + \{n_1 = n_2\}$
$\Delta M = 12\,|n_1 - n_2| + 2\,(1 - \{n_1 = n_2\})$

SAWn_1COMPASSn_2:
$\cap M = 1 + 3\min(n_1, n_2)$
$\Delta M = 9\min(n_1, n_2) + 12\,[n_1 - n_2] + 3\,[n_2 - n_1] + 5$

SAWn_1SPIKESn_2:
$\cap M = 3 + 6\min(n_1, n_2) + \{n_1 = n_2\}$
$\Delta M = 6\min(n_1, n_2) + 12\,[n_1 - n_2] + 6\,[n_2 - n_1] + 2\,(1 - \{n_1 = n_2\})$

SAWn_1STARn_2:
$\cap M = 3 + 5\min(n_1, n_2)$
$\Delta M = 11\min(n_1, n_2) + 12\,[n_1 - n_2] + 9\,[n_2 - n_1] + 3$

COMPASSn_1COMPASSn_2:
$\cap M = 2 + 3\min(n_1, n_2) + \{n_1 = n_2\}$
$\Delta M = 3\,|n_1 - n_2| + 2\,(1 - \{n_1 = n_2\})$

COMPASSn_1SPIKESn_2:
$\cap M = 2 + 3\min(n_1, n_2)$
$\Delta M = 3\min(n_1, n_2) + 3\,[n_1 - n_2] + 6\,[n_2 - n_1] + 5$

COMPASSn_1STARn_2:
$\cap M = 2 + 3\min(n_1, n_2)$
$\Delta M = 8\min(n_1, n_2) + 3\,[n_1 - n_2] + 9\,[n_2 - n_1] + 3$

SPIKESn_1SPIKESn_2:
$\cap M = 3 + 6\min(n_1, n_2) + \{n_1 = n_2\}$
$\Delta M = 6\,|n_1 - n_2| + 2\,(1 - \{n_1 = n_2\})$

SPIKESn_1STARn_2:
$\cap M = 3 + 5\min(n_1, n_2)$
$\Delta M = 5\min(n_1, n_2) + 6\,[n_1 - n_2] + 9\,[n_2 - n_1] + 2$

STARn_1STARn_2:
$\cap M = 4 + 9\min(n_1, n_2)$
$\Delta M = 9\,|n_1 - n_2|$

These formulas were used to derive the computed similarities for the experiment. The computed similarities were to be compared with the similarity judgements made by human subjects. The experiment and results are described in the next section.

4.4 A final caveat

In the foregoing analysis, only morphisms are used to derive the similarity formulas. The diagram objects are simply considered part of the morphisms for which they serve as domains or codomains. This is justified by the fact that it is really the morphisms that determine a category. However, our use of this justification neglects one property of diagrams. That is, to be fully correct according to category theory, diagrams include not only the morphisms typically shown, but also the identity morphisms and also those obtained by composition of the diagram morphisms. These are typically not shown because (a) they are given based upon what is shown and (b) it is inconvenient to show them because the diagrams typically shown are there to make a point and it is desirable not to clutter an illustration associated with them. The effect of including these neglected morphisms can be investigated in a future effort. However, including them would be an unusual practice, and it was decided to try achieving a positive experimental result without adding them in.

5 Empirical Work

In this section of the paper we report the results of an empirical study that compared human judgments of similarity to similarity measures derived from category theory. In order for category theory to be viewed as a viable model of human cognition, its predictions of basic cognitive processes, such as similarity, must be empirically validated.

5.1 Stimuli

As mentioned previously, we created a set of 16 geometrical shapes to serve as visual stimuli (see Figure 6). The shapes were created using the colimit construct described in Section 3 and were used to compute similarity judgments for pairs of shapes (pairs of line drawings). The computed similarity judgment for a pair of complex shapes is based upon the relative amount of common structure in their base diagrams. The relative amount of common structure is defined as the relative number of morphisms the two base diagrams have in common. In this experiment, we tested not only the applicability of category theory, but our application of it based upon our definition of relative amount of common structure.

The diagrams were carefully constructed so that the 16 stimuli could be classified into four groups in either of two ways. The first was by overall, complex shape structure (labeled saw, compass, spikes, and star), with four stimuli in each each group. Within each group, the four stimuli varied by the number of arms (3, 4, 5, or 6) emanating from the center of each figure. The second method of categorizing the stimuli was by number of arms (3, 4, 5, or 6), with four shapes (saw, compass, spikes, and star) in each group. This was done so that the stimuli would be simple enough to avoid confounding effects due to multiple hidden variables, yet numerous enough to contain multiple exemplars in each grouping and complex enough to allow hypotheses about individual grouping preferences to be tested.

5.2 Participants

There were 53 undergraduate students at the University of New Mexico who voluntarily participated in the study for class credit.

5.3 Procedure

Participants first read and signed a statement of informed consent. They were then seated in a small room with a personal computer. Participants viewed pairs of geometrical shapes presented on the computer screen and rated the similarity of each pair by selecting a number along a rating scale from 1 (least similar) to 5 (most similar). They selected the rating values with a mouse click and could change their selection while the pair remained on the screen. When they were satisfied with their judgment they clicked a Continue button to clear the screen and present the next pair of shapes. Participants were instructed to make quick, intuitive judgments of similarity and not to spend much time viewing the shapes. Participants took around 30 minutes on average to complete the task.

Each participant viewed the 120 $(16 \cdot 15/2 = 120)$ unique pairs of the 16 geometric shapes plus 15 of these pairs chosen at random and repeated at the end of the sequence to give a total of 135 stimulus pairs. The additional 15 pairs were used for a reliability check. The 120 pairs of stimuli were presented in a unique random order for each participant. Participants were not informed that there would be repeated pairs at the end of the task.

6 Results

The reliability of each participant was obtained by computing the correlation (all correlations reported here are Pearson correlation coefficients) between the first and second sets of ratings for the 15 repeated stimulus pairs. The reliabilities ranged from 0.18 to 0.96 with a mean of 0.64. Nine participants with reliabilities below 0.50 were dropped from the study leaving 44 participants for the following analyses.

We first examined how well the participants agreed with one another on their perceived similarity of the geometric shapes. We correlated each participants 120 ratings with every other participants ratings, and then for each participant, we defined an agreement index as the average correlation between that individuals ratings and the individual ratings of the remaining 43 participants. These agreement indices ranged from 0.46 to 0.63, with a mean value of 0.55. This measure of overall agreement among peoples judgments of perceived similarity established a baseline against which to compare other models of similarity. The variability inherent in human judgments of similarity establishes an upper bound on how well a formal model can predict these judgments.

The diagram correspondence criterion defined in Section 4 was used to compute the similarity between each pair of the 16 geometric shapes based upon the correspondence of their base diagrams. To determine how well these category-theory-derived judgments of similarity matched human perception, we correlated the 120 derived similarities with each participants similarity ratings. These correlations ranged from 0.24 to 0.77 with a mean of 0.51 and a standard deviation of 0.11.

For comparison with an alternative similarity judgement method as mentioned in the Introduction, we encoded each of the 16 geometric shapes as a two-element vector consisting of the numbers 1 - 4. The values 1, 2, 3, and 4 were assigned to star, spikes, saw, and compass, respectively for the first dimension; the values 1, 2, 3, and 4 were assigned to three, four, five, and six, respectively for the second dimension. Hence the compass with five arms had vector representation (3, 4). Euclidean distance was computed between pairs of stimuli using their vector representations and the resulting pairwise distances were then correlated with the human judgments of similarity. These correlations ranged from -0.39 to -0.63 with a mean of -0.52 and a standard deviation of 0.06.

7 Discussion

The similarities derived from applying colimits to the geometric shapes provided a good fit to the human judgments of similarity. The mean correlation of 0.51 between the similarities derived from colimits and those given by the participants needs to be interpreted in light of the amount of variance in the ratings across the participants. The correlation of 0.55 (the mean agreement among the participants) indicates that only about 30 per cent $((.55)^2)$ of the variance in the ratings can be accounted for by a linear relation. In light of this, a correlation of 0.51 with the category-theoretic model is seen to be quite high.

Although the similarities derived from the creation of feature vectors correlated highly as well, it should be noted that these feature vectors were derived in an ad hoc manner by attempting to identify the most psychologically salient features of the stimuli. In contrast, the colimit similarities were derived from a mathematical theory of structure together with a means of calculating numbers to represent the correspondence between diagrams expressing shape structure. We propose that the latter approach explicates the semantics of the cognitive process underlying similarity judgements in categorization.

8 Conclusion

Hopefully, the foregoing exposition has suggested the value of a mathematical model of concepts and their categories of exemplars. Admittedly, introducing mathematics to a subject can itself be a complicating factor because of the time and attention required to absorb it. For many, this is especially true when the mathematical discipline used is novel and the application unusual. Category theory is new to most scientists and is an unfamiliar discipline even for many mathematicians, although it has long been applied in certain branches of mathematics (notably algebraic topology) and more recently has seen increasing application in many fields. We propose that the value lies in the advantages of having a mathematical model in any science, and that cognitive science is no exception. In our view, category theory is an appropriate mathematics for the subject matter of cognitive science, although not to the exclusion of other fields of mathematics; indeed, category theory offers a capacity to work together with any branch of mathematics. A more detailed mathematical exposition of theories, their morphisms, colimits, and other notions of category theory is given in any of [3, 9, 6, 19, 20, 7, 5]. The latter two references explore examples of theory colimits and other notions of our theoretical model.

References

[1] W. H. Acton, P. J. Johnson, and T. E. Goldsmith. Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*, 86:303–311, 1994.

[2] Jiri Adamek, Horst Herrlich, and George Strecker. *Abstract and Concrete Categories*. Cambridge University Press, 1990.

[3] Roy L. Crole. *Categories for Types*. Cambridge University Press, 1993.

[4] T. E. Goldsmith and D. M. Davenport. Assessing structural similarity of graphs. In Schvaneveldt [17], pages 75–87.

[5] Michael J. Healy and Thomas P. Caudell. Generalized lattices express parallel distributed concept learning. In Vassilis G. Kaburlasos and Gerhard X. Ritter, editors, *Computational Intelligence Based on Lattice Theory*, volume 67 of *Studies in Computational Intelligence*, pages 59–77. Springer, 2007.

[6] Michael J. Healy and Keith E. Williamson. Applying category theory to derive engineering software from encoded knowledge. In Teodor Rus, editor, *Algebraic Methodology and Software Technology, 8th International Conference, AMAST 2000, Iowa City, Iowa, USA, Proceedings*, pages 484–498. Springer-Verlag, 2000. Lecture Notes in Computer Science, Volume 1816.

[7] Michael John Healy and Thomas Preston Caudell. Ontologies and worlds in category theory: Implications for neural systems. *Axiomathes*, 16(1-2):165–214, 2006.

[8] P. J. Johnson, T. E. Goldsmith, and K. W. Teague. Locus of the predictive advantage in pathfinder-based representations of classroom knowledge. *Journal of Educational Psychology*, 86(4):617–626, 1994.

[9] R. Jullig and Y. V. Srinivas. Diagrams for software synthesis. In *Proceedings of KBSE '93: The Eighth Knowledge-Based Software Engineering Conference*, pages 10–19. IEEE Computer Society Press, 1993.

[10] F. W. Lawvere and S. H. Schanuel. *Conceptual Mathematics: A First Introduction to Categories*. Cambridge University Press, 1995.

[11] Saunders Mac Lane. *Categories for the Working Mathematician*. Springer-Verlag, 1971.

[12] John Macnamara and Gonzalo E. Reyes, editors. *The Logical Foundations of Cognition*. Oxford University Press, 1994.

[13] Barbara C. Malt and Eric C. Johnson. Do artifact concepts have cores? *Journal of Memory and Language*, 31:195–217, 1992.

[14] Douglas L. Medin. Concepts and conceptual structure. *American Psychologist*, 44(12):1461–1481, 1989.

[15] Gregory L. Murphy and Edward J. Wisniewski. Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4):572–586, 1989.

[16] B. C. Pierce. *Basic Category Theory for Computer Scientists*. MIT Press, 1991.

[17] R. W. Schvaneveldt, editor. *Pathfinder Associative Networks: Studies in Knowledge Organization*. Ablex Publishing Corporation, Norwood, NJ, 1990.

[18] Amos Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

[19] Keith E. Williamson and Michael J. Healy. Deriving engineering software from requirements. *Journal of Intelligent Manufacturing*, 11(1):3–28, 2000.

[20] Keith E. Williamson, Michael J. Healy, and Richard A. Barker. Industrial applications of software synthesis via category theory–case studies using specware. *Automated Software Engineering*, 8(1):7–30, 2001.