Mathematical Models of Cognitive Space and Time

Joseph Goguen

University of California at San Diego
Dept. Computer Science & Engineering
9500 Gilman Drive, La Jolla CA 92093–0404 USA

Abstract. This paper explores reasoning about space and time, e.g., in metaphors of time as space; an important method is to find minimal assumptions needed to reach the same conclusions that humans reach. Some mathematical language, including the notion of triad, is introduced for this purpose, formalizing and generalizing the cognitive semantics approaches to conceptual spaces (in the senses of both Fauconnier & Turner and of Gärdenfors), blending, and metaphor; in particular, continuous mathematics is used to model space and time. A new explanation of emergent structure in blend spaces is also discussed, and proposed as a source of creativity. Four main examples illustrate the approach, and an appendix encapsulates the most difficult mathematics.

1 Introduction

Space and time are very fundamental human concepts, but despite great advances in physics, and extensive research in experimental psychology, our understanding of how humans actually use these concepts is still very incomplete, and the best cognitive science results are all rather recent. Although practical human reasoning about space and time is quite efficient, machine reasoning is often very poor; we would like to understand how humans do it, and to help machines do it better. Recent work in cognitive linguistics has found that human reasoning is very often mediated by metaphor[1]. Therefore the study of reasoning with metaphors that involve space and time is of particular interest.

Other recent research in cognitive linguistics claims that metaphor is best seen as the result of integrating (blending) conceptual spaces[2], because this generalizes and subsumes prior work of Lakoff and others on metaphor as a map from a more concrete source space to a more abstract target space, by introducing a "blend space" that includes relevant content from both the source and target spaces, plus perhaps new "emergent" structure [6]. Under either view, for metaphors involving time or

[1] This theme runs throughout [6], and [29] is an extended study of metaphor in mathematics.

[2] The word "space" is used metaphorically here as a "container" of symbolic entities; conceptual spaces are explained in detail in the next section.

space, it is necessary to represent those concepts in conceptual spaces, in a way that supports inference. Surprisingly[3], we have found that familiar mathematical representations of space, time and motion are adequate for this purpose: time can be represented by intervals I from either the real numbers or the integers, space can be represented by a manifold $\mathcal{M}$ (i.e., a smooth[4] subset of n-dimensional Euclidean space $\mathbb{R}^n$, where $\mathbb{R}$ denotes the space, i.e., the line, of all real numbers), and motion by a function $I \to \mathcal{M}$ (which should be smooth when I is a subset of $\mathbb{R}$).

This paper explores four examples in some detail, each exhibiting a different kind of ambiguity. The first (Example 2) is Peter Gärdenfors' skin color example [9], which uses geometrical conceptual spaces. The second (Example 3) discusses the many different blends of "house" and "boat" [11]. The third (in Section 4) is a puzzle from [6], in which a Buddhist monk ascends and then descends a sacred mountain; this is also where we discuss recruitment and creativity in blending. The fourth example (in Section 5) discusses reasoning about time using spatial metaphors [34]; sentences like "The Wednesday meeting was moved forward two days" have been studied experimentally by Núñez [33], and found to be ambiguous in interesting ways[5]; the method is to determine minimal assumptions needed to derive particular conclusions, such as that there are one (or two, or three) different dates for the rescheduled meeting; when the conclusion reached is the same as that of human subjects, this provides a method for validating models. Readers not interested in the mathematics should focus on these examples, but should also look at the other, non-main, examples for continuity. Natural language understanding is not addressed, because we focus on reasoning about understandings as represented by models.

Section 2 reviews some cognitive science research on concepts and conceptual spaces. Section 3 introduces Unified Concept Theory (abbreviated UCT) [14, 16] in a somewhat informal way that avoids the category theory used in [16]. Section 4 uses the Buddhist monk example of [6] to introduce the basics of our modeling methods for space and time, demonstrating the need to enrich the conceptual spaces of Gilles Fauconnier with types, functions and axioms, and to enrich the machinery of cognitive linguistics with (what we call) triads and triad blending; a solution is also suggested

[3] In view of the phenomenological research of Husserl, Merleau-Ponty and others, though of course it is not likely to surprise physicists.

[4] Technically, this means that the nth derivative exists and is continuous for every integer n.

[5] Note that the spatial term "forward" is metaphorically applied to time in this sentence.

to the mystery of how "emergent structure" appears in blends. Section 5 discusses time as space metaphors, focussed on examples like those in [33], and Section 6 draws some conclusions, while Appendix A provides further mathematical details of UCT; readers who are not very familiar with category theory and institutions[6] will find a definition of institution that uses no category theory in Appendix A, followed by a brief exposition of some basic notions on category theory, with some illustrative diagrams.

This paper does not aim to contribute new theories of human cognition; rather, it aims to contribute a new language in which certain kinds of theories of human cognition can be expressed and explored with much greater precision than has previously been possible. Although the examples given to illustrate this are theories of human cognition in certain very particular situations, they are not supported with new experimental or linguistic data, but instead, their properties are examined mathematically, and compared with common sense, or with experiments done by others. This situation is analoguous to that in physics: to establish his theory of gravitation, Newton needed not only experimental data, but also the mathematical language of the calculus, to precisely express his theories and derive their consequences for the solar system. Similarly, the contributions of this paper are like the mathematical theory of differential calculus rather than the physical theory of gravitation, though of course no claim is made that it is equally significant!

A major feature of UCT is its use of "triads," which can combine discrete symbolic with continuous geometrical representations, and thus can reap the benefits of both. Models are built for particular purposes, and need only be adequate to those purposes; this view is ubiquitous in engineering and applied science, and is opposed to the view that mathematical models should in principle be able to capture every aspect relevant to every possible situation, which is associated with philosophical realism, as in the linguistic theories of Montague, Barwise & Perry, Chomsky, and others (see [15] for further discussion of this topic).

2 Cognitive Science of Concepts

This section surveys some cognitive research on concepts. In a series of papers that are a foundation for contemporary cognitive semantics, Eleanor Rosch designed, performed, and carefully analyzed innovative

[6] Although rather technical, institutions are needed because each of our main examples uses a different logic.

experiments, resulting in a theory of human concepts that differs greatly from the Aristotelian tradition of giving necessary and sufficient conditions, based on properties. Rosch showed that concepts exhibit *prototype effects*, e.g., degrees of membership that correlate with similarity to a central member. Moreover, she found that there are what she called *basic level concepts*, which tend to occur in the middle of concept hierarchies, to be perceived as gestalts, to have the most associated knowledge, the shortest names, and to be the easiest to learn. Expositions in [27, 28, 24] give a concise summary of research of Rosch and others on conceptual categories. This work served as a foundation for later work on metaphor by George Lakoff and others [27, 28, 30, 26]. One significant result from this research is that many metaphors come in families, called *basic image schemas*, that share a common sensory-motor pattern. For example, MORE IS UP is grounded in our everyday experience that higher piles contain more dirt, or more books, etc. Metaphors based on this image schema are very common, e.g., "That raised his prestige." or "This is a high stakes game."

Fauconnier's *mental spaces* [5] (also called *conceptual spaces* [6]) do not attempt to formalize concepts, but instead capture the important idea that concepts are used in clusters of related concepts. This idea can be formalized as a very simple logic, consisting of individual constants, and assertions that certain relations (mostly binary) hold among certain of those constants; it is remarkable how much natural language semantics can be encoded with this framework (see [5, 6]).

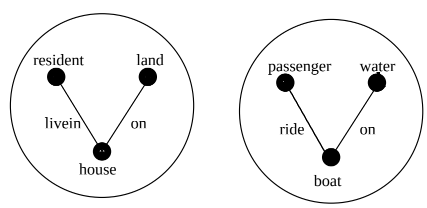

Fig. 1. Two Simple Conceptual Spaces

Example 1. Figure 1 shows two simple conceptual spaces, the first for "house" and the second for "boat." These do not give all possible information about these concepts, but only the minimal amount needed for a particular application, which is further discussed below. The "dots" represent the individual constants, and the lines represent true instances of relations among those individuals. Thus, the leftmost line asserts *livein(resident, house)*, which means that the relation *livein* holds between these two constants. We will soon see good reasons for assigning "sorts" (also called "types") to constants and relations. For example, *resident* and *passenger* can be given the sort *Person*, and *house* and *boat* the sort *Object*. □

Peter Gärdenfors [9] proposes a notion of "conceptual space" that is very different from that of Fauconnier, since it is based on geometry rather than logic. An intriguing hypothesis in [9] is that all conceptual spaces are *convex*[7]. Although [9] aims to reconcile its geometric conceptual spaces with symbolic representations like those of Fauconnier, it does not in fact provide a unified framework. However, such a unification can be done in two relatively straightforward steps. The first step is to introduce *models* in addition to many sorted logical theories, where a model provides a set of instances for each sort, a function for each function symbol, and a relation for each relation symbol; since we are interested in the models that satisfy the axioms in the theory, an explicit notion of *satisfaction* is also needed; special cases of such relations are called *classifications* in [1] and *formal contexts* in [8]. This leads to the basic UCT notion of triad, which is discussed in detail in Section 3.

The second step is to fix the interpretations in models of certain sorts to be particular geometrical spaces (the term "standard model" is often used in logic for such a fixed interpretation). For example, a sort *color* might be interpreted as the set of points in a fixed 3D manifold representing human color space, coordinatized by hue, saturation and brightness values, as in Figure 2, which is shaped like a "spindle," i.e., two cones with a common base, one upside down. This provides a precise framework within which one can reason about properties that involve colors, as in the following:

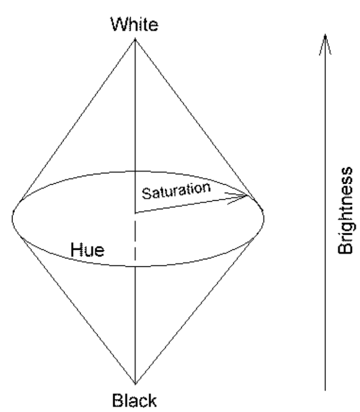

Fig. 2. Human Color Manifold

[7] A subset of Euclidean space is *convex* if the straight line between any two points inside the subset also lies inside the subset; this generalizes to non-Euclidean manifolds by using geodesics instead of straight lines.

Example 2. A nice example in [9] concerns the (English) terms used to describe human skin tones (red, black, white, yellow, etc.), which have a very different meaning in that context than e.g., in a context of describing fabrics. Gärdenfors claims that this shift of meaning can be explained by embedding the space of human skin tones within the larger color manifold[8], and showing that in this space, the standard regions for the given color names are the closest fits to the corresponding skin color names. Technically, it is better to view the two geometrical spaces as related by a canonical projection from the spindle to the subspace, because Gärdenfors' convexity hypothesis plus the reasonable assumption that each space has a "reference point" (a "zero color") guarantees that such a canonical projection exists[9]. Gärdenfors does not give a formal treatment of the color terms themselves, but we can view them as unary predicates in a theory (or "ontology"), and view the relationship between skin color terms and colors in the color spindle as a satisfaction or classification relation; note that many colors will not have any corresponding skin tone name. □

$$
\begin{tikzcd}
 & B & \\
I_1 \arrow[ur] & & I_2 \arrow[ul] \\
 & G \arrow[ul] \arrow[ur] \arrow[uu] &
\end{tikzcd}
$$

Fig. 3. Information Integration over a Shared Subobject

The most important recent development in the tradition of Rosch, Lakoff, and Fauconnier is *conceptual blending*, claimed in [6] to be a fundamental cognitive operation, which combines different conceptual spaces into a unified whole. The simplest case is illustrated in Figure 3, where for example I_1, I_2 might be mental spaces for "house" and "boat" (as in Figure 1), with G containing so-called "generic" elements such that the

[8] Methodologically, it seems reasonble to take the standard color spindle as a neutral ground from which deviations due to context, such as priming, can be viewed as deviations by projection mappings.

[9] Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be convex 3-manifolds in $\mathbb{R}^3$ that have reference points p_1, p_2, respectively. Then $\pi \colon \mathcal{M}_1 \to \mathcal{M}_2$ is defined as follows: Given $m_1 \in \mathcal{M}_1$, let L_1 be the ray (half infinite line) starting at point p_1 and passing through m_1, let m'_1 be the point where L_1 intersects the surface of $\mathcal{M}_1$, and let L_2 be the ray parallel to L_1 in $\mathcal{M}_2$ from p_2, intersecting the surface of $\mathcal{M}_2$ at m'_2. Now define $\pi(m_1) = m_2$ where m_2 is the point on L_2 such that the ratios of line lengths $p_1m_1/p_1m'_1$ and $p_2m_2/p_2m'_2$ are equal.

maps $G \to I_i$ indicate which individuals should be identified. Some "optimality principles" are given in Chapter 16 of [6] for judging the quality of blends, and hence determining which blends are most suitable, although these distillations from numerous examples are far from formal.

Example 3. Blends are not determined uniquely, not even up to isomorphism. For example, if we blend the two spaces in Figure 1 (for the concepts "house" and "boat"), then B could be[10] "houseboat," or "boathouse," or some other combination of the two input spaces (see below). The blend diagram for "houseboat" is shown in Figure 4. As in Figure 3, the bottom space is the generic or base space, the top is the blend space, and the other two are the input spaces, in this case for "house" and "boat." The arrows between circles indicate *conceptual maps*, which describe how to map entities in the source space to entities in the target space; in general, they are partial, not total. In this simple example, all four spaces have graphs with the same "vee" shape, and the maps simply preserve that shape, e.g., each maps the bottom node of the "vee" in its source space to the bottom node in its target; this is not typical of more complex examples.

Figure 5 shows the "boathouse" blend of the same two concepts. In it, the boat ends up in the house. Notice that mapping *resident* to *boat* does not type check (this presupposes the type assignments given in the table on page 10 unless *boat* is "cast" to be of type *Person*; otherwise, the boat could not live in the boathouse. This is the kind of metaphor called *personification* in literary theory, in which an object is considered a person. A third blend is similar to (in fact, symmetrical with) the above "boathouse" blend; in it, a *house/passenger* ends up riding in the boat (e.g., where a boat is used to transport prefabricated houses across a bay for a housing development on a nearby island).

A fourth blend is less familiar than the first three, but has very good preservation properties. This is an amphibious RV (recreational vehicle) that you can live in, and can ride on land and on water. A fifth blend has an even less familiar meaning: a livable boat for transporting livable boats; this was found by the Alloy blending algorithm developed by Fox Harrell and I, and perhaps only an algorithm could have discovered this counter-intuitive blend [19]. A sixth blend gives a boat used on land for a house; it omits axioms that a house/boat be on water and a passenger ride a house/boat. Alloy also found 42 other, less obvious blends, most of which are far from optimal.

[10] It is unusual that there are such convenient names for two of these blends.

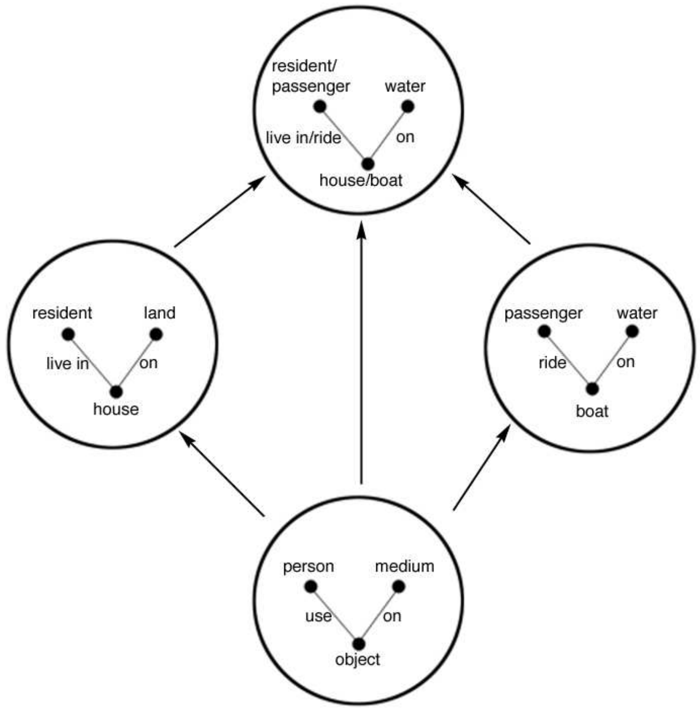

Fig. 4. Houseboat Blend Diagram

The extent to which a mapping preserves source space features can be used as a formal optimality criterion [11, 18], and for this example, our intuitive sense of the relative purity of the blends, and the degree to which they seem boat-like and house-like, corresponds to the degree to which the appropriate morphisms preserve entities and axioms from the input spaces. See [19] for a detailed discussion. □

Blending theory [6] refines the metaphor theory of Lakoff, by explaining the metaphorical map from I_1 to I_2 as a kind of "side effect" of a blend B of I_2 and I_2. This theory notes that a metaphor really constructs a new blend space in which only certain parts of I_1 and I_2 appear, and in which some new structure found in neither I_1 nor I_2 may also appear; the usual formulation of metaphor as a "cross space mapping" $m\colon I_1 \to I_2$ is the reflection of the identifications that are made in B, i.e., if i_1, i_2 are constants in I_1, I_2 respectively, that map to the same constant in B, then we set $m(i_1) = i_2$. See Figure 6.

Example 4. In the metaphor "the sun is a king," the more concrete input space I_1 is for "a king" while I_2 is for "the sun;" the constants "sun" and "king" from their respective input spaces are identified in the blend, so "king" maps to "sun," but the fact that kings may collect taxes is not

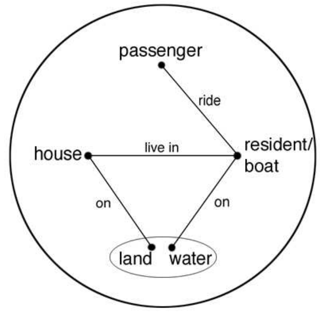

Fig. 5. Boathouse Blend Space

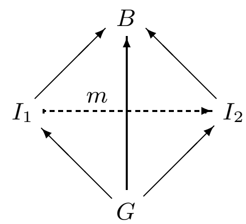

Fig. 6. Cross-Space Mapping

mapped up or across. However, if we add "the corona is his crown," then the element "crown" in I_1 is identified with the element "corona" from I_2 in the blend space, and so "crown" is mapped to "corona"; the same must be done for the "ownership" relations. □

3 Theories and Triads

This section summarizes parts of unified concept theory, beginning with a review of *algebraic theories* [11], which provide additional features over conceptual spaces that are needed for applications to areas such as user interface design [11], where many signs have parts, and these parts can only be put together in certain ways. For example, consider icons, windows, scrollbars, etc.; or consider words in a sentence, or the visual constituents of diagrams, such as Figure 1. Algebraic theories may have *constructor functions*, which build complex signs from simpler signs; for example, a window constructor could have arguments for a scrollbar, label, and content. Then one can write *W1 = window(SB1, L1, C1)*; there could also be additional arguments for color, position, and other details of how these parts constitute a particular window. This approach conveys information about the relations between parts and wholes in a much more explicit and useful way than just asserting *has-a(window, scrollbar)*, and it also avoids many of the problems that plague the *has-a* relation and its axiomatizations in formal mereology[11] ([36] explains some of these problems).

Sorts (or types) in algebraic theories serve to restrict the structure of signs: each declared constant has a sort, and each relation and function has restrictions on the sorts that its arguments may take. In example 1, the relation *own* can only take arguments of sorts *Person* and *Object* (in that order). In algebraic theories, relations are represented as Boolean valued functions. Allowing sorts to have *subsorts* provides a more effective way to support inheritance than the traditional *is-a* relation. For example, *Person* might have a subsort *Adult*. *Order sorted algebra* [22] provides a mathematical foundation that integrates inheritance with whole/part structure (using constructor functions instead of the *has-a* relation) in an elegant and computationally tractable algebraic formalism that also captures some subtle relations between inheritance and whole/part relations[12].

Example 5. Here we explain the notation of the algebraic theories that we use to represent conceptual spaces. Algebraic theories may have conditional equations as axioms to further constrain the space of possible signs; for example, certain houses might restrict their residents to be adults. Fauconnier's mental spaces are the special case of order sorted algebraic theories with no functions, no sorts or subsorts, and with only atomic relation instances as axioms. The table below gives the theory forms of the two conceptual spaces in Example 1:

$resident\colon \to Person$ $house\colon \to Object$ $land,\ water\colon \to Medium$ $livein\colon Person\ Object \to Bool$ $on\colon Object\ Medium \to Bool$	$passenger\colon \to Person$ $boat\colon \to Object$ $land,\ water\colon \to Medium$ $ride\colon Person\ Object \to Bool$ $on\colon Object\ Medium \to Bool$
$livein(resident, house)$ $on(house, land)$	$ride(passenger, boat)$ $on(boat, water)$

This notation is similar to that of functional programming, or more precisely, an equational programming language like OBJ [21]. Each of the two theories has two parts, one for declaring sorts, constants, and functions, and one for asserting axioms to serve as constraints on interpretations. The first three lines declare constants with their types[13]. The next two

[11] Mereology is the study of whole/part relations.

[12] E.g., the monotonicity condition on overloaded operations with respect to subsorts of argument sorts described in [22].

[13] The arrow appears because technically it is better to view constants as functions with no arguments.

lines declare relations as Boolean valued functions, in this case, each with two arguments. The last two lines give axioms, which here assert that certain relations hold on certain constant arguments; there is an implicit "*= true*" after each relation instance. The set of declarations of types and functions of a theory is called its *signature*. □

There is much experience using algebraic theories to specify and verify computer-based systems (e.g., [21, 20, 4]), and various extensions have been devised to facilitate this. One such extension allows certain sorts to have fixed interpretations (such as *Bool* and *Medium* in the two theories above, or the natural numbers), while other sorts are allowed arbitrary interpretations. Hidden algebra [20] provides additional features to better handle dynamic systems with states, which are a central feature of computer-based systems. *Semiotic spaces* further extend algebraic theories by adding priority relations on sorts and constructors, information that helps greatly with user interface design applications [11, 12]. Semiotic spaces are also called *semiotic systems* or *semiotic theories*, because they define systems of signs, not just single signs, for example, all possible displays on a particular digital clock, or a particular cell phone.

If T is a theory (e.g., that for "house"), then a *model* M for T provides concrete instances for all declared functions (which we recall include the constants) in T, in such a way that all the axioms in T are satisfied in M. In this case, we write $M \models_C T$, where C is a *context* that restricts the theories and/or models that are allowed. For example, contexts might consist of sets of declarations (so they are signatures), with $M \models_C T$ restricted to those T with axioms that only use symbols from C, M restricted to those models that instantiate the symbols in C, and with $M \models_C T$ holding if and only if M satisfies all the axioms in T.

Unified concept theory [14] (abbreviated *UCT*) uses the term *triad* (in honor of Charles Sanders Peirce) for a combination of a symbolic space, a context, a geometrical space (or set of spaces), and a relation among them, denoted $\models$. We can illustrate this with two triads from Example 2. Here contexts are sets of color names viewed as unary predicates. The first triad consists of the color spindle $\mathcal{C}$, a context C of color names, and an ontology theory O for color names in C, such that each color name is true (under $\models_C$) on a certain convex submanifold of the color spindle[14], while the second triad consists of the skin color submanifold $\mathcal{S}$, a set C' of skin color names contained in C, an ontology O' for C', with elements of C'

[14] It makes sense for this to be a convex fuzzy submanifold, in conformance with Rosch's work on prototype effects; see [14] for details of this notion.

again interpreted as convex submanifolds. The projection map $\pi\colon \mathcal{C} \to \mathcal{S}$ defined in Example 2 can be seen as giving a *triad morphism* along with a suitable inclusion map $i\colon O' \to O$, since we have

$$\pi(m) \models_{\text{skin}} s \text{ iff } m \models_{\text{spindle}} i(s)$$

for all colors m in $\mathcal{C}$ and all skin color names s in O'. (To allow $\models$ to have fuzzy values for color name predicates, e.g., in the unit interval $[0, 1]$, the "iff" in the above should be replaced by an equality relation on the fuzzy values.) An equivalent form that better fits the general notion of triad given below is

$$\pi(\mathcal{C}) \models_{\text{skin}} a' \text{ iff } \mathcal{C} \models_{\text{spindle}} i(a')$$

for all axioms a' in O', noting that in this case, $\pi(\mathcal{C}) = \mathcal{S}$ and the axioms in O' define the unary predicates s for skin colors. These formulae are also similar to the nicely named "infomorphisms" of [1], which in fact are a special case of triad morphisms.

More formally: contexts are sets C of color names; C-models are convex 3-manifolds $\mathcal{M}$ in $\mathbb{R}^3$ with a given reference point and a given convex submanifold for each $c \in C$; C-theories T are sets of axioms about the unary predicates in C; and $\models_C$ tells whether a model $\mathcal{M}$ satisfies an axiom, and by extension, whether it satisfies a theory T. Thus a *triad* is a triple $(\mathcal{M}, C, T)$ such that $\mathcal{M} \models_C T$, where $\mathcal{M}$ is a C-model and T is a C-theory. A *morphism* from one such triad $(\mathcal{M}', C', T')$ to another, $(\mathcal{M}, C, T)$ is a pair (Φ, Ψ) where $\Phi\colon T' \to T$ and $\Psi\colon \mathcal{M} \to \mathcal{M}'$ such that

$$\Psi(m) \models_C a \quad \text{iff} \quad m \models_{C'} \Phi(a)$$

for all $m \in M, a \in T$. Then (i, π) is a morphism $(\mathcal{C}, \Sigma, O') \to (\mathcal{S}, \Sigma, O)$ in this sense, where Σ is the set of skin colors. A more general definition of triad is given in Appendix A.

Triads and sorts with fixed interpretation help solve the *symbol grounding problem*[15] [25]. Our approach to this problem is consistent with Peirce [35], who said that signs must be interpreted in order to refer, and that interpretation only occurs in some pragmatic context of signs being actually used. Sensors, effectors, and world models ground elements of conceptual spaces in reality, where the world models are geometrical spaces. This implies that the symbol grounding problem is artificial, created by a desire for something that is not possible for purely symbolic systems, as in classic logic-based AI, but which is natural for embodied systems.

[15] This is the problem of how abstract computational symbols can be made to refer to entities in the real world.

4 A Buddhist Monk Meets Himself

One of the most striking examples in [6], called "the Buddhist monk," is not a metaphor, but a clever puzzle. It is posed as follows: A Buddhist monk makes a pilgrimage to a sacred mountain, leaving at dawn, reaching the summit at dusk, spending the night there in meditation, then departing at dawn the next day, and arriving at the base at dusk. The question then posed is: is there a time such that the ascending monk and the descending monk are at the same place at that time? This question calls forth a blend in which the two days are merged into one, but the one monk is split into two! The reasoning needed to answer the question cannot be done in a logic-based blend space, because some geometrical structures are needed to model the path of the monk(s), in addition to the individuals and relations that are given logically. The table below shows the semiotic spaces for the first and second day in its first and second columns, respectively; notice the explicitly given types, which are needed to constrain possible interpretations of the declared elements.

$Time = [6, 18]$ $Loc = [0, 10]$ $m\colon Time \to Loc$	$Time = [6, 18]$ $Loc = [0, 10]$ $m\colon Time \to Loc$
$m(6) = 0$ $m(18) = 10$ $(\forall\ t, t'\colon Time)\ t > t' \Rightarrow$ $m(t) > m(t')$	$m(6) = 10$ $m(18) = 0$ $(\forall\ t, t'\colon Time)\ t > t' \Rightarrow$ $m(t) < m(t')$

The first two lines of each theory are type definitions, while the third declares a function; here and hereafter, we assume such functions are smooth (i.e., continuously differentiable of all orders); the two type declarations mean that these types have fixed interpretations in all models. After that, each theory has three axioms, the third of which uses the notation $(\forall t, t'\colon\ Time)$ to introduce two variables, t, t', with their type, $Time$, for use in that axiom.

A model for the theory of the first day will interpret $Time$ as the fixed interval [6,18] (for dusk and dawn, in hours); it will also interpret Loc as another fixed interval, [0,10] (for the base and summit locations, in miles). Then m is interpreted as some continuous function $[6, 18] \to [0, 10]$, giving the monk's distance along the path as a function of time. The key axiom is the last one, a monotonicity condition, which asserts that the monk always makes progress along the path, though without saying how quickly or slowly. Each such function m corresponds to a different model of the theory. The theory for the second day is similar except the last three axioms assert that the monk starts at the top and always descends until reaching the bottom. The types $Time$ and Loc must be given exactly the same interpretations on the two days, but the possible paths are necessarily different. The blended theory is shown in the table below, in which m indicates the monk's locations on the first day and m' on the second day.

$Time = [6, 18]$	$Loc = [0, 10]$
$m, m', d\colon Time \to Loc$	
$t^*\colon \to Time$	
$m(6) = 0$	$m(18) = 10$
$m'(6) = 10$	$m'(18) = 0$
$(\forall\, t, t'\colon Time)\ t > t' \Rightarrow\ m(t) > m(t')$	
$(\forall\, t, t'\colon Time)\ t > t' \Rightarrow\ m'(t) < m'(t')$	
$(\forall\, t\colon Time)\ d(t) = m'(t) - m(t)$	
$d(t^*) = 0$	

To answer the puzzle, we have to solve the equation $m(t) = m'(t)$. If we let t^* denote a solution and let $d(t) = m'(t) - m(t)$, then the key "emergent" structure added to the blend space is $d(t^*) = 0$, since this allows us to apply the version of the Intermediate Value Theorem which says that a strict monotone continuous function which takes values a and b with $a \neq b$ necessarily takes every value between a and b exactly once. In this case, d is strict monotone decreasing, $d(6) = 10$ and $d(18) = -10$, so there is a unique time t^* such that $d(t^*) = 0$.

Blending theory [6] speaks of "*recruiting*" new spaces to the blend in order to create emergent structure, but it does not explain how this happens. We suggest that emergent structure arises by integrating new triads that match important non-integrated concepts in the input spaces. In this particular example, the locations of the monk on the two days are clearly important, since they are mentioned in the formulation of the puzzle, but they are not integrated by the generic space; the notion of being in the same place at the same time is also mentioned in the puzzle. We therefore suggest that the missing piece of the puzzle is the "meeting space" shown below, and that it is "recruited" by searching for a triad that matches key non-integrated key concepts in the blend:

$a, b, d\colon Time \to Loc$
$t^*\colon \to Time$
$(\forall\, t\colon Time)\ d(t) = a(t) - b(t)$
$d(t^*) = 0$

This just says that two individuals, a, b, meet if they are at the same place at the same time, where t^* is that time. To do the integration, let us call the two theories in the first table of this section I and I', and let the generic space G contain just the first two lines of these theories, with the maps $G \to I$ and $G \to I'$ inclusions. Next, let M denote the meeting space, let $M \to I$ map a to m, and let $M \to I'$ map b to m'. Then identifying the elements that are mapped, and just copying the others gives exactly the blend space of the second table in this section; see Figure 7. A good hypothesis is that this provides a general explanation for emergent structure in blends. (This example is also a good illustration of blending with more than two input spaces.)

$$\begin{tikzcd} & & B & \\ I \arrow[urr] & M \arrow[l] \arrow[ur] \arrow[rr] & & I' \arrow[ul] \\ & & G \arrow[uu] \arrow[ull] \arrow[ur] & \end{tikzcd}$$

Fig. 7. Buddhist Monk Blend Diagram

There are some surprises if we weaken the monotonicity axioms to become non-strict, so that the monk may stop and enjoy the view for a time, as formally expressed for the first day by the axiom

$$(\forall\, t, t' \colon \mathit{Time})\ t > t' \Rightarrow m(t) \geq m(t')$$

then (by another version of the Intermediate Value Theorem) the monk can meet himself on the path for any fixed closed proper subinterval $[a, b]$ of [6,18] (i.e., with $6 \leq a \leq b \leq 18$ with either $a \neq 6$ or $b \neq 18$). Moreover, if we drop the monotonicity assumption completely but still assume continuity, then (by the most familiar version of the Intermediate Value Theorem) there still must exist values t^* such that $m(t^*) = m'(t^*)$, but these t^* are no longer confined to a single interval, and can even consist of countably many isolated intervals. It seems safe to say that such observations would be difficult to make without a precise mathematical analysis like that given above[16]; indeed Fauconnier and Turner had not realized

[16] Of course, some aspects of this analysis are unrealistic, due to assuming that the monk can move arbitrarily quickly; a velocity restriction could be added (e.g., $|dm/dt| \leq 4$), but the extra complexity can only be justified if there is a specific need for it. This is a good illustration of the pragmatic character of model making, in particular, its sensitivity to how models are used, and the need for trade-offs.

that the monk could meet himself at more than a single instant before they saw this analysis. It is interesting and perhaps now even surprising that human subjects invariably reach the conclusion that is supported by the rather non-standard strict monotone version of the intermediate value theorem.

The structures in this example, consisting of a many sorted logical theory and a class of possible models for that theory (with some sorts having fixed interpretations in the models), can be seen as classifications in the sense of Barwise and Seligman [1]; but it is better to consider these structures as triads, because the geometry of the models is important here, and because the different theories have different contexts. Moreover, the example blends not just theories, but also classes of models, and hence triads, including their contexts; this is spelled out in the appendix.

5 Reasoning about Time as Space

There is enormous cross-cultural evidence that time is primarily conceptualized as space[17]: in all languages studied so far, the vocabulary for time is primarily spatial [34]. Here are two simple examples in English: "The end of the year is approaching" and "We are coming to the end of the year." Notice that in the first, time is moving, while in the second the deictic[18] reference point (conventionally called "ego") is moving while time is fixed. These are both dynamic metaphors, but space as time metaphors can also be static, as in "The due dates of the reports are too close together," where neither time nor ego is moving. Times can also be moving with respect to other times, as in "December follows November."

These examples (and many many others that are similar) show that there are three dualities:

1. Ego RP *vs.* Time RP (where "RP" abbreviates "Reference Point");
2. Static *vs.* Dynamic; and
3. Landmark (the deictic reference point, which is fixed) *vs.* Trajector (that which moves) [34].

Of course, there can also be metaphoric blends involving time, as in "Time flies like an arrow." So things can get very complex! They were recently made even more complex by the discovery that Aymara, a language of the Peruvian Andes, has a static metaphor of time as space, where the future

[17] Kant might be surprised to learn that one of his analytic a priori categories is significantly more fundamental than another.

[18] The term "deictic" is used for words where meaning depends on the context where they appear; the deictic reference point is the location of this appearance.

is behind Ego RP, rather than ahead of it[19], as in all other known languages; [34] gives conclusive evidence using gesture, with control groups of Spanish speakers and even bilingual speakers. However, for all dynamic cases, even in Aymara, the future is in front of Ego RP.

The question that motivated the material in this section was, under what conditions does the sentence "The Wednesday meeting was moved forward two days" have just two solutions (as days of the week)? We will use many sorted equational logic, where declarations are for sorts and operations, and theories consists of sets of equations over a given signature. We also let contexts be theories, and let triads be theory extensions, i.e., inclusions of the corresponding sets of declarations and of axioms. We also allow some sorts to have fixed interpretations. The following theory defines the basic ingredients that are needed to formalize the problem posed above:

$$\begin{array}{|lll|}
\hline
\multicolumn{3}{|l|}{Time = \mathbb{Z}} \\
\multicolumn{3}{|l|}{Su, M, Tu, W, Th, F, Sa \colon \ \to Day} \\
\multicolumn{3}{|l|}{day \colon Time \to Day} \\
\multicolumn{3}{|l|}{E,\ T \colon \to RP} \\
\multicolumn{3}{|l|}{f_2 \colon Time\ Time\ RP \to Time} \\
\hline
day(0) = Su & day(1) = M & day(2) = Tu \\
day(3) = W & day(4) = Th & day(5) = F \\
day(6) = Sa & & \\
\multicolumn{3}{|l|}{(\forall\ t, t' \colon Time)\ day(t) = day(t') \text{ if } |t - t'| = 0 \text{ mod } 7} \\
\multicolumn{3}{|l|}{(\forall\ t, t' \colon Time)\ f_2(t, t', E) = t - 2 \text{ if } t \le t'} \\
\multicolumn{3}{|l|}{(\forall\ t, t' \colon Time)\ f_2(t, t', E) = t + 2 \text{ if } t > t'} \\
\multicolumn{3}{|l|}{(\forall\ t, t' \colon Time)\ f_2(t, t', T) = t + 2} \\
\hline
\end{array}$$

In the first line, the sort $Time$ is declared to represent the set $\mathbb{Z}$ of all integers, while the second line declares the days of the week, the third declares a function that maps times to days, the fourth declares two constants for the two different kinds of reference point, and the fifth declares a function to give the result of moving a time "two days forward." For the axioms, the first eight define how days of the week are assigned to integers, and the next three define how f_2 works, following the discussion at the beginning of this section. The notation $|t - t'|$ in the eighth axiom indicates the absolute difference of t, t', and the axiom says that two integers have the same day if they differ by some multiple of 7.

Perhaps the following extension of the above theory is the most obvious way to set up the problem posed above:

[19] I.e., the experiencer faces the flow of events, as in "March is coming soon."

$m, m', e \colon \to Time$ $r \colon \to RP$
$f_2(m, e, r) = m'$ $day(m) = W$

Models for this extended theory instantiate the constants m, m', e, r with values that are consistent with the axioms. One difficulty with this formulation is that there will be an infinite number of solutions; to get around this, we can just specify that $m = 3$. Under these assumptions, we can indeed prove that there are just two solutions, namely $m' = 1$ and $m' = 5$. However, this is still unsatisfactory in a way, because the initial assumption $m = 3$ (i.e., that the locator is the originally scheduled time of the meeting) precludes exploring other options, such as that the locator is the rescheduled time of the meeting ($m' = 3$).

The simplest way forward is to consider two theories, which will serve as contexts, one for each choice of the locator:

$m, m', e \colon \to Time$ $r \colon \to RP$	$m, m', e \colon \to Time$ $r \colon \to RP$
$f_2(m, e, r) = m'$ $m = 3$	$f_2(m, e, r) = m'$ $m' = 3$

These theories are the same except for their last lines; let us denote them A and B, respectively. We now do another case split, on whether $r = E$ or $r = T$; let the resulting four theories be denoted $A.1, A.2, B.1$ and $B.2$. Finally, due to the form of the axioms for f_2, it is convenient to do one more case split, on whether $m > e$ or $m \leq e$; let the resulting 8 cases be denoted $A.1.1, A.1.2, ..., B.2.2$. Each of them is a context within which we can ask which days are possible solutions. In the B cases, we are asking about the solutions for m, the original meeting time, rather than m', the rescheduled meeting time. The following table summarizes the results for each case:

	r	e	m'	m
$A.1.1$	E	$< m$	5	3
$A.1.2$	E	$\geq m$	5	3
$A.2.1$	T	$< m$	1	3
$A.2.2$	T	$\geq m$	5	3
$B.1.1$	E	$< m'$	3	1
$B.1.2$	E	$\geq m'$	3	1
$B.2.1$	T	$< m'$	3	1
$B.2.2$	T	$\geq m'$	3	5

Thus we see that in the most general context, there are three possible values for m', but only two under that assumption that $m = 3$; similarly, there are only two possible solutions for m under the assumption that $m' = 3$. If we use the notation $A.*.*$ for the context that includes the first four cases (where $m = 3$), and $B.2.*$ for the context where $M' = 3$ and $r = T$, $*.*.*$ for the most general context, etc., then we can also describe the results of the above table in terms of contexts.

Further research should consider static metaphors including the unusual one in Aymara. Other topics include the case where the meeting is normally scheduled for the same day of every week, and unusual models of time, such as occur in science fiction[20], e.g., parallel universes, and time travel (as in the movie "Back to the Future").

6 Conclusions

Humans are quite efficient at practical reasoning about space and time, but such reasoning is largely unconscious, and is therefore difficult to understand. This paper attempts to improve our understanding by introducing a formal language called unified concept theory, in which models of cognitive space and time can be expressed, including metaphors of space as time. We have studied four examples in some detail, each exhibiting an interesting, but different, kind of ambiguity. In each case, humans quickly reach a single conclusion. We have constructed models for each example, such that the conclusions that can be deduced from the models are the same as those reached by humans, and we have explored the assumptions and deductions needed to reach those conclusions.

Although UCT is relatively simple, the formal models and deductions may seem more complex than is warranted by the apparent simplicity of the examples; but this is typical of common sense human processes that are rapid but unconscious, including natural language understanding and vision. It is also similar to what happens when relatively simple physical theories are applied to real systems, such as aircraft wings, ecologies, and proteins: the underlying physical theories are much simpler than the practical models.

Our models of human concepts use triads, which can combine discrete symbolic with continuous geometric representations, so that reasoning can draw on the advantages of each, and the relation between them; this has a precise and very general foundation in the theory of institutions

[20] It is striking how easily readers accept such far from ordinary possibilities.

(see the appendix). Unified concept theory also extends conceptual integration (blending) to triad integration, with a precise notion of context and a precise theory of its role in blending. The Buddhist monk example demonstrates how "emergent structure" arises in blends: through the integration of additional triads that match important non-integrated concepts in the input spaces; this is a major new hypothesis of this paper. Although much remains to be done, we consider the results so far encouraging support for the promise of applying UCT to a wider range of problems in cognitive science.

References

1. Jon Barwise and Jerry Seligman. *Information Flow: Logic of Distributed Systems.* Cambridge, 1997. Tracts in Theoretical Computer Science, vol. 44.
2. David Borgo and Joseph Goguen. Sync or swarm: Group dynamics in musical free improvisation. In Richard Parncutt, Annekatrin Kessler, and Frank Zimmer, editors, *Proceedings, Conference on Interdisciplinary Musicology*, pages 52–53. Dept. Musicology, University of Graz, 2004. Held in Graz, Austria, 15–18 April 2004.
3. David Borgo and Joseph Goguen. Rivers of consciousness: The nonlinear dynamics of free jazz. In Larry Fisher, editor, *Jazz Research Proceedings Yearbook*, pages 46–58. IAJE Publications, 2005.
4. Răzvan Diaconescu and Kokichi Futatsugi. *CafeOBJ Report: The Language, Proof Techniques, and Methodologies for Object-Oriented Algebraic Specification.* World Scientific, 1998. AMAST Series in Computing, Volume 6.
5. Gilles Fauconnier. *Mental Spaces: Aspects of Meaning Construction in Natural Language.* Bradford: MIT, 1985.
6. Gilles Fauconnier and Mark Turner. *The Way We Think.* Basic, 2002.
7. José Luiz Fiadeiro. *Categories for Software Engineering.* Springer, 2004.
8. Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations.* Springer, 1997.
9. Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought.* Bradford, 2000.
10. Joseph Goguen. A categorical manifesto. *Mathematical Structures in Computer Science*, 1(1):49–67, March 1991.
11. Joseph Goguen. An introduction to algebraic semiotics, with applications to user interface design. In Chrystopher Nehaniv, editor, *Computation for Metaphors, Analogy and Agents*, pages 242–291. Springer, 1999. Lecture Notes in Artificial Intelligence, Volume 1562.
12. Joseph Goguen. Semiotic morphisms, representations, and blending for interface design. In *Proceedings, AMAST Workshop on Algebraic Methods in Language Processing*, pages 1–15. AMAST Press, 2003.
13. Joseph Goguen. Musical qualia, context, time, and emotion. *Journal of Consciousness Studies*, 11(3/4):117–147, 2004.
14. Joseph Goguen. What is a concept? In Frithjof Dau and Marie-Laure Mungier, editors, *Proceedings, 13th Conference on Conceptual Structures*, volume 3596 of *Lecture Notes in Artificial Intelligence*, pages 52–77. Springer, 2005. Kassel, Germany.

15. Joseph Goguen. Ontotheology, ontology, and society. *Int. J. Human Computer Studies*, page to appear, 2006. Special issue on ontology, ed. by Christopher Brewster and Kieron O'Hara.
16. Joseph Goguen. Information integration in instutions. In Lawrence Moss, editor, *Memorial volume for Jon Barwise*. Indiana, to appear 2006.
17. Joseph Goguen and Rod Burstall. Institutions: Abstract model theory for specification and programming. *Journal of the Association for Computing Machinery*, 39(1):95–146, January 1992.
18. Joseph Goguen and Fox Harrell. Information visualization and semiotic morphisms. In Grant Malcolm, editor, *Multidisciplinary Studies of Visual Representations and Interpretations*, pages 93–106. Elsevier, 2004. Proceedings of a workshop held in Liverpool, UK.
19. Joseph Goguen and Fox Harrell. Foundations for active multimedia narrative: Semiotic spaces and structural blending, to appear. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*.
20. Joseph Goguen and Kai Lin. Behavioral verification of distributed concurrent systems with BOBJ. In Hans-Dieter Ehrich and T.H. Tse, editors, *Proceedings, Conference on Quality Software*, pages 216–235. IEEE Press, 2003.
21. Joseph Goguen and Grant Malcolm, editors. *Software Engineering with OBJ: Algebraic Specification in Action*. Kluwer, 2000.
22. Joseph Goguen and José Meseguer. Order-sorted algebra I: Equational deduction for multiple inheritance, overloading, exceptions and partial operations. *Theoretical Computer Science*, 105(2):217–273, 1992. Drafts exist from as early as 1985.
23. Joseph Goguen and Grigore Roşu. Institution morphisms. *Formal Aspects of Computing*, 13:274–307, 2002.
24. Rebecca Green. Internally-structured conceptual models in cognitive semantics. In Rebecca Green, Carol Bean, and Sung Hyon Myaeng, editors, *The Semantics of Relationships*, pages 73–90. Kluwer, 2002.
25. Stevan Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
26. Masako Hiraga. *Metaphor and Iconicity: A cognitive approach to analyzing texts*. Palgrave Macmillan, 2004.
27. George Lakoff. *Women, Fire and Other Dangerous Things: What categories reveal about the mind*. Chicago, 1987.
28. George Lakoff and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic, 1999.
29. George Lakoff and Rafael Núñez. *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics into Being*. Basic Books, 2000.
30. George Lakoff and Mark Turner. *More than Cool Reason*. Chicago, 1989.
31. Saunders Mac Lane. *Categories for the Working Mathematician*. Springer, 1998. Second Edition.
32. Till Mossakowski, Joseph Goguen, Razvan Diaconescu, and Andrzej Tarlecki. What is a logic? In Jean-Yves Beziau, editor, *Logica Universalis*, pages 113–133. Birkhauser, 2005. Selected papers from the First World Conference on Universal Logic.
33. Rafael Núñez, Benjamin Motz, and Ursina Teuscher. Time after time: The psychological reality of the Ego- and Time-Reference-Point distinction in metaphorical construals of time. *Metaphor and Symbol*, accepted for publication.
34. Rafael Núñez and Eve Sweetser. Aymara, where the future is behind you: Convergent evidence from language and gsture in the cross-linguistic comparison of spatial construals of time. *Cognitive Science*, accepted for publication.

35. Charles Saunders Peirce. *Collected Papers*. Harvard, 1965. In 6 volumes; see especially Volume 2: Elements of Logic.
36. Simone Pribbenow. Merenymic relationships: From classical mereology to complex part-whole relations. In Rebecca Green, Carol Bean, and Sung Hyon Myaeng, editors, *The Semantics of Relationships*, pages 35–50. Kluwer, 2002.

A More Mathematics of Unified Concept Theory

We begin with the notion of *institution* [17], which axiomatizes the possible kinds of theory and model that can be used for UCT. This is important because our examples used several different kinds of logic and model, and other applications will no doubt require still other logics and models. First, we assume there are *contexts*[21], and *context morphisms*, which have a (partially defined) composition operation (if $\varphi\colon\ C \to C'$ and $\psi\colon\ C' \to C''$ then $\varphi;\psi\colon\ C \to C''$ denotes their composition) that is associative (i.e., $(\varphi;\psi);\xi = \varphi;(\psi;\xi)$ whenever these compositions are defined) and has identities (i.e., given a context C, there is a context morphism 1_C such that $1_C;\varphi = \varphi$ and $\psi;1_C = \psi$ whenever these compositions are defined)[22].

Next, we describe how context changes affect theories and models; this also is done axiomatically[23]: We assume that for each context C, there is a given set $Sen(C)$ of axioms (or sentences) for that context, and a given set $\mathbf{Mod}(C)$ of models[24] for that context. If x is an axiom for context C and if $\varphi\colon\ C \to C'$, then we let $\varphi(x)$ denote the result of moving a to the context C', and we assume that $1_{C'}(a) = a$ and $\varphi;\psi(x) = \psi(\varphi(x))$ for $\psi\colon\ C' \to C''$. Similarly, given a model M' for C', let $\varphi(M')$ denote the C'-model that results from the change of context (notice that this translation is "contravariant" rather than "covariant"), and we assume that $1_C(M') = M'$ and that $\psi;\varphi(M') = \psi(\varphi(M'))$. Finally, given $\varphi\colon\ C \to C'$, we assume the *satisfaction condition*, that

$$M' \models_{C'} \varphi(x) \ \text{ iff } \ \varphi(M') \models_C x$$

for all C'-models M' and all C-axioms x. See Figure 8. It seems that any logic (that has a notion of model) is an institution in this sense; a detailed argument for this is given in [32], which also describes how a

[21] Most institution literature uses the term "signature," but "context" is more appropriate for the wider range of applications now being explored for institutions, many of which exhibit similarity to Peirce's triadic semiotics [35], in which the context of a sign is important for its interpretation.
[22] More technically, we are assuming a given *category* of contexts.
[23] More technically, we assume two *functors* on the category of contexts.
[24] More technically, a class, in the sense of Gödel-Bernays set theory.

great deal of metamathematics can be done at the level of institutions. (We mention that institutions can be generalized in various ways: one is to let the satisfaction relation take values other than true and false, e.g., in a lattice; another is to let model and sentence classes have more structure, such as a category; see [23] for details.)

$$
\begin{tikzcd}
 & \mathbf{Set} \\
\mathbf{Sign} \arrow[ur, "Sen"] \arrow[dr, "\mathbf{Mod}"'] & \models \\
 & \mathbf{Set}^{op}
\end{tikzcd}
$$

$$
\begin{tikzcd}
C \arrow[d, "\varphi"'] & \mathbf{Mod}(C) & C & Sen(C) \arrow[d, "Sen(\varphi)"] \\
C' & \mathbf{Mod}(C') \arrow[u, "\mathbf{Mod}(\varphi)"] & \models_C & Sen(C')
\end{tikzcd}
$$

Fig. 8. Structure of an Institution

A C-*theory* is a set of C-axioms. We now define a *triad* over an institution $\mathcal{I}$ to be a triple $(\mathcal{M}, C, T)$ where $\mathcal{M}$ is a set[25] of C-models and T is a C-theory such that $\mathcal{M} \models_C T$. Then a *triad morphism* from $(\mathcal{M}, C, T)$ to $(\mathcal{M}', C', T')$ is a pair of maps $\Phi\colon\, T \to T'$ and $\Psi :\, \mathcal{M}' \to \mathcal{M}$ such that

$$M' \models_{C'} \Phi(x) \;\text{ iff }\; \Psi(M') \models_C x$$

for all models $M' \in \mathcal{M}'$ and all axioms $x \in T$; this condition is similar to the satisfaction condition. An equivalent form is

$$M' \models_{C'} \Phi(T) \;\text{ iff }\; \Psi(M') \models_C T$$

for all M' in $\mathcal{M}'$, where $\Phi(T) = \{\Phi(x) \mid x \in T\}$. It is easy to see that every context morphism induces a triad morphism; but triad morphisms are more general, as is shown by the skin color example. In the special case where all functions are unary predicates and there is just one context, triad morphisms degenerate to the *infomorphisms* of [1].

Categories represent structures, such as automata (with their homomorphisms), groups (with their homomorphisms), and vector spaces (with linear transformation). The description of contexts and their morphisms im the first paragraph actually constitutes a precise definiton of the category notion. For example, **Set** denotes the category of sets, and **Cat**

[25] This is more general than the definition in the body of the paper.

denotes the the category of (small) categories. Similarly, functors represent constructions on structures, such as the formal language accepted by an automaton, or the lattice of normal subgroups of a group; and the description of how context changes sentences and models contains precise definitions for covariant and contravariant functors. Finally, natural transformations represent relations between functors, such as that one constructs less structured objects than another; the vertical arrows in Figure 9 enforce the mutual consistency of the relationships. More intuitions and examples for categorical concepts can be found in [10], and in many other places.

$$\begin{tikzcd} F(C) \arrow[r, "a_C"] \arrow[d, "F(\varphi)"'] & G(C) \arrow[d, "G(\varphi)"] \\ F(C') \arrow[r, "a_{C'}"'] & G(C') \end{tikzcd}$$

Fig. 9. A Natural Transformation from F to G

Colimits abstractly capture the notion of "putting together" objects to form larger objects, in an optimal way that takes account of shared substructure[26]. Whereas colimits are determined uniquely up to isomorphism, blends are not, as shown by Example 3, where B could be "houseboat," or "boathouse," or some other combination of the two input spaces. Therefore colimits are not an adequate formalization of blending. This raises the mathematical challenge of weakening colimits so that multiple, non-isomorphic solutions are allowed. $\mathbf{\frac{3}{2}}$**-colimits** are (somewhat tentatively) suggested as a model of blending in [11], based on $\frac{3}{2}$-categories, which have a partial ordering relation on morphisms; this relation can be used to reflect the quality of morphisms, and thus can represent certain values.

In the Buddhist monk example, the institution is equational logic, contexts C for triads are theories, and C-theories are *extensions* of C (i.e., they contain all the declarations and axioms of C), while C-models have signatures that extend C, and $M \models_C T$ holds iff M satisfies T. Thus triads here are $(\mathcal{M}, C, T)$ where T extends C, and $\mathcal{M}$ is a set of models of C plus the declarations of T.

To reconcile these triads with the definition of triad over an institution $\mathcal{I}$ given above, we construct a new institution $\mathcal{TX}(\mathcal{EQ})$ from the institution $\mathcal{EQ}$ of many (or order) sorted equational logic. Actually, since it is

[26] See [10] for an intuitive discussion of the mathematics, and [31, 7] for details.

just as easy, we do the construction for an arbitrary institution $\mathcal{I}$: The contexts of $\mathcal{TX}(\mathcal{I})$ are theories of $\mathcal{I}$, i.e., pairs $C = (\Sigma, A)$ where Σ is an $\mathcal{I}$-context and A is a set of Σ-axioms of $\mathcal{I}$; the C-axioms of $\mathcal{TX}(\mathcal{I})$ are theory extensions (Σ', A') of C (i.e., $C \subseteq C'$ and $A \subseteq A'$); the C-models of $\mathcal{TX}(\mathcal{I})$ are Σ'-models (where $\Sigma \subseteq \Sigma'$) that satisfy A; and $M \models_C (\Sigma', A')$ holds for $\mathcal{TX}(\mathcal{I})$ iff $M \models_{\Sigma'} A'$ for $\mathcal{I}$.

It is interesting to notice that the blend theory in the Buddhist monk example is a colimit, and the set of models in the blend triad is a limit (the dual notion to colimit) of the sets of models in the component triads; this is a special case of a result that holds for any institution: colimits of triads are computed by taking colimits of theories and limits of model sets. The institution, contexts, and triads are the same for the example of Section 5, and it is interesting to notice that here the model set in the triad of the blend for the most general context (denoted $*.*.*$) corresponds to the final table in Section 4, viewed as a set of tuples. It may also be interesting to know that it would be easy for a suitable automatic theorem prover to automatically find the solutions for each context.

An important open problem is to formalize the optimality principles of [6], and a promising approach is to use continuous mathematics, e.g., potential functions on phase space, as in our research on the qualitative segmentation of music [2, 3, 13]; this connection suggests that complexity functions might be able to measure the novelty of blends. In addition, it seems likely that image schemas can be modeled using triads, with geometrical spaces for the sensory-motor aspects, and semiotic spaces for the conceptual aspects. So there are still many interesting issues for future research.