# Mapping Spatial Relations

by

Rajesh Kasturirangan

MSc, Mathematics
Indian Institute of Technology, Kanpur, 1993



Submitted to the Department of Brain and Cognitive Science in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[February 2004]
June 2004



Author .......................................................................................
Rajesh Kasturirangan
Department of Brain and Cognitive Science
February 5, 2004

Certified by .......................................................................................
Whitman Richards
Professor of Cognitive Science
Thesis Supervisor

Accepted by .......................................................................................
Earl Miller
Chairman, Department Committee on Graduate Students

## Abstract

One of the fundamental issues in cognitive science is the problem of grounding concepts in the perceptual world. In this thesis, I present a computational theory for how spatial relations are grounded in the perceptual world.

Three constraints are critical to this theory: *abstractness*, *groundedness* and *flexibility* all of which need to be satisfied in order to explain the structure of spatial concepts. I then show how a formal framework, based on the mathematical notions of category theory can be used to model the grounding problem. The key computational ideas are that of *minimal mapping* and *derivations***.** A *minimal mapping* of two categories, A and B, is the "smallest' category, C, that contains A and B. A derivation is a sequence of categories that follow a minimal mapping rule.

Derivations and minimal mappings are used to model three domains – the semantics of prepositions, the structure of a toy "Jigsaw World" and the semantics of generic terms and quantifiers. In each case, I show how the computational theory gives rise to insights that are not available upon a purely empirical analysis. In particular, the derivational account shows the importance of stable, non-accidental features and of multiple scales in spatial cognition.

Thesis Supervisor: Whitman Richards
Title: Professor, Department of Brain and Cognitive Sciences

## Table of Contents


| | |
|---|---|
| Acknowledgements | 5 |
| 1. Introduction | 6 |
| 1.1 Orientation | 6 |
| 1.2 Spatial Relations | 9 |
| 1.3 The Structure of the Mapping Problem for Spatial Relations | 13 |
| 1.4 The Main Contributions of the Thesis | 20 |
| 1.5 Summary of the Thesis | 25 |
| 2. Spatial Syntax | 27 |
| 2.1 Preliminary Considerations | 27 |
| 2.2 Spatial Syntax | 31 |
| 2.3 Category Theory | 35 |
| 2.4 Categories and Natural Kinds | 41 |
| 3. Semantics of Space | 43 |
| 3.1 Introduction | 43 |
| 3.2 Semantic Constraints | 45 |
| 3.3 The Global Model | 47 |
| 3.4 Statics | 50 |
| 3.5 Dynamics | 54 |
| 3.6 Mappings and Derivations | 58 |
| 4. The Semantics of Prepositions | 65 |
| 4.1 Introduction | 65 |
| 4.2 A Theory of Prepositional Semantics | 68 |
| 4.3 Generating the Meaning of Prepositions | 69 |
| 4.4 A Full Derivation for ACROSS | 88 |
| 4.5 Summary | 100 |
| 5. Tangrams: A Test bed for Spatial Cognition | 102 |
| 5.1 Introduction | 102 |
| 5.2 Orientations | 102 |
| 5.3 Tangrams | 104 |
| 5.4 Tangrams: Computational Theory | 107 |
| 5.5 Tangrams: Emergent Properties | 114 |
| 5.6 Tangrams: Algorithmic Analysis | 118 |
| 5.7 Multiple Scale Graphs | 120 |
| 5.8 Deriving Path Length Behavior | 121 |
| 5.9 Jigsaw World: Simulations | 123 |
| 5.10 Conclusions | 127 |

**6. Generics and Extensions** **128**
**6.1 Introduction** **128**
**6.2 An Introduction to Generics and Quantifiers** **128**
**6.3 Extensions** **132**
**6.4 Quantifiers** **134**
**6.5 Discussion** **137**

**7. Endgame** **139**
**7.1 General Observations** **139**
**7.2 Psychological Implementations** **142**
**7.3 Concluding Remarks** **146**

**Appendix: Testing the Multiple Scale Hypothesis** **148**

**References** **154**

**Acknowledgements**

My primary debt is to my advisor, Whitman Richards for encouraging me to think for myself and chart my research directions, while at the same time being generous with his advice. It has been a pleasure working with him. I also want to thank Jacob Feldman for many fruitful discussions over the years, especially during my visits to Rutgers. Two other people who have helped me enormously are Steven Tainer and Piet Hut.

This thesis is the outcome of discussions with too many people to name, but here is a partial list – Michael Coen, Roland Fleming, Vinay Kumar, John Mikhail, Sourabh Niyogi, Richard Russell, Peter Schein, Jaspal Singh, Manish Singh.

Thanks also go to my brother, Sriram, for lending me his laptop when it was really needed, for programming help and for regular pizza pilgrimages to New York. Finally, I want to thank my wife, Rebecca, for supporting me through these difficult times and for taking this journey together. Beck, I hope to return the favor soon.

# 1. Introduction.

## 1.1. Orientation

Computational modeling of mental faculties is based on two broad traditions, roughly, the "philosophical-scientific" tradition and the "engineering" tradition. The philosophical-scientific tradition asks "Where does knowledge come from?" Historically, the answer to this question has been either "knowledge is innate," the rationalist answer, or "knowledge is learnt," the empiricist answer. The philosophical-scientific tradition is based on the intuition that there are rules governing mental processes and that these principles can be found using computational modeling. The goal of the scientific enterprise is to characterize the *principles* that characterize a human being's knowledge. Within the domain of high-level cognition and perception (the domain that interests me), the goal of the philosophical-scientist is:

(1) To elucidate the principles that govern human cognition of high-level systems such as space-time, causality and concepts, without regard to how the knowledge is used by human beings in real world situations;

(2) To describe how these principles are acquired by children as they develop.

On the other hand, the engineer asks "Why are we so smart?" Human beings perceive, cognize and act on the world so effortlessly that it is hard to see the complexity underlying these tasks. However, when engineers have tried to build machines that emulate human performance, they have invariably run into enormous difficulties. A good exposition of the difficulties involved in constructing computers with common sense is described by Dennett ( Dennett, 1998).

> Once upon a time there was a robot, named R1 by its creators. One day its designers arranged for it to learn that its spare battery, its precious energy supply,

was locked in a room with a time bomb set to go off soon. R1 located the room and the key to the door, and formulated a plan to rescue its battery. There was a wagon in the room, and the battery was on the wagon, and R1 hypothesized that a certain action which it called PULLOUT(WAGON, ROOM) would result in the battery being removed from the room. Straightaway it acted, and did succeed in getting the battery out of the room before the bomb went off. Unfortunately, however, the bomb was also on the wagon. R1 knew that the bomb was on the wagon in the room but did not realize that pulling the wagon would bring the bomb out along with the battery. Poor R1 had missed the obvious implication of its planned act.
Back to the drawing board. " The solution is obvious," said the designers." Our next robot must be made to recognize not just the intended implications of its acts, but also the implications about their side effects, by deducing these implications from the descriptions its uses in formulating its plans." They called their next model, the robot-deducer R1D1. They placed R1D1 in much the same predicament that R1 had succumbed to, and as it too hit upon the idea of PULLOUT(WAGON, ROOM) it began, as designed, to consider the implications of such a course of action. It had just finished deducing that pulling the wagon out of the room would not change the color of the room's walls, and was embarking on a proof of the further implication that pulling the wagon out would cause its wheels to turn more revolutions than there were wheels on the wagon -- when the bomb exploded.

Back to drawing board. "We must teach it the difference between relevant implications and irrelevant implications," said the designers, "and teach it to ignore the irrelevant ones." So they developed a method of tagging implications as either relevant or irrelevant to the project at hand, and installed the method in their next model, the robot-relevant-deducer, or R2D1 for short. When they subjected R2D1 to the test that had so unequivocally selected its ancestors for extinction, they were surprised to see it sitting, Hamlet-like outside the room containing the ticking bomb, the native hue of its resolution siclied o'er with the pale cast of thought, as Shakespeare (and more recently Fodor) has aptly put it. "Do something!" they yelled at it. "I am," it retorted. "I'm busy ignoring some thousands of implications I have determined to be irrelevant. Just as soon as I find and irrelevant implication, I put it on the list of those I must ignore, and ..." the bomb went off.

The three "thought experiments" in the above quote illustrate the problems of modeling common sense. The predicament of R1 is due to its being ungrounded – it cannot take into account the effect of its actions in the real world. On the other hand, R1D1 is too grounded – it cannot abstract away the relevant aspects of the world. R2D1 is appropriately grounded, but it is too inflexible. Human cognition on the other hand, is abstract, grounded and is used flexibly from situation to situation. What is it that humans have that computers do not (as yet)?

My own bias is a combination of the two approaches. I agree with the engineer in the statement of the problem. Everything we know about biological systems suggests that they have to be analyzed functionally, keeping in mind the following questions:

(1) What is the system for?
(2) What contexts does it operate in?
(3) How does an organism use the information in the environment to perform its tasks?

Organisms, including human beings live in rapidly changing, complex, dynamic environments and therefore any understanding of their "smarts" has to take these environmental conditions into account. The traditional "philosophical-scientific" models with their emphasis on static, internal structure, do not say much about the dynamic character of cognitive faculties, whether they be low-level or high-level. On the other hand, I agree with the philosopher in emphasizing the role of principles. While organisms may be functional entities, there is an intuition that these functions are not mere hacks – they are based on relatively sound, rational design criteria. It is often argued that the difficulty of modeling "common sense" is due to a lack of understanding of the principles that regulate higher levels of human cognition (Minsky, 1986). In this regard, there is no reason to expect the problems of high-level cognition to be any different from other areas such as linguistics and computer vision where research became much more sophisticated when they were based on a sound theoretical framework (Chomsky, 1956; Marr, 1982). The goal of this thesis is to elucidate some of the design principles behind the versatility of human cognition in one domain, the representation of spatial relations.

## 1.2. Spatial Relations

Human spatial cognition is distinguishable from the spatial abilities of other animals (perhaps excluding apes) by its flexibility. At any given time, we are engaging in a series of tasks, typically using several different representations for each task. For example, we might be talking to a family member on a cell phone while walking to work, switching back and forth between a wide range of spatial, physical and social faculties. A simple example of this ability to engage multiple representations to solve just one task is as follows.

> **Puzzle.** John, Joe, Mary and Sam are sitting on a bench. John is sitting in the left hand corner. John is next to Mary, Sam is next to Joe, and Joe is next to John. What is the order in which they are sitting?

It is quite hard to perform this task with verbal instructions alone. On the other hand, if we are also given some visual information in the form of a board with four dots representing the four people (figure 1.1) the task is much easier. One can ask several questions about this mapping from language to vision: How is the linguistic input set in correspondence with visual features? How is the solution recognized as being true?

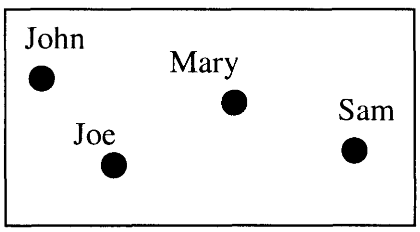



**Figure 1.1**

The above puzzle illustrates a fundamental problem for an agent in a complex environment, the problem of mapping representations in one representational format on to representations in another format. The mapping problem is quite central to human cognition; not only does it occur in spatial puzzles, it occurs in other domains of human cognition as well. In fact, if one is to believe theorists such as Lakoff (Lakoff, 1987), Fauconnier (Fauconnier, 1997) and Gentner (Gentner, 1983), mappings from one representation to another, in the form of analogy or metaphor, are central to human thought.

While there are several phenomenological and experimental investigations of analogy and metaphor as well as computer programs that try to simulate human capacities for analogy, there isn't much work on the intrinsic computational problems raised by the problem of mapping different representations on to each other. Formally, we can state the computational problem of mapping representations as follows:

> **The Mapping Problem**: Suppose $R_1$ and $R_2$ are two representational systems. Then,
> (1) What are the computationally plausible maps $F: R_1 \rightarrow R_2$ from $R_1$ to $R_2$?
> (2) What are the rules governing these maps from $R_1$ to $R_2$?

The main computational issue is that there is often a *qualitative* difference between the two representational systems that are being mapped. Consider the following examples, one from language and the other from navigation.

*Language Example*:

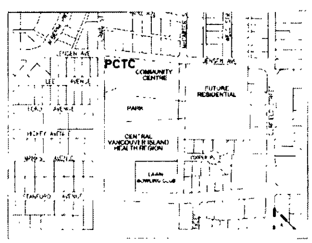
(1) The village is on the other side of the river.
(2) John lives in Chicago

What is the meaning of the preposition "*on*" in sentence 1 and the preposition "*in*" in sentence 2? The answer to these questions is not obvious, since there is no obvious correspondence between the linguistic input and the external 3D world since language is quite schematic and represents very little spatial information. For example, in the first sentence, the term "village" is plausibly represented as a point. In the second sentence, the preposition *in* does not identify an exact spatio-temporal location, only a certain region of space. In both of these sentences, spatial language represents only a tiny fraction of the spatial information available about Chicago or villages. At the same time, both of these sentences are true of a large number of world situations, because an infinite number of states of a three dimensional world are compatible with sentence 1 as well as sentence 2. In other words, the prepositions ON and IN quantify over a large number of world situations. Therefore, when prepositions are grounded in the perceptual world, the human mind is mapping highly abstract, "quantifier-like" relations to very concrete perceptual representations. How is this done?

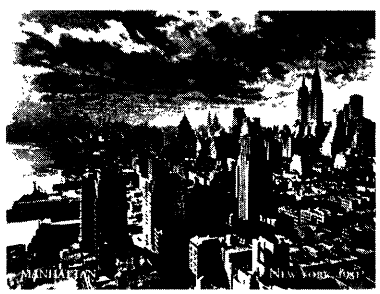
*Visual Example*

The same issues arise in visual mappings as well, such as the relationship between the map of a city neighborhood and the actual three-dimensional structure of the neighborhood (figure 1.2 a-b), which is quite schematic. Maps often represent locations as points and features such as roads and rivers as paths, ignoring their 3D shape. In some cases, such as a subway map, the distance between points may be ignored altogether. In other words, visual maps are also highly impoverished and abstract representation of the world. Furthermore, the same map can be used for a large number of tasks at various times and spatial locations. How are human beings able to use maps that are schematic to aid navigation in a three dimensional world?

Like representations used in prepositions, visual representations used in spatial cognition are *qualitatively* different from the concrete perceptual representations in whom they are (ultimately) grounded. How is this done? What features of the world does a visual map represent? These are the kind of questions that need to be addressed by a computational theory of mapping.




(a)



(b)

**Figure 1.2**

In this thesis, I elucidate some of the principles behind the versatility and of "mappings" in one domain, the representation of spatial relations. The representation of spatial relations is one of the core components of our cognitive systems. The computation of spatial relations is important for spatial reasoning (Cohn & Hazarika, 1996), navigation (Gallistel, 1990), and spatial language (Talmy,2000; Herskovits, 1985, 1986; Jackendoff, 1996). Given the importance of spatial relations for reasoning, action and language, it is not surprising that they have been the focus of attention in several fields, namely: linguistics, robotics, animal learning and child development. It is hard to imagine human cognition without spatial representations. Some scholars (Lakoff &Johnson, 1998; Talmy, 2000) have even argued that the creativity of human cognition in general is grounded in the capacities derived from our representation of spatial relations. Therefore, a computational understanding of spatial representations that throws light on these properties will also help us understand conceptual knowledge in general.

### 1.3. The Structure of the Mapping Problem for Spatial Relations

As the examples from language and vision show, the relationship between spatial cognition and the external world is complex. The representation of spatial relations (spatial representations for short) share three seemingly contradictory properties – they are *abstract*, *grounded* and *flexible*, properties that need to be explained by a computational theory of spatial relations.

Spatial representations are abstract because:

(1) Spatial representations do not represent objects/scenes in the world faithfully. That is to say, they represent some aspects of the world schematically, while ignoring other world features altogether;
(2) Spatial relations are categories, not particulars. While early perceptual systems deliver a representation of particular objects and events in the immediate spatio-temporal environment of an organism, spatial relations are represented as categories that club together a large set of individual objects and events.

Spatial Relations are grounded because:

(1) They are always coupled to features of our immediate environment and
(2) They are used in a way that is appropriate to the context.

For example, a map of a city represents those aspects of the city that are needed for navigation. The same goes for football coach who draws a sequence of plays using a chalk-board. In other words, spatial representations are grounded in the external world. This point may seem tautological; if spatial representations did not represent important aspects of the spatial environment, what are they good for? Yet, given the abstractness of spatial representations, it is far from obvious as to how they are grounded.

Finally, spatial representations are flexible because:

(1) They can be combined rapidly and creatively to solve new tasks
(2) They are robust with respect to noise.

For example, suppose you want go to a friend's house in an unknown part of Boston. Your friend might tell you which subway stop is nearest to his house and provide you with walking directions from the subway stop. You use a subway map to get to the right subway stop and then switch to the walking directions for the walk from the subway to your friend's house. In other words, the spatial representations being used at a given moment depends on the spatio-temporal context and the current task.

Abstractness, groundedness, and flexibility have been described in several fields but these efforts have, for the most part, been descriptive or implementational. In linguistics, many authors (Talmy, 2000; Lakoff & Johnson, 1999; Herskovits, 1986) have noticed that spatial language is schematic. In cognitive science, the literature on mental models has focused on the idea of abstract representations (Johnson Laird, 1983). In AI, frames and schemas (Minsky, 1974; Schank and Abelson, 1977) have been central to the modeling of common sense. In vision, Ullmans pioneering work on routines (Ullman, 1996) recognized the abstractness as well as the flexibility of visual cognition. Similarly, several authors have commented on the dynamic, embodied, structure of human and animal activity (Varela et. al., 1991; Agre, 1997; Beer, 2000).

Despite these pioneering approaches to the mapping problem, these explanations have been largely descriptive. What we need is a computational theory of mapping which integrates abstractness, groundedness and flexibility in one framework. In other words, we need a theory that answers the following questions:

(1) How is it possible for a representational system to be abstract and grounded at the same time?
(2) What procedures are capable of generating the set of spatial representations that are abstract as well as grounded?
(3) What constraints enable a system to perform a complex task in a flexible, robust manner, i.e., unexpected interruptions do not disrupt the task?

Furthermore, if the "static" problem of relating two fixed representations is hard, then the dynamic version of the mapping problem is even harder, where by 'dynamic mapping problem', I mean the problem of mapping a sequence of representations over a period of time to produce a final result or to achieve a predetermined goal. The difficulties in understanding the dynamical mapping problem are illustrated vividly by Plato in Meno (Cooper & Hutchinson, 1997) where he cleverly uses a dialogue speaker (Socrates) and another speaker (a slave boy) argues for the impossibility of transmission of information from one person to another. Note that any exchange between two agents leading to a conclusion is a dynamic mapping problem, for it involves the mapping of representations from one agent to another over time and produces a final result.

Plato's arguments have been enormously influential, affecting cognitive science to this day. I reproduce below fragments from Socrates' dialog with the slave boy, where he "proves" a version of the Pythagorean theorem and argues that the boy could not have learnt the basic propositions from him, hence, geometric knowledge must be innate. A series of diagrams used by Socrates in the argument is given in figure 1.3 below.

**Fragment 1**

SOCRATES: Attend now to the questions which I ask him, and observe whether he learns of me or only remembers.

MENO: I will.

**Fragment 2**

SOCRATES: And you know that a square figure has these four lines equal?

BOY: Certainly.
SOCRATES: A square may be of any size?

BOY: Certainly.

**Fragment 3**

SOCRATES: And does not this line, reaching from corner to corner, bisect each of these spaces?

BOY: Yes.

SOCRATES: And are there not here four equal lines which contain this space?

BOY: There are.

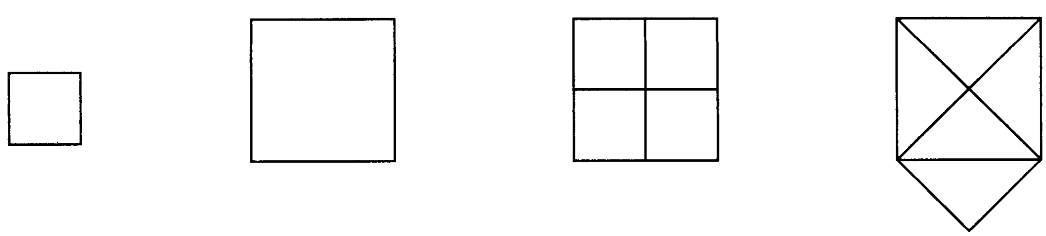


**Figure 1.3**

**Fragment 4**

SOCRATES: That is, from the line which extends from corner to corner of the figure of four feet?

BOY: Yes.

SOCRATES: And that is the line which the learned call the diagonal. And if this is the proper name, then you, Meno's slave, are prepared to affirm that the double space is the square of the diagonal?

BOY: Certainly, Socrates.

SOCRATES: What do you say of him, Meno? Were not all these answers given out of his own head?

MENO: Yes, they were all his own.

SOCRATES: And yet, as we were just now saying, he did not know?

MENO: True.

SOCRATES: But still he had in him those notions of his--had he not?

MENO: Yes.

SOCRATES: Then he who does not know may still have true notions of that which he does not know?

MENO: He has.

What information was transmitted from Socrates to the slave boy? What did the slave boy learn? It is hard to see how an information theoretic model can help us understand these questions. On the other hand, Plato's arguments remain convincing even now. How else can we explain the fact that we "know" that squares have equal sides or that they can have any size – given that real world objects have neither of those two properties? The basic argument is based on the intuition that correct responses to questions are predicated on intrinsic knowledge. However, what is interesting for my purposes is that Plato's arguments are incomplete, perhaps because of his philosophical agenda to understand the origin of knowledge. In particular, while Plato's arguments explain the abstractness of spatial representations, they do not explain how they are grounded or flexible. Nevertheless, the above dialog serves as a vivid account of their flexibility and their groundedness as well.

The flexibility of geometric representations is illustrated by the skill with which Socrates interrogates the slave boy and makes him come to an understanding of the Pythagorean theorem, a theorem that he surely did not possess before. The analysis of the exchange between Socrates and the boy is misleading; while the slave boy does produce the right answers, *he does not ask the questions or direct the process of inquiry and he does not know when the task has ended*, both of which are done by Socrates. The creative combination of geometric propositions is entirely mediated by the world, in the form of Socrates. Furthermore, the world also defines the goal of the exercise and provides essential resources for its solution. Both Socrates and the boy ground their arguments in an external representation –the diagrams in figure 3. It is hard to see how the slave boy could have followed the arguments without the use of visual aids. Even though the diagrams make the situation more complex, introducing new elements that need to be represented and tracked, mapping the problem on to a larger structure allows for a more transparent solution.

This procedure of augmentation – introducing new features that are not formally mandated – is part of the mutual dependence of abstract geometric structures (the slave boy's innate knowledge of triangles and squares) and concrete individual spatial elements (the figures of specific shape and size illustrated in the diagrams). By pairing an internal deductive process with an external form of representation the boy gains tremendous computational power.

Therefore, while one might agree with Plato on the existence of innate ideas, one can also make the following observations:

(1) The world (in the form of Socrates) contains the goal and also chooses the hypotheses under consideration by the mind.
(2) Abstract geometric representations have to be mapped onto external spatial structures (the diagrams) in order for the task to be completed successfully.

I argue that one has to take the *entire* system – the internal structures, the external structures and the map between the two – in order to understand spatial cognition. The cognitive scientist who gives a purely internal description of the geometric structures behind spatial cognition, is modeling only one of the three components of the system, a strategy that is likely to miss crucial elements. In contrast, for me the basic unit of analysis is the triad of mind-map-world, as illustrated in figure 1.4.

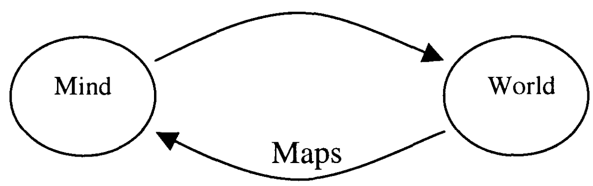



**Figure 1.4**

A computational approach to spatial representations has to be able to model the entire spectrum from abstract to concrete. On the one hand, we need spatial schema that capture abstract geometric features and their underlying, quasi-topological[1] principles. On the other hand, we need spatial representations that represent concrete individual figures in space and time; these figures have specific shape and size as well a particular location. Furthermore, abstract and concrete schema should be combinable at any given time, depending on the demands of the task.

## 1.4. Main Contribution of the Thesis

The dialog between Socrates and the slave boy in the previous section raises serious questions about the way in which abstract spatial concepts can be related to concrete external examples. The main contribution of this thesis is to show that qualitatively different aspects of spatial concepts and spatial semantics, such as abstract and concrete features, can be brought together within a computational theory. A computational theory of spatial concepts within a domain, for example, prepositions, should answer the following two questions:

(1) What spatial features are represented by the concepts in question?

(2) Why are those features represented and not others?

As I argued before, these questions are harder to answer than appears at first glance. To take the example of prepositions, how can one explain the fact that, on the one hand, there is no English preposition that distinguishes between standing on two legs versus standing on one leg (figure 1. 5(a)), simple topological distinction, while at the same time being capable of distinguishing between relatively subtle aspects of geometry as show in figure 1.5 (b), where the path of John in the right hand figure seems much

[1] Their quasi-topological character is illustrated in the fact that the principles governing the abstract structures are typically invariant under many spatial transformations such as scale invariance: squares can be of any size.

more likely to be deemed "across the park" than the path of John in the left hand figure, even though the paths are the same.

**Figure 1.5(a): John is standing on the ground**

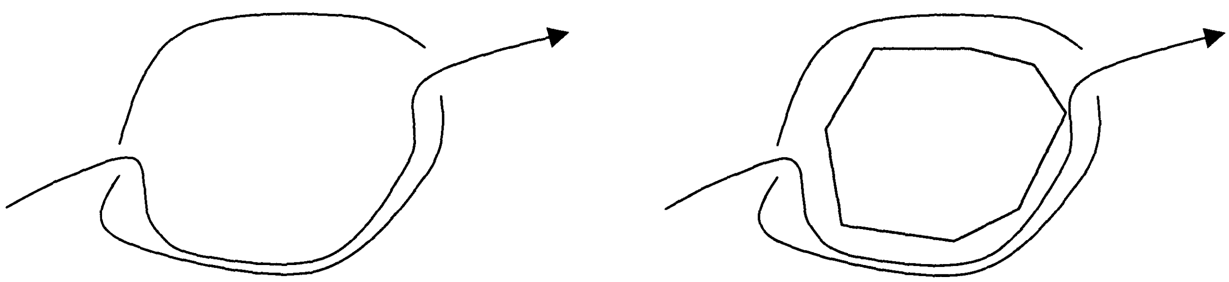


**Figure 1.5(b). John walked across the park**

As I show in chapter four, category theory offers a precise way of saying why prepositions can have schematized core meanings that are highly restrictive, and yet admit context dependent concrete meanings that are quite flexible. Furthermore, both the context-independent and the context-dependent meanings fall under the same generative framework and come from the same global model. Without the formal power of category theory, one cannot even state the abstract features that are needed for the unification of abstract and concrete structures. For example, once we have the notions of "topological category", "partial ordering generated by a category" and "depth in a partial order" (see chapters three and four), we can state the following constraints on the meaning of prepositions:

(1) The core meaning of spatial prepositions is within a topological category.
(2) Each category generates a partial ordering
(3) Prepositions represent features of depth $\leq 1$ in the partial order generated by the topological category of objects in three dimensions.

It turns out that contact is a topological feature of depth 1 and standing on two legs is a feature of depth 2, which is why ON does not distinguish between standing on one leg or standing on two legs. The machinery of category theory makes this distinction possible.

The depth constraint is an example of an abstract "competence" constraint on representations of spatial concepts. The competence-based approach of this thesis differs from previous accounts of spatial concepts and semantics both in its focus on computational theory and in its ability to unify qualitatively different aspects of semantics. In order to give the reader an idea of the relationship between my work and other approaches to spatial concepts involving computer models, let me a give a brief summary of Terry Regier and Laura Carlson's (Regier and Carlson, 2001) "attentional vector-sum" model of spatial relations and Annette Herskovits' (Herskovits, 1986) work on modeling spatial prepositions.

Regier and Carlson try to show how the computation of spatial relations can be implemented in geometric operations that are plausibly connected to visual attention. They argue that visual attention mediates representations that compute the formal operation of vector summation and that these vector-sums are central to the computation of spatial relations. In their recent paper (Regier and Carlson, 2001), they show how the attentional vector-sum model makes the right predictions about human judgement in the case of the preposition ABOVE.

However, the attentional vector sum model is not a model of human competence. In terms of David Marr's (Marr, 1982) three level (computational theory- algorithm-implementation) analysis of explanations of cognitive and perceptual phenomena, the attentional vector-sum model is an example of an explanation at an algorithmic level along with some evidence for its implementation in terms of attention based representations. A computational theory would ask the following questions of the attentional vector-sum model:

(1) Why does the representation system compute vector-sums and not something else?
(2) Where are these vector-sums generated?

My goal is to be able to answer questions of the kind given above. In this sense, Regier and Carlson's work is approaching the problem of spatial semantics at a different level of analysis than I do. Herskovits' work on prepositions is closer to a computational theory. She shows how the problem of modeling the meaning of prepositions is hard, by showing how prepositions are hard to define. In order to account for all aspects of the meaning of a preposition, Herskovits developed an account based on two kinds of meanings of prepositional concepts

(3) The idealized geometric meaning of a preposition, which captures the core meaning and then pragmatic factors and,
(4) Use types that capture the deviations from the core.

While I believe this subdivision is essentially right, it still leaves us with the following two questions:

(1) What do the core meanings and the use types represent?
(2) Is there any systematic way to generate core meanings and use types?

In chapter 4, I show how the core meaning as well as the use types can be generated from the same framework, based on category theory. Accordingly, I draw different conclusions from the division between core meanings and use types than the conclusions made by Herskovits. While she argues that there is no 'neat' computational theory of the meaning of prepositions, I argue that there is a unifying theory, but it not available at the level of core meanings. Instead, the unification takes place at a higher level, a level in which concrete and abstract meanings are brought together.

Note that Herskovits' account of spatial concepts is the converse of Socrates's account of spatial concepts. While Herskovits wants to cover both the abstract and concrete aspects of spatial semantics and she does not care whether there is a unifying theory that connects the two, Socrates cares for theory (in fact his theory has been very influential, i.e., that 'ideas' are innate), and he is willing to ignore concrete concepts in order to justify the innateness of abstract concepts. My claim is that we can have the cake and eat it too. Category theory gives us a way of understanding the commonalities between qualitatively different kinds of representation in ways that would me impossible to see without a similar formal framework. In other words, while Herskovits argues for a 'loose' theory of abstract and concrete concepts, and Socrates argues for a 'tight' theory of abstract concepts, I present a tight theory of abstract and concrete concepts, which shows the category-theoretic structures underlying the representation of spatial semantics.

### 1.5. Summary of the Thesis: An outline of the solution to the mapping problem.

I argue that the mapping problem with its three puzzles of abstractness, groundedness and flexibility has three aspects to a computational solution. These are *qualitative mapping*, *generativity* and *dynamicity*. Let me briefly explain these three properties and how my thesis approaches them.

Qualitative mapping is the property that the computational theory must be able to map qualitatively different kinds of representations on to each other – for the different representations that we use, such as language and perception are not just quantitatively different, but qualitatively different. Note that qualitative mapping, as a computational problem, is essentially a *syntactic problem*, i.e., a problem whose solution is purely formal, requiring a representational language that can accept qualitatively different representations. In chapter 2, I show how the framework of category theory is the appropriate formal language and how the qualitative mapping problem can be addressed within category theory.

While qualitative mapping is a problem of syntax, generativity and dynamicity are essentially semantic, i.e., they are driven by the nature of the world we live in. Generativity is the property that the qualitatively different representations, if they are to reflect the structure of the world, cannot be random (in a qualitative sense). That is to say, the qualitative categories must fall under some regularity, if they are capable of being mapped on to each other, as well as being carriers of useful information about the world.

Furthermore, if the qualitative categories are regular, not random, then there must be a systematic way to generate them. In chapter 3, I show how the ideas of non-accidentalness and partial ordering (Richards, Jepson & Feldman, 1996) lead to a generative framework for categories.

As the conversation between Socrates and the slave boy suggests, the mapping problem, if it is to explain the flexibility and robustness of spatial cognition, must be solved within a dynamic framework, not a static framework. In chapter 3, I show how the idea of *derivations* offers a natural way to solve the dynamic mapping problem in the framework of category theory.

Finally, I apply these general computational ideas to model cognitive phenomena in three domains – the meaning of spatial linguistic terms such as prepositions in chapter 4, a toy world of jigsaw puzzles to test out the dynamical framework in chapter 5 and a generalization of these ideas to a more abstract domain of quantifiers and generics in chapter 6.

# Chapter 2. Spatial Syntax

## 2.1. Preliminary Considerations.

In the introduction, I discussed the relationship between spatial concepts and the world in terms of three puzzles – abstractness, groundedness and flexibility. My intention is to use spatial concepts to make progress in modeling concepts as a whole –their origin, their structure and their relationship to perception. In other words, the ultimate goal of the enterprise is to understand spatial *semantics*. My goal is to understand the human ability use the right (spatial) concept in a given context, i.e., the mapping of spatial concepts on to the world, in a dynamic context[2].

Even the simplest spatial concepts demonstrate the need to take the mapping problem seriously. Let us take any English preposition, say ON or IN. Prepositions, like other concepts, are abstract and universal in a way that seems wholly inconsistent with perceptual experience. It is hard to see what set of perceptual features can capture the meaning of ON since ON applies to an *infinite* number of perceptual situations, with objects of various shapes and sizes that may share no perceptual feature whatsoever. Furthermore, prepositional statements capture truths that last far longer than the fleeting experience that gave rise to them – how else can one explain the durability of even a simple statement like “The cup is on the table” when compared with the rather limited experience of a cup and a table. How can one possibly extract such universal properties from such a concrete experience?

[2] Traditionally, the semantic problem of interest to philosophers is “ Under what conditions is a proposition true (spatial or otherwise).

At the same time, despite their abstraction, prepositions such as IN and ON are used to make meaningful claims about concrete world situations. When asked "Where is the cup?" the response "The cup is on the table" seems eminently *concrete* as well as informative. In general, human beings do *apply* prepositions correctly in an infinite number of real world situations. If the prepositions are so abstract, how can human beings ever learn to map them onto an infinite number of world situations?

If one accepts Socrates' claim, i.e., that the abstract meaning of a preposition has to be an innate property of the human mind, one also has to accept that in each of the infinite situations that licenses the use of a preposition (most of whom are never experienced by a human) there is something worthy of mapping the abstraction on to. What is this concrete entity? In other words, the mapping problem also includes the converse of Socrates' claim – Socrates asks "How can abstract ideas ever be learned from fleeting experience?" to which one can respond "How can an abstract idea ever be applied to an infinite number of concrete situations[3]?"

In fact, an important case of the mapping problem is the systematic application of abstract concepts in concrete situations. The problem of spatial semantics –from the perspective of the mapping problem- is not that of understanding the structure of abstract concepts or of modeling concrete situations but one of how human beings combine abstract and concrete representations in order to solve particular tasks. Going back to the proof of the Pythagorean theorem in the previous section, we need to understand three things:

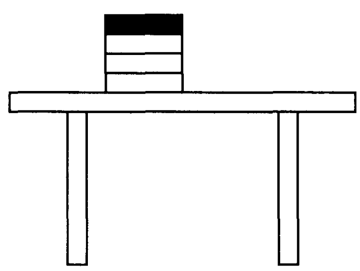
---

[3] I want to point out that it is a mistake to call the first problem "competence" and the second "performance" (the classic philosophical division) when they are logically symmetric.

(1) How humans come to know truths about abstract objects such as triangles
(2) How they recognize a particular drawing as an triangle
(3) How they can skillfully combine abstract ideas (such as "triangles can be of any size") and concrete ideas (such as "this square is four times the area of that triangle") to prove theorems.

I claim that it is this combination of abstract and concrete that allows humans to learn new things about abstract concepts from *experience* (surely we are not all born with the knowledge of the Pythagorean theorem) as well as apply abstract ideas to concrete situations, thereby handling situations that we never encountered before. Furthermore, I think that both the abstract and the concrete are governed by the same principles. One of the goals of this thesis is to show that the same global models of the world are behind human representation of abstract concepts as well as concrete perceptual categories.

One of the ways in which global models influence spatial semantics is by situating spatial concepts within a framework of physical-causal relations. For example, geometry alone can never capture the meaning of a spatial preposition; we need knowledge of other domains such as physics and function as well. Otherwise how can one explain why the sentence "The black book is on the table" is acceptable in figure 2.1 below?



The black book is on the table

**Figure 2.1**

The acceptability of the above sentence can be understood if we situate human knowledge of geometry in the context of causal physical models of the external world –in particular, it is very likely that notions of *support* enter into the meaning of ON.

The role of causal models in concept learning has been popularized in the form of the 'theory-theory' of concepts (Gopnik & Meltzoff, 1997). Similarly, the role of causal models in perceptual categorization has been recognized by many authors (Feldman, 1997; Richards et. al., 1996). Nevertheless, the mapping between concepts and perceptual categories has not been studied within a common framework. In general, there is a perfect symmetry between the abstract and the concrete – there are an infinite number of concepts and infinite number of world situations; each concept that has to be abstracted is also a concept that has to be applied. I propose that the right way to model this symmetry is not by studying concepts and percepts separately, but rather by making sure that the computational framework is agnostic to the difference between abstract and concrete (or mind and world if one sees it that way). What kind of framework can solve this problem for us?

One can address the issue of a global framework for connecting percepts and concepts at two levels, roughly, the *syntactic* and the *semantic* level. The syntactic component delineates the formal structures required to capture the elements needed in order to model spatial concepts. This component is quite general – it applies to concepts as a whole. The semantic component models the constraints on the structure of concepts, percepts as well as the constraints on the map between the two.

In the next two sections, I address the problem of spatial syntax. At the syntactic level, I argue that the formal framework needed to model the map from concepts to percepts motivates a different definition of concept learning. While the classical question of concept learning is "How can I learn the extensional meaning of a concept from examples?" in my reformulation the question is "How can I learn to map the various clusters of a concept on to each other". I argue that the semantics of spatial concepts is driven by *global models* of the world, where by 'global model', I mean a representation of all the features of the world that are relevant in a particular context. For example, a global model of the physical world may look (approximately) as follows:

(1) It is three dimensional
(2) Contains rigid objects with solid interiors and smooth boundaries.
(3) Objects move in piece-wise smooth trajectories.
(4) Collisions obey the law of conservation of momentum.

While the constraints (1)-(4) above do not enumerate all the features of the physical world, they do provide a template for all the top-level features of the physical world that are shared by rigid bodies moving in three dimensional space. In order to formalize the notion of a global model, we need a sophisticated syntactic language, namely, the language of category theory.

### 2.2. Spatial Syntax.

In order to understand spatial syntax, let me start with the standard account of abstract concepts. Suppose we start with trying to understand the meaning of the word "line". It is quite possible that children have one understanding of lines while adults have another. Nevertheless, for a conversation between the child and the adult to be initiated, the two must share *something* –even if the common element is structured in different ways by the two parties. This common element is often called a *natural kind* (Putnam, 1975; Carey, 1985; Gopnik & Meltzoff, 1997; Macnamara, 1999) which in this particular case is the kind LINE. A natural kind is an abstract mental representation that is the primary reference of the word "line". A particular case of a line is individuated by the natural kind LINE (figure 2.2).

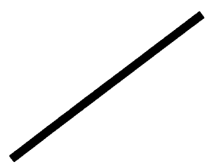


**Figure 2.2**

A natural kind is only a placeholder – it might be structured in different ways depending on the amount of experience an individual has with that kind. Typically, each kind is structured by a theory – i.e., a causal story of what that kind is about. For example, a newborn infant may know next to nothing about lines, and so its LINE kind is not structured at all, a normal adult may know a lot more, in which case the LINE kind may be structured by a folk theory which says something like "lines are geometric elements that connect points" while an adult mathematician may structure the kind LINE as "a one dimensional element of a vector space". Nevertheless, what everybody shares is a natural kind term LINE, as well as a theory formation device that can be bootstrapped from perceptual evidence.

However, the above analysis says nothing about how the concept is actually related to real world examples of lines. The issue of mapping from natural kinds to the world is not solved. How are the percepts of lines acknowledged and structured? In particular, if we look at figure 2.2 above, the perceptual input is a black, somewhat thick straight bar. In what way is this bar a line? There is a clear distinction between the perceptual category "Straight bar" and the abstract, geometric concept "line" and yet the two are closely linked. What is their relationship?

We cannot understand the relationship between perceptual categories and the abstract concepts unless we can put them on the same footing. In the same way that two people cannot even begin talking about an object unless they have a shared natural kind term, the perceptual and conceptual structures cannot map on to each other unless they share similar structures. In other words, there is a mapping problem at the level of categories, i.e., not between individual concepts and the external world but between *categories* of concepts and *categories* of world situations.

In this regard, note that the classical account of natural kinds has several elements that are essentially independent of their abstraction –natural kind terms are essentially placeholders that can be structured (perhaps in complex ways) by some kind of input. Given this intuition, the symmetry between mind and world suggests that for each perceptual experience, we will need

(1) A place holder (say, P-LINE) that is called up whenever an instance of that perceptual concept is detected – a perceptual kind so to speak.
(2) An (causal) explanation of what that perceptual input is about. Furthermore, in order to map the perceptual kind P-LINE on to the natural kind N-LINE and vice versa, we need
(3) A map F: P-LINE →N-LINE as well as a map G: N-LINE → P-LINE and an account of how these maps are structured.

To summarize, the above arguments suggest that a mapping between two kinds of representations can take place only when there is a common knowledge structure shared by the two representations. Therefore, we need two levels of representational structure:

(1) One level consisting of individual concepts and individual (concrete) world states.
(2) Another level consisting of *kinds* – conceptual as well as concrete.

Similarly, the problem of mapping spatial concepts to world situations (figure 2.3) has two aspects:

(1) The mapping of individual concepts to individual world situations.
(2) The mapping of categories of concepts to categories of world situations.

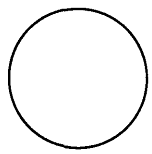
| | | |
|---|---|---|
| KIND | EQUATION | CIRCLE |
| INDIVIDUAL | $x^2 + y^2 = 1$ |  |

**Figure 2.3**

In other words, there is a mapping problem at the level of kinds as well as individual categories, i.e., the qualitative mapping problem. In order to solve the mapping problem, we need a formal language that can cope with qualitative mapping, i.e., a formal language in which we can handle abstract entities and concrete entities, kinds as well as individuals and the maps between all of these.

Devising an appropriate formal language is a non-trivial problem. For example, kinds are structured – the natural kind term LINE is structured by our theory of lines. The need for internal structure eliminates the traditional formal approaches to natural kinds, based on set theory because the basic elements of set theory, sets, have little internal structure and hence cannot possibly model even the simplest perceptual kinds. Instead, I believe that the correct language to talk about the mind-world mapping is category theory (Maclane, 1998, Lawvere and Schanuel, 2002). In the next section, I explain what a category is and show how categories can model natural kinds, perceptual kinds and the map between the two.

### 2.3. Category Theory.

**Definition**: A Category is a collection of *objects, a, b, c …* and *arrows, f, g, h …*such that

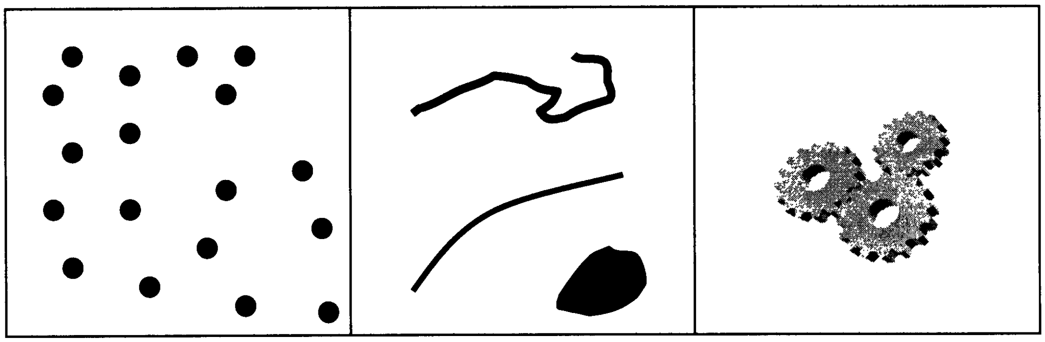
(1) For each arrow *f* we have an object *a* which is its domain, dom(*f*), and another object, *b* its codomain, codom(*f*). Graphically, we write $a \xrightarrow{f} b$.

(2) For each object *a,* we have an Identity arrow $Id_a$, graphically represented as, $a \xrightarrow{Id_a} a$.

(3) For each pair of arrows *f, g* such that codom(*f*) = dom(*g*), we have a composition operation, $\circ$ , which gives rise to a new arrow $g \circ f$ with domain, dom(*f*) and codomain, codom(*g*). Graphically if $a \xrightarrow{f} b$ and $b \xrightarrow{g} c$ then $a \xrightarrow{g \circ f} c$.

(4) Finally, the composition operation satisfies the law of associativity so that if $a \xrightarrow{f} b \xrightarrow{g} c \xrightarrow{h} d$ then $k \circ (g \circ f) = (k \circ g) \circ f$.

The definition of category is broad enough to cover all kinds of conceptual and perceptual domains, both structured and unstructured. For example, we have the category SET whose objects are all sets and whose arrows are functions from sets to sets. The category TOPOLOGY consists of all those geometric features and relations, such as *containment*, that are maintained when we subject an object or collection of objects to transformations such as stretching, expansion etc.



SET TOPOLOGY RIGID

**Figure 2.4. Three kinds of Categories**

To give an example of a structured category, consider the category RIGID(3D) of all three-dimensional solids. Here, the objects are rigid three-dimensional solids and the arrows are rigid transformations (translations, rotations) (figure 2.4).

Category theory affords a rich algebraic structure. A detailed discussion of the algebra of category theory is beyond the scope of this chapter (see Maclane, 1998; Lawvere & Schanuel, 2002) but I show how some of the algebraic structures of category theory can be used for my purposes. For example, since kinds have individuals and individuals fall into kinds, we should be to individuate categories as well as to map individuals to their categories. Similarly, we also need to map categories on to each other – think of mapping the kind LINE to the particular perceptual object in figure 2.2. In category theory, individuation and membership is accomplished using *individuator* and *member* operators. More precisely

**Definition:** Let $C$ be a category. Then the individuator operator Ind acts on $C$ and gives rise to an object $a$ in $C$. Graphically, Ind($C$) = $a$. Similarly, if $a$ is an object of a category $C$ then the membership operator Mem is defined as Mem($a$) = $C$.

Mapping between categories is done by a *Functor*.

**Definition**: A *Functor* is a map $F : A \to B$ where $A$ and $B$ are both categories such that
(1) For each object $a$ in $A$ we have an object $F(a)$ in $B$.
(2) For each arrow $\to$ in $A$ we have an arrow $F(\to)$ in $B$. Furthermore,
(3) $F(Id_a) = Id_{F(a)}$ where $Id$ is the identity operator.
(4) $F(g \circ f) = F(g) \circ F(f)$

The simplest functor is the *inclusion* functor from categories contained within each other. For example, there is a functor from the category RIGID(3D) (of all 3D solids) to the category SET(3D) (consisting of all subsets of 3D space). In this case the map is nothing but identity because each 3D solid is also a subset of 3D space. Therefore, the simple procedure of forgetting all the rigid structure converts a rigid object into a set. These operations of adding/removing structure are important – I will say more about them in my discussion of semantics.

Finally, we need ways of creating new categories from a given set of categories. For my purposes, the three important operations are *product*, *merge* and *separate*.

**Definition**. Let $C_1$ and $C_2$ be two categories and let $F_{12}$ be a functor mapping $C_1$ to $C_2$. Then
(1) The product operation $\otimes$, gives rise to a new category $C_1 \otimes C_2$ whose objects are pairs $(c_1, c_2)$ where $c_1$ and $c_2$ are objects in $C_1$ and $C_2$ respectively and whose arrows are pairs $(\to_1, \to_2)$ where $\to_1$ and $\to_2$ are arrows in $C_1$ and $C_2$ respectively.
(2) The merge operation $\oplus_{F_{12}}$ gives rise to a new category $C_1 \oplus_{F_{12}} C_2$ whose objects are pairs $(c_1, F_{12}(c_1))$ and whose arrows are pairs $(\to_1, F_{12}(\to_1))$ where $c_1$ is an

object and $\rightarrow_1$ is an arrow in $C_1$. Note that $C_1 \oplus_{F_{12}} C_2$ is a subcategory of $C_1 \otimes C_2$. Informally, the merge operations binds together the objects in one category with their images under a functor.

(3) The separate operations $\prod_1$ and $\prod_2$ act on $C_1 \otimes C_2$ such that $\prod_1(C_1 \otimes C_2) = C_1$ and $\prod_2(C_1 \otimes C_2) = C_2$, where $\prod_1(c_1, c_2) = c_1$ and $\prod_2(c_1, c_2) = c_2$ and $\prod_1(\rightarrow_1, \rightarrow_2)_1 = \rightarrow_1$ and $\prod_2(\rightarrow_1, \rightarrow_2)_1 = \rightarrow_2$ respectively.

Let us see how the category theory language can be used to model familiar spatial categories. In fact, it will be clear that even the simplest spatial concepts have a vast categorical structure implicitly hidden inside them. For example, consider the statement "Line B is longer than line A" illustrated in figure 2.5 below. It is clearly true, almost transparently so for most people.

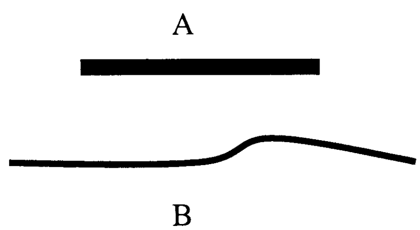



**Figure 2.5. Line B is longer than line A**

However, this outward simplicity hides a tremendously complex set of syntactic structures. On the one hand, the labels "Line A" and "Line B" refer to particular configurations of ink on paper, occupying particular locations. Therefore, the syntactic structure has to be able to represent concrete objects – their shape, size and location. Otherwise, how can the concepts have any reference in the world? On the other hand, the two shapes are not ideal straight lines –they have both thickness and curvature (clearly visible in case of line B). Furthermore, the two objects are being compared along an abstract property – length. Consequently, the syntactic structure has to be able to represent highly abstract properties as well. Let us see how this can be done using categories. In fact, the syntactic representation of the line has to be able to represent the various levels (concrete, abstract) *simultaneously*. How is this done?

We can model the above situation using three categories, that I call PIXEL, SEGMENT and LINE respectively. The definitions of the three categories are given below.

(1) The category PIXEL represents the distribution of black pixels in a coordinate frame defined by the white page boundaries. To be more precise PIXEL consists of a set of objects $(\alpha,\beta)$ where $\alpha$ is a parameter that can take two values WHITE and BLACK and $\beta$ is a parameter that represents the location of the object in the 2D coordinate frame of the paper. Furthermore, the PIXEL category also has metric information in the form of *similarity*. Given two distributions of pixels $\alpha_1$ and $\alpha_2$ there is a similarity metric $M(\alpha_1,\alpha_2)=\phi$ which measures the similarity between the two distributions.
(2) The category SEGMENT consists of 1D straight line segments in 2D euclidean space. The SEGMENT category also has metric information, in terms of *scale*. To each segment , $S$ , in 2D one can associate a number L( $S$ ) which is the length of the segment in the metric determined by the scale.
(3) The category LINE consists of all intervals (0,p) that are subsets of the infinite straight line $(0, \infty)$. In this category we have no metric information. However, LINE represents qualitative information as an interval calculus –i.e., the relations <, +, > between line segments that are defined as follows:
$(0,p) < (0,q)$ if and only if $p < q$.
$(0,p) = (0,q)$ if and only if $p = q$.
$(0,p) > (0,q)$ if and only if $p > q$.

One can describe functors between the PIXEL and SEGMENT as well as between SEGMENT and LINE. The functor between PIXEL and SEGMENT is the composition $\Omega = \Gamma \circ \Phi$ of two operators –

(3) The smooth derivative operator, $\Phi$, (a convolution of the normal derivative with a smoothing operator, say, a gaussian, see figure 2.6) that gives the smooth version of the derivative of a set of pixel values.

(4) The integral operator, Γ, (figure 2.6) integrates the derivatives to produce a line segment.

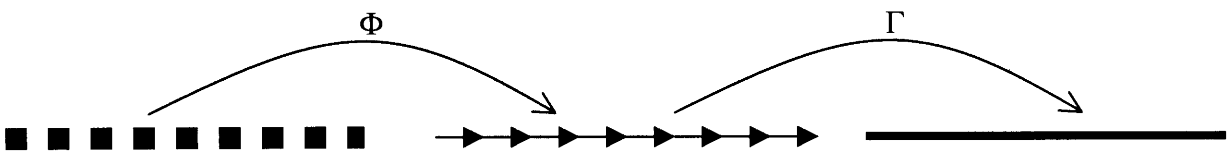



**Figure 2.6. Functorial transformations from PIXEL to LINE**

Similarly, there is a functor, Ψ, between the SEGMENT to the LINE category, where for each segment, $S$ , such that L( $S$ ) =p, Ψ( $S$ ) = (0,p).

Of course, the PIXEL-SEGMENT-LINE schema is not the only way to formalize the notion of line. We could, instead of PIXEL, have a concrete representation of a line that is temporal, say, the time taken to trace a mark on a piece of paper at constant speed. Nevertheless, however we formalize the concept of line, it will require multiple categories, some abstract and others concrete. For example, if we measure lines by the time taken to trace them, we need an abstract category that will tell us when two concrete 'traces' are the same line.

We can now answer the question posed in figure 2.5 using the three categories defined above. Each "line-object" is represented as an object in the merged category PIXEL $\oplus_{\Omega}$ SEGMENT $\oplus_{\Psi}$ LINE. The "line-objects" A and B have one component in PIXEL, another component in SEGMENT (where they become segments of length L(A)and L(B) respectively) and a component in LINE. In LINE, the comparison (0,L(A)) < (0,L(B) is available, answering the question posed in figure 2.5. We can see some interesting *emergent* phenomena here because the same term can refer to multiple "objects" - blotches of ink, a particular length and an abstract comparison – an option unavailable in PIXEL, SEGMENT and LINE individually, but becomes available when we form a merged category combining the three categories.

## 2.4. Categories and Natural Kinds.

As we saw earlier, even such an unproblematic concept as a line has a complex syntactic structure. Nevertheless, the formalism of category theory is adequate to the task of modeling the various meanings of the concept line. We can now generalize the approach to the relationship between *kinds* (that are conceptual objects) and *categories* (that are formal objects).

> **Definition**. The syntactic representation of a *kind* is a category *C*. The syntactic representation of a map from kinds to kinds is a functor between the two associated categories.

From the category theory perspective, the mapping problem from concepts to the world and back can be seen as mapping an abstract natural kind (for example, the kind represented by the category LINE in the analysis of lines in the previous section) on to a perceptual kind (for example, the kind represented by the category PIXEL in the analysis of lines). Note that this model is quite different from the classical model where the world is seen as providing *tokens* for an abstract *type*. In my model both the world and the mind consist of categories regulated by causal procedures. This is good, for the type-token distinction seems like an idealization unnecessarily skewed towards abstract concepts. The classical claim is that types individuate tokens, or in terms of the ideas in this chapter, abstract concepts individuate perceptual categories. But how is that possible?

Types are abstract structures while tokens are defined in terms of perceptual features. How can the link between the two be understood without invoking the mapping problem?

Instead, it seems reasonable to model concepts as well as perceptual categories within category theory, where objects, arrows and functors replace the classical model of concepts which models them as a combination of abstract entities (the intension of a concept) and exemplars (the extension of a concept. In this approach, each concept is a bundle of categories and maps, some which are long-term stable aspects, perhaps the one encoded in language, and the others that are more fleeting and are given in immediate perception. My claim is that concept learning is not learning the extension of a concept but rather learning the map between various categories associated by the concept. So how is this map structured? How is it learnt? These are the key issues for spatial semantics in the category theory approach. I argue that the map is learnt because there are global causal-spatio-temporal structures that constrain the categories and their functors. In the next chapter, I show how spatio-temporal constraints in the form of stability can be imported into the category theory language as constraints on categories and functors.

# Chapter 3: Semantics of Space

## 3.1 Introduction.

In the previous chapter, I argued that the syntactic aspects of spatial cognition should be modeled using category theory. Category theory offers a unified method to model the various levels involved in the representation of a concept, from the abstract to the concrete as well as the maps relating the levels to each other. As I showed, even a simple geometric concept such as **line** has to be modeled at multiple levels with a rich structure of maps connecting the various levels.

However, category theory by itself is only a language – it is a means by which the 'qualitative mapping' of structures that involve multiple levels and maps between levels can be modeled. As such, categories has no *content*, they are placeholders that can take any values whatsoever. In order to understand spatial concepts, we need to specify the content of the categories being used to model them, i.e., to specify which categories are necessary for modeling spatial concepts and which categories are not. For example, in order to model the concept **line**, I used three categories - PIXEL, SEGMENT and LINE, with two sets of maps connecting them to each other. While it might be possible the properties of the concept **line** with a different set of categories and mappings, it is clear that we do not need the category SET of sets in order to model the concept **line**. How do we know which category is useful and which is not?

This is a subtler issue than might seem at first. On the one hand, since the problem of conceptual representation is immensely complex, we have to make sure that the representational framework is rich enough that it can handle *all* the categories that are needed. On the other hand, there must be a scheme by which only those categories that are relevant to a particular task are called up – otherwise the problem of indexing into the right set of categories becomes too hard. The need for a framework that can incorporate all the relevant categories echoes Jerry Fodor's arguments for a language of thought (Fodor, 1975). Fodor points out that there is no end to the number of concepts needed to situate any given concept. In fact, Fodor claims that all concepts are irreducible, thereby saying that the semantics of concepts has no structure whatsoever. I argue that within category theory one can formulate a generative account of concepts that is sufficiently powerful that it can model all the relevant aspects of a concept and yet be constrained enough that it can be used to make inferences and generalizations. The determination of a generative framework satisfies the dual criterion of providing a global framework in which concepts are situated as well as providing a means by which individual concepts are learnt and augmented, without reference to the entire space of concepts.

One might ask, why is a generative framework needed? I have already given two answers to that question: namely, that a generative framework is needed to explain the regularities that we see in our categories and their inter-relations and also to reconcile the global context of a given concept with its local usage at given time and place. We need a generative framework for other reasons as well – here my arguments parallel Fodor's arguments for the existence of an innate language of thought[4]. The arguments have three main planks

(1) Sensibility: Without some intrinsic structure, the incoming sensory stimulus is unorganized and hence, random. Some structure is needed to make sense of the data.

[4] Except that I am arguing for a global structure of conceptual representation. The existence of global structure does not imply that the global structure is located in an individual human being's mind. The structure could well be distributed across the mind-world system

(2) Poverty of stimulus: The incoming stimulus is not enough induce the rather abstract structures constituting human knowledge of a particular domain.
(3) Particularity: The particular patterns of usage/behavior are not consistent with a general purpose system, i.e., additional constraints are needed to explain the pattern of behavior.

### 3.2 Semantic Constraints.

All of the above arguments apply to the case of global conceptual structure as well. I introduced the first argument in the previous chapter where I said that the initial structure has to involve kinds –both abstract and concrete. Without this structure, the relevant concepts cannot even be represented. What does this minimal structure needed to represent the space of concepts? In the previous chapter I argued for a minimal syntactic structure of categories –objects, maps and functors. However, syntactic structure by itself, is not enough. Let us see why.

Suppose we are trying to model a conversation between two people –say the conversation between Socrates and the slave boy. In this conversation, both Socrates and the slave boy need representations for natural kinds – concrete natural kinds such as particular figures drawn in the sand and abstract natural kinds such as squares that can take any size. On top of the individual elements of the dialogue, there is a global structure to the conversation as well, namely, the intent to prove the Pythagorean theorem. Since the conversation is non-random at the global level as well, there must be some representation of global structure, say, the knowledge that a theorem is going to be proven at the end. Local structure does not imply the existence of global structure.

In fact, iterating this argument, one can see that structure at any given spatio-temporal scale does not imply structure at some other spatio-temporal scale. Therefore, in order to represent a complex global regularity, there needs to be *some structure at all* *scales*. To summarize, purely on the grounds of sensibility, one can say that there must be categorical structure at all scales of a given task[5].

Secondly, echoing the poverty of stimulus argument, it is clear that observers have very little experience of any of the features –concrete categories, abstract categories and maps from one to the other. Consider the induction of abstract features - after given a few examples, people generalize the spatial concept to examples far removed from the examples presented to them, suggesting that the abstract structure was not learnt from data. Similarly, the fact that new concrete instances can be classified as belonging to the concept suggests that the map from abstract to concrete was also present in some form.

Finally, every spatial task, whether it is proving a theorem or cooking a dish, involves stringing together a particular list of spatial primitives while ignoring irrelevant features of the environment. One of the most puzzling aspects of human behavior is the way in which we move smoothly from task to task, each one of which requires a specialized set of spatial skills. The smooth movement from task to task suggests the existence of *dynamical* representations of task structure, at *multiple spatio-temporal scales*. All of these considerations put together argue for the existence of a *global generative procedure* satisfying the following constraints (Witkin & Tennenbaum, 1985; Watanabe, 1985).

(1) *Completeness*. Any generative procedure should be able to generate *all* the categories needed to model a spatial domain. In particular, the generative procedure should be able to generate abstract as well concrete categories.
(2) *Closure*. Since the generative procedure is complete, none of the elements in the procedure should come from the outside, because (by definition), there are no categories outside the generative procedure. Therefore, all the elements in the procedure must be definable internally, i.e., it should be a closed system.

[5] There is the question of how this structure at all scales is assembled for a given task so quickly – this is a hard problem, but I will put this question off till the issue of dynamics takes center stage.

(3) *Dynamicity.* As I argued before, the elements in a particular task are generated on the spot and used in creative ways. Therefore, a generative procedure must be task specific, i.e., it combines elements necessary for a particular task, from the local features (that can be both abstract and concrete) to the global features (which can also be abstract and concrete) such as the theme for the task. All of these elements have to accessed dynamically and then combined to solve the task.

In the rest of this chapter, I show how these constraints can be satisfied within the formal framework of category theory. Before I go on to more technical discussions, let me describe informally, what needs to be done. Since all of the concepts are spatial concepts and presumably make reference to spatial relations in the real world, we need to generate a list of spatial categories that are consistent with spatial relations in the world.

We can think of each category as a *model* of the world. Furthermore, each "world" generates a set of categories, namely all the categories that are models of the "world". Therefore, the set of spatial categories should be all the models of the world. What should this world be?

### 3.3 The Global Model.

First, let me show how any category can be used to generate further categories by means of a natural partial ordering on categories. After that I show how this abstract, algebraic construction (given below) is equivalent in the case of geometric categories to a topological construction using the notions of invariance and stability. To start with the algebraic construction, define the functorial and inclusion relations between categories as follows:

**Definition:** Let *C* and *D* be two categories. Then:

(1) *C* is functored in *D*, formally $C \prec D$, if there exists a functor $F: C \to D$.

(2) *C* is included in *D*, formally $C \mapsto D$, if there exists a functor $F: C \to D$ and a functor $G: D \to C$ such that $G \circ F = Id_C$.

Note that $\prec$ and $\mapsto$ are both transitive operations, hence the functorial and inclusion relations are *partial orderings* among categories..

**Definition:** Let *C* be a category. Then define the Categorical Hierarchy and the Restricted Categorical Hierarchy as follows:

(1) The Categorical Hierarchy of *C* consists of the set of all categories $B_C$ that C functors in, i.e., the set $\langle B_C : C \prec B_C \rangle$.

(2) The Restricted Categorical Hierarchy is defined with respect to a larger category *D*. Formally, the Restricted Categorical Hierarchy of *C* and *D* consists of the set of all categories $B_C$ that C functors in, i.e., the set $\langle B_C : C \prec B_C \mapsto D \rangle$.

Informally, the restricted hierarchy is a way of generating all the categories bounded on the one end by C and on the other by D. Therefore, given two categories *C* and *D* where *C* is included in *D*, there is a whole set of new categories that are generated by *C* and *D*.

The above algebraic treatment can be fleshed out in a concrete situation in many ways. For our purposes, it seems reasonable to assume that spatial concepts are meant to aid the functioning of human beings acting in a dynamic spatio-temporal environment. If that is the case, the appropriate "world" is the world of an organism moving in 3D space in a field of other objects. Typically, animals are continuously moving within a physical environment where important stimuli (food, shelter, mates, predators) are continuously changing their location. Nevertheless, when thought of as physical objects, all of these stimuli share a basic structure, which bind together most spatial tasks. These common elements are:

1. A gravitational environment with gravity pointing up.

2. An autonomous agent (the perceiver), and possibly other agents.
3. Solid objects of randomly distributed shapes, sizes and locations that satisfy the following physical laws:
    a. An object is at one place at one time
    b. An object has fixed shape and size and mass.
    c. An object moves in a continuous path.
4. Most of the objects are motionless during the task, though some may move due to forces exerted by an agent or by some other object.
5. All of the agents and objects are located in an environment suffused with light obeying the principles of geometric optics.

For the moment, if we restrict ourselves to the purely geometric elements of the above parameters, we can formalize the spatial environment as follows:

(a) The ground plane can be modeled as an infinite flat plane. Surfaces (of objects) can be modeled as smooth two-dimensional manifolds distributed over the ground plane. Because the scene is always viewed by an observer, we also have one coordinate frame in which the observer occupies the origin of the ground plane, though this is not necessary in all coordinate frames.

(a1) Formally, the ground plane, *GP*, is defined below:

$$GP = (x, y, z) : z \geq 0.$$

(a2) World surfaces form a collection of smooth manifolds $S_i$.

(a3) In some cases, the observer, *O*, maps onto the origin (0,0,0).

(b) Since organisms move constantly, the sensible portion of the environment is a dynamic entity. For simplicity, let us assume that all objects in the spatial environment are rigid. Then, transformations of the environment can be divided into two types:

(i) Transformations due to rotation in depth.

(ii) Transformations due to translation.

Rotation in depth is caused by changes in the observers' viewpoint and also by rotating objects. Translation is caused by moving toward or away from an object or by an object moving toward or away from the observer. We can model rotations and translations in the following manner:
Environmental transformations are parametrized by the space of affine transformations - $T_{aff}$ - of the ground plane. $T_{aff}$ splits into rigid rotations around the Z axis and into translations along the ground plane.

(c) The spatial environment is intrinsically gravitational with a universal downward direction and a ground plane that supports objects. Therefore, the problem of spatial representation is not a purely geometric problem, but also a physical one, with gravitational notions such as support entering into spatial representation from the beginning. Formally, there is a designated direction vector $v_g$ given by the positive $z$ axis.

Summing (a)-(c) above, we can define the spatial environment as

$E = GP \wedge T_{aff} \wedge O \wedge v_g \wedge S_i$.

In the language of category theory, we can describe the category $E$ as the category defined by the upper half space in three dimensions, whose *objects* are solid objects bounded by 2D surfaces and whose *arrows* are rigid transformations of 3D space. I claim that this model of the environment, $E$, can serve as a generative structure for spatial categories.

In our case, since we want the categories to be spatial, we can impose the following constraint:

*All the categories are included the universal model* $E = GP \wedge T_{aff} \wedge O \wedge v_g \wedge S_i$.

In the next two sections, I show how the universal model E above can be used to generate spatial categories, first in the static situation and then in the dynamic situation.

**3.4 Statics**.

In the static case, a classification of geometric categories based on techniques from topology (Poston & Stewart, 1978) show that the *algebraic* definitions of inclusion and functoriality are equivalent to *geometric* definitions of categories based on the mathematical ideas of **co-dimension** and **invariance**. Co-dimension is defined as follows:

**Definition**. Let $M$ be a n-dimensional manifold and let $S$ be a m-dimensional sub-manifold of $M$. Then we say that $S$ has co-dimension m-n.

Now suppose that a group $G$ of transformations acts on $M$. We can then classify sub-manifolds of $M$ by looking at the action of $G$ on the sub-manifolds. In particular, given a sub-manifold $S$, an important set is the subset of transformations $G_S$ of $G$ that leave $S$ invariant, i.e., takes points of $S$ to points of $S$. Formally,

**Definition**: $g \in G_S$ if and only if $g(s) \in S, \ \forall s \in S$.

Then one can define an ordering, $\prec$ , on the space of sub-manifolds, $\Sigma_M$ , in the following way:

**Definition**: $S \prec T$ if and only if $G_S \subseteq G_T$.

It is easy to verify that $\prec$ is a partial ordering on $\mathbf{\Sigma_M}$ and consequently, $\prec$ defines a partially ordered lattice $L_M$ on $\Sigma_M$. I call $L_M$ the *Structure Lattice*. Going up the structure lattice is equivalent to finding sub-manifolds that are more invariant with respect to $G_M$. Let us take a concrete example in order to understand $\prec$ . Consider the space of triangles $T$. This space is a three-dimensional manifold, with each point corresponding to a triangle (figure 3.1). Now suppose your measurement system has two features:

(1) The first feature, =, measures when two sides of a triangle are the same.
(2) The second feature, $\perp$ , measures when there is a right angle in the triangle.

How is the space of triangles categorized by the combination of these two features? Picking a triangle at random almost always implies (with probability 1) that none of the sides of the triangle will have the same length, or that any of the angles will be a $90^\circ$ angle. In fact, the space of isosceles triangles - $T_{iso}$ - forms a two dimensional sub-manifold of $T$. Therefore, - $T_{iso}$ has co-dimension 1. The space of right-angled triangles – $T_{right}$ is also two dimensional, and therefore, has co-dimension 1. Finally, the space of equilateral triangles as well as the space of right-angled isosceles triangles – $T_{equi}$ , $T_{r\text{-}iso}$ - are 1-dimensional sub-manifolds and consequently have co-dimension 2. Now consider the group of transformations given by linear shears and dilations, where a shear is a transformation that consists of moving a vertex along a predetermined direction.

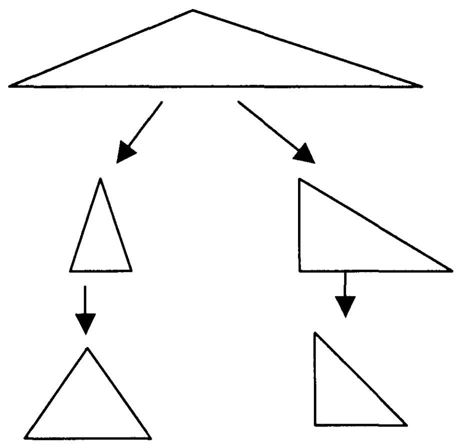


**Figure 3.1: Space of Triangles**

$T$ is invariant under all of these transformations. $T_{iso}$ is invariant under all dilations as well as those shears that are based on the vertex between the two equal sides and in a direction perpendicular to the bisector. $T_{equi}$ is invariant only under dilations. Therefore the following relation holds:

$$T_{equi} \prec T_{iso} \prec T$$

The above techniques can also be applied to $E$. Classifying environmental properties according to their invariance with respect to $T_{aff}$ leads to subdivision of $E$ into three broad categories: Coordinate frames, Topological Structure, and Metric Structure. Coordinate frame and Topological Structure are qualitative features of **E** while Metric

Structure is a quantitative feature of **E**. Coordinate Frame, Topological Structure and Metric Structure satisfy the following relationships.

(1) Metric Structure ∠ Coordinate Frame
(2) Metric Structure∠ Topological Structure

Coordinate Frame, Topological Structure and Metric Structure are complex representations with the following sub-representations:

(1) Coordinate Frame Representations: In order of decreasing invariance, we have three Coordinate frame representations- Gravitational Frame, Blob-location, Principal Axis-location and Minor Axis-location.

(1a) Gravitational Frame, (GF), is the universal vertical frame defined by gravitation. It is a global coordinate frame since it is valid at all spatial locations.

(1b) Blob-location (BL), represents a Blob at a particular location. Similarly, Principal Axis-Location (PA), represents a principal axis of an object at a given location while Minor Axis-Location (MA), represents a Minor Axis at a given location.

(2) Topological Structure Representations:

(2a) Blobs (B), are undifferentiated spatial representations, with no internal structure. Therefore, they are the most invariant representation of an object.

(2b) Dimension (D), gives the inherent dimension of an object, i.e., whether it is 0, 1, 2 or 3 dimensional.

(2c) Generic spatial invariants (GI), are those spatial invariants that do not change when the stimulus is perturbed slightly. For example, Containment and Closure are generic invariants. Non-generic invariants, (NGI) are those spatial invariants that change when perturbed along a certain direction. However, non-generic invariants are still topological invariants because they do not depend on metric properties of surfaces. Contact is a good example of a non-generic invariant.

(2d) Higher Topological Invariants are topological invariants such as Sphere-like and Torus-like that involve computing a global topological property of the stimulus.

(3) Metric Structure Representations. While the previous two categories were both qualitative since they do not involve numerical quantities, Metric Structure captures the quantitative structure of the environment.

(3a) Global Properties, (GP) are quantitative properties -such as Area and Diameter- that are properties of an object as a whole.

(3b) The Part Structure Hierarchy (PSH) contains the decomposition of the object into parts. Furthermore, by looking at the major and minor axes one can classify a part as convex or concave. This classification in terms of convexity gives rise to the Convexity Hierarchy, (CH). (3c) Finally, Riemannian Structure (RS) captures the detailed surface structure of objects. This is the level that is often called the 2.5 D sketch (Marr, 1982).

When classified according to invariance, Coordinate Frame, Topological Structure and Metric Structure can be further decomposed into the partially ordered categories given below:

(1) Coordinate Frame: MA $\prec$ PA $\prec$ BL $\prec$ GF.
(2) Topological Structure: HTI $\prec$ NGI $\prec$ GI $\prec$ D $\prec$ B.
(3) Metric Structure: RS $\prec$ CH, PSH $\prec$ GP.

The Structure lattice of **E** is the collection of all the partially ordered sequences given above. To conclude, a classification of spatial categories based on stability leads to a partially ordered set of spatial categories.

**3.5 Dynamics.**

So far my analysis of categories has been purely geometric. I showed how spatial categories (where the term "category" is defined in the technical sense of chapter 2) generate a partial order that is the same partial as the one by using stability as a criterion, based on non-accidental features. We can extend the above generative procedure to the dynamic case, where along with geometry, we also have notions of time and change (Lowe, 1985).

For example, take the concept *across*. Like other prepositions (Talmy, 2000), across can be understood as a spatial relation of the form:

Across Spatial-Relation(Figure, Ground)

The applicability of *across* in a given situation, depends not only on the shape of the figure and ground objects, but also the manner in which the figure object moves and the physical relations between the path of the figure and the ground (figure 3.2).

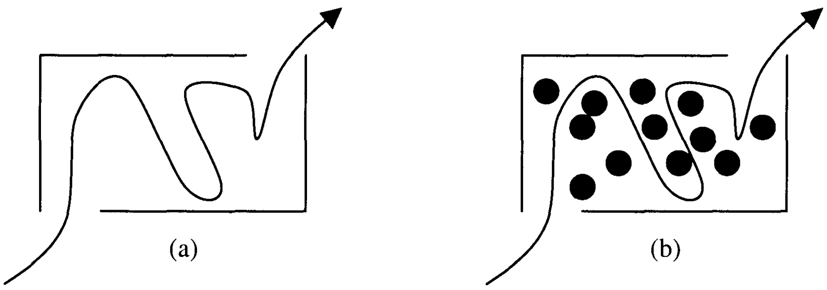



Figure 3.2: Assume that a person, John, is standing at the location marked 1 and he needs to walk across a rectangular "Park" to the location marked 2. Is the curve in 3.2(a) a valid case of "John walked across the park? What about the path in figure 3.2(b)?

I believe that most spatial concepts are like *across*, highly relational and grounded in the physical world. Any account of such concepts must include an account of their physicality. Another reason to pay attention to dynamics is because spatial concepts, whether seemingly independent of context or not, are used by humans in contexts that seem appropriate. Think of a city map –on the one hand it is a spatial configuration that represents an aspect of a cities spatial structure independent of where an observer is, but on the other hand, it is used creatively by individuals to make their way in a particular spatial context. How can we carve up the dynamical world into categories? Once again, I turn to the environmental model *E* and try to understand its dynamical regularities. The basic intuition is the following:

> **Modal Hypothesis** (Richards, 1988). The world is intrinsically structured into stable dynamical states and events, that recur again and again due the nature of

the world itself (for example, unsupported objects will always fall to the ground). Therefore, ground (mental) spatial categories in intrinsic dynamical regularities of the world.

More formally,

**Definition**. A dynamical system as a smooth function $F$: $\mathbf{R}^N \rightarrow \mathbf{R}^M$ for some M and N.

**Definition**. A stable invariant $I$ of the dynamical system $F$ is a function satisfying $I \bullet F = I$, i.e., $I$ does not change when $F$ is applied to it.

If we agree that stability is the link between categories and the physical world, we can categorize physical systems using stability as a classificational device. The classification is general and applies to any dynamical system (Poston and Stewart, 1978).

**Definition**. Let us assume that the dynamical system is smooth. Every smooth dynamical system (locally) can be expanded in a Taylor series of the form:

$F = F_0 + F_1 + F_2 + F_3 + \ldots$ where $F_i$ is the i$^{th}$ derivative of $F$.

Then,

(1) The stable states of a dynamical system are the same as the extrema of the derivative of the system.
(2) From the Taylor series formula and (1) above, it follows that the stable invariants are those states for which $F_i = 0$ for some i. Then define the set of stable dynamical states, Stable($F_i$) as follows
(3) Stable($F_i$) = The set of states $\{x: F_i(x) = 0\}$.

For complexity reasons (see Feldman, 2001), we need not represent higher order terms of the dynamical system, which allows one to assume that the system is quadratic. In this case, the only stable states represented for categorization are those for which $F_1 = 0$ or $F_2 = 0$, i.e, Stable($F_1$) and Stable($F_2$).

**Definition**: A dynamical category is either a first order or second order stable set of a dynamical system.

These are (see figure 3.3), in increasing order: the fixed points Stable($F_1$), invariant sets Stable($F_1$), stable motion Stable($F_1$) (often with a fixed direction) and the second derivative, i.e., sharp changes in trajectories Stable($F_2$).

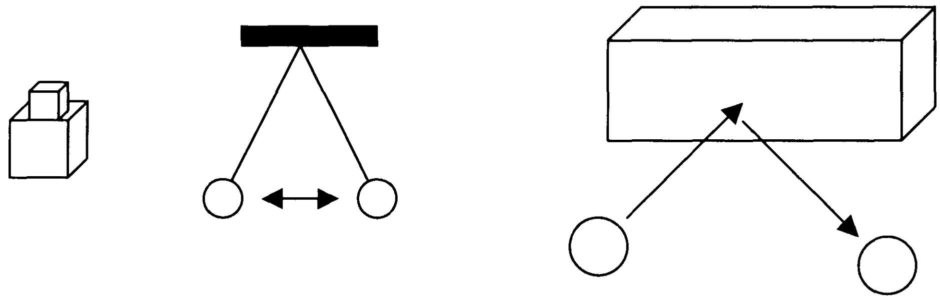


**Figure 3.3**

However, dynamics is not about the stable states alone. After all, the world does change! From a categorical standpoint, we can assume that change is registered only if it leads to a change of a stable state. A ball travelling in a straight line is dynamically irrelevant.

> **Definition**: Change is a transition from one stable state in Stable($F_1$)$\cap$Stable($F_2$) to another stable state in Stable($F_1$)$\cap$Stable($F_2$).

Finally, we note that change can be localized, i.e., sometimes, we can assign an object responsible for the change. In the formalism developed above, we can define a dynamical map as follows:

> **Definition**: A dynamical map $D$ is a map $D$: Stable($F_1$)$\cap$Stable($F_2$) $\rightarrow L$ where $L$ is the set of loci.

This concludes the exposition of the generative model of categories, i.e., the formal content of the semantics of spatial categories. In the next section, I show how these categories can be mapped on to each other and how these mappings can be assembled sequentially to perform spatial tasks.

### 3.6 Mappings and Derivations.

In the previous section, we saw how global models generate categories using their partial orders. In a given task, the subject will combine many of these categories to perform the task. Suppose you are in your kitchen, cooking and eating dinner with a few friends. The objects in the kitchen change location and function as the evening proceeds. The kitchen table, which is against a wall, starts as a surface for cutting vegetables. Then it becomes a surface around which the family gathers and eats. At this time it is moved to the center of the kitchen so that everybody can gather around the table. Finally, it may serve as a repository for dirty dishes. The same goes for other kitchen furniture and utensils. During the period of an evening, you and your guests move smoothly from one location to another interacting with multiple objects in the kitchen which may well be changing location as well. Nevertheless, a functional adult human being will smoothly transition between all these activities. While I cannot claim to understand the computational structure behind such complex tasks, we can try to see how such dynamic processes can be modeled in simple situations.

The dynamical mapping problem consists of two parts:

(1) Given two Categories $C_1$ and $C_2$, how can we map them on to each other, to create a new category $C_3$?
(2) Given a sequence of Mappings, $\{F_i : C_{i1} \rightarrow C_{i2}, i = 1, 2...n\}$, how can we sequence them together to perform a task?

Let us look at question (1) above first. Consider the two circles in figure 3.4a. Suppose we want to combine the two elements together into one picture. How many distinct states are created when they are combined? Intuitively, it seems that the distinct states are given by the possibilities in figure 3.4 b.

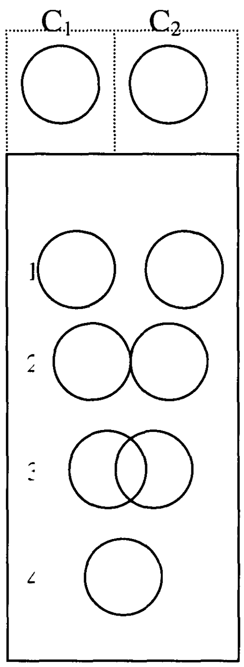



Figure 3.4a

Figure 3.4b

**Figure 3.4**

The mapping problem arises in several domains –in categorization, in generalization, in mapping linguistic categories to perceptual categories. In the formal framework of Category theory, devised in the earlier chapters, there are four kinds of maps that concern us:

(1) Maps from Categories to Categories
(2) Maps from Categories to Objects
(3) Maps from Objects to Categories
(4) Maps from Objects to Objects.

For example, the Pythagorean theorem is a statement about the category of triangles. However, all the examples given by objects are individual objects within the category of triangles. Therefore, learning the theorem is equivalent to learning a map from objects to the underlying category (of triangles). One can ask whether there are any general rules constraining the mapping of categories. In general, there cannot be an answer to this question. However, within the spatial domain, one can make several assumptions constraining the mapping of categories. These are

(1) The State Space of all possibilities is a manifold **M**. For example, in the case of the two circles in figure 1, the state space is four dimensional, corresponding to the four degrees of freedom of the two circles (each circle is allowed to move independently in 2D space).

(2) The State space is generated by codimension and deviations from transversality. Co-dimension is defined as follows (Feldman, 1997):

 **Definition**. Let **M** be a **m**-dimensional manifold and let **N** be a m-dimensional sub-manifold of **M**. Then we say that **N** has co-dimension m-n.

 Transversality is defined as follows:

 **Definition**. Let **M** be a **m**-dimensional manifold and let **S** and **T** be sub-manifolds of **M** of dimensions **s** and **t** respectively. Then **S** and **T** intersect transversally if
 $\text{Codimension}(\mathbf{S} \cap \mathbf{T}) = \mathbf{s} + \mathbf{t} - \mathbf{m}$
 Similarly, **S** and **T** intersect non-transversally if
 $\text{Codimension}(\mathbf{S} \cap \mathbf{T}) < \mathbf{s} + \mathbf{t} - \mathbf{m}$
 In every manifold, the non-transversal intersections are transitions between regions of transversal intersections. Constraint 2 above implies that spatial categories are marked by transversal and non-transversal intersections of sub-manifolds.

(3) The categories have low codimension. In principle, if we have many sub-manifolds, the total number of choices explodes exponentially, with the number of manifolds involved. In order to cut down the complexity of the categories, we need to impose a constraint that prevents highly complex categories (those with high codimension) from being represented.

(4) All information is meaningful. Information is provided to the observer in the form of examples of the two categories that are being mapped. The meaningfulness constraint can be divided into two parts:
   (a) Any information provided explicitly is seen as a necessary condition for the category.
   (b) Any information not provided explicitly is seen as irrelevant to the category.

   For example, in the case of figure 3.4, we can assume that the fact that the figures are circles is important, since it is given explicitly, while the fact that they are not given any designated location means that their location is irrelevant. One can also see this constraint as saying that information explicitly specifies the *free* and *bound* variables. The free variables are those that can take any value whatsoever, while the bound variables are those that have to take a particular value or range of values.

Now we can see how these constraints can be used to generate the set of categories in figure 3.4b. Constraint 4 suggests that the state space should take into account all the free variables (the position of the two circles), while restricting the mapped category to the bound variables, namely, two circles. Furthermore, from the point of view of transversality, we are looking at the space of all possible intersections of two 1-dimensional submanifolds (the two circles) in two dimensional space. If we look at the space of all possible intersections, the lowest codimension is for situation 1 and for situation 3 that are both transversal. Situations 2 and 4 are non-transversal and are at the boundary of transversal intersections. Therefore, the categories generated by mapping the two circles on to each other is the set illustrated in figure 3.5.

While the informal reasoning given above appears plausible, one could ask whether there is a formal basis for making the above judgement. The answer, of course, is "yes". To see how the formation of categories has a formal basis, one should go back to the correspondence between the geometric relations of transversality and the algebraic relations of partial order.

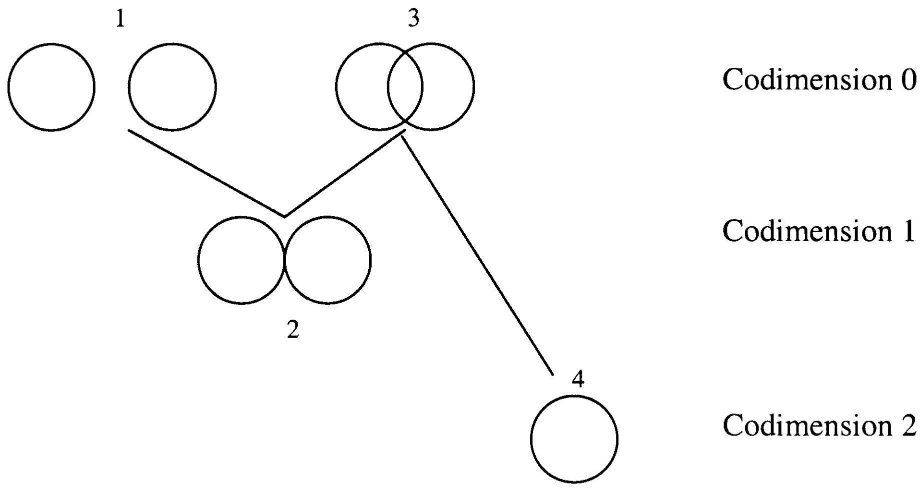



**Figure 3.5**

The classification of geometric categories based on techniques from topology (Poston & Stewart, 1978) shows that the *algebraic* definitions of inclusion and functoriality are equivalent to *geometric* definitions of categories based on the mathematical ideas of co-dimension and invariance. In other words, the generative procedure based on transversality is a *partial order.*

Therefore, from an algebraic point of view, one can restate the fundamental question, (1), in the language of partial orders as follows:

> (1') Suppose there are two partially ordered categories $C_1$ and $C_2$, how can we map them on to each other, to create a new partially ordered category $C_3$?

In principle, there are an infinite number of ways to generate a new partially ordered set. Recall that we have assumed that all our spatial categories are subcategories of the spatial environment $E = GP \wedge T_{aff} \wedge O \wedge v_g \wedge S_i$. Therefore, given any two categories $C_1$ and $C_2$ the new category $C_3$ can be any partially ordered subcategory of $E$. However, given constraint 4 above, the correct choice of a new category should be one that takes into account the information in the two categories being mapped and nothing else. Consequently, we can define a minimal category $C_3$ as follows:

> **Definition**: Let $C_1$ and $C_2$ be two partially ordered categories that are both subcategories of a supercategory (in our case the spatial environment *E*). Formally,
>
> $C_1 \mapsto E, \quad C_2 \mapsto E$
>
> Then the *mapped* category $M(C_1,C_2)$ is the *minimal* subcategory $C_3$ of E that includes both $C_1$ and $C_2$. Formally,
>
> $C_1 \mapsto M(C_1,C_2), \quad C_2 \mapsto M(C_1,C_2)$ *and* $\forall\, C : C_1 \mapsto C \,\&\, C_2 \mapsto C; M(C_1,C_2) \mapsto C.$

One can see how this definition gives rise to the correct answer for the mapping problem given in figure 3.4. In both categories, the free variables are all possible rotations and translations of the circles, while the bound variables are the topological states of the figures (interior, exterior, boundary). The constraint of minimality suggests that the mapped category should observe the free as well as the bound variables of the two underlying categories. Therefore, the mapped category is the category of distinct topological states of two circles that are allowed to translate and rotate freely in space, which is exactly the set given in figure 3.5. In this way, one can see how the four constraints provide a computational theory for mapping spatial categories.

Finally, we need to see how maps can be combined in sequence, in order to understand the mapping problem (Feldman, 2001). I argue that the dynamic problem can be solved using "derivations". By *derivation*, I mean the following:

> **Definition**: Let $\Gamma_1,\Gamma_2,\ldots\Gamma_n$ be categories (in the sense of chapter 2) and let $P_1,P_2,\ldots P_n$ be the partial orders generated by these categories. Let $C_1,C_2,\ldots C_n$ be concepts, i.e., nodes in the respective partial orders. Furthermore, let the categories $\Gamma_1,\Gamma_2,\ldots\Gamma_n$ satisfy the following constraint
>
> $\Gamma_n \mapsto \Gamma_{n-1} \mapsto \ldots\ldots \mapsto \Gamma_1$

(i.e., for all i, $\Gamma_i$ is included in $\Gamma_{i-1}$ where *inclusion* is defined in the sense of chapter 2)
Then, a *derivation* is a sequence $C_1 \rightarrow C_2 \rightarrow \ldots C_n$ of concepts where $C_i \in \Gamma_i$ and the arrows, $\{\rightarrow\}$, are dynamical maps in the sense of section 3.5.

Intuitively speaking, a derivation is a "fleshing out" of a core concept, which in this case would be $C_1$. Each further concept $C_2, \ldots C_n$ in the sequence is a richer version of the previous concept. Furthermore, the fleshing out consists of a sequence of non-accidental features linked via maps that are dynamically non-accidental, i.e., stable dynamical maps. In the next three chapters, I show how the derivational framework can be used to model domains where spatial relations are mapped on to the world.

## Chapter 4. The Semantics of Prepositions

### 4.1. Introduction.

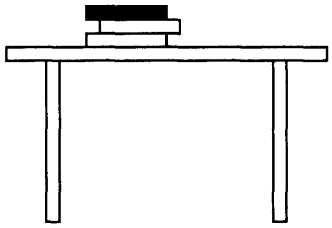
In this chapter I want to use the derivational method to model the semantics of spatial prepositions in English such as ON, IN, ACROSS and others. In particular, I want to understand

(1) How prepositions are used to convey information about concrete reality.
(2) How the meaning of prepositions can be generated systematically.

Prepositions have been studied rather extensively within the semantic literature because they are often seen as the easiest and simplest relational terms for which one can give an account for their meaning (Landau & Jackendoff, 1993; Herskovits, 1986). In addition, unlike verbs and nouns, prepositions are a rather small class of words in all languages, perhaps even a closed class, i.e., a class of terms that is fixed and cannot be extended. Initial attempts tried to model the meaning of a preposition as a relation between two objects (Miller & Johnson-Laird, 1976). For example, one can try to define ON as follows:

**Definition 1**: ON(A,B) ≡ CONTACT(A,B)

Unfortunately, this definitional account of prepositions fails rather systematically. A counterexample to the definition (1) of ON given above is shown in figure 5.1.

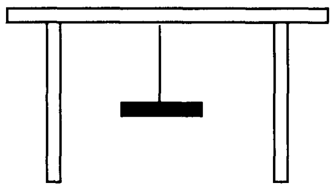


**Is the black book on the table?**
**Figure 5.1**

Most people would agree that the book is on the table, however the book does not touch the table. One could try to modify the definition as follows:

> **Definition 2**: ON(A,B) ≡ SUPPORT(A,B).

This definition also fails as is shown in figure 5.2.



**Is the black book on the table?**
**Figure 5.2**

And so do the definitions

> **Definition 3**. ON(A,B) ≡ CONTACT(A,B) ∨ SUPPORT(A,B) (falsified by figure 5.2)
> **Definition 4**.ON(A,B) ≡ CONTACT(A,B) ∧ SUPPORT(A,B) (falsified by figure 5.1)

In general, there seems to be no simple way to define prepositions using features and logical connectives (Siskind, 1991). An alternative approach to the modeling of prepositions is to assume that prepositional meanings are based on idealized cognitive models (ICM's for short), an approach popular in much of cognitive linguistics (Lakoff, 1987). In the ICM approach, prepositions have idealized meanings but these idealized meanings are not definitions, i.e., they allow exceptions. While the ICM approach can solve problems raised by the definition approach, it is not satisfactory from a theoretical perspective. The definitional account has the virtue of having a strong computational principle guiding the theory, namely:

> **Principle**: If **P** is a preposition and **f** is a feature, then **f** ∈ Meaning(**P**) if and only **f** ∈ Definition(**P**).

For the ICM approach, while there is a computational principle guiding the core of the meaning of the preposition (the meaning is given by the idealized cognitive model, there is no principle stating how the meaning of a prepositions can depart systematically from an idealized meaning. Therefore, the ICM approach is open to the criticism that the use of idealized meanings is ad hoc. In this chapter, I will show how to generalize the ICM approach using the derivational method in a way that systematically extends the meanings of prepositions away from idealized meanings.

Another problem for both the definitional and idealized meaning approaches is that these approaches do not say anything about the way the meanings of prepositions are generated. Are the meanings of prepositions random or are they systematically generated using some principle? If so, can one arrange their meanings on some kind of structure? For example, is there a systematic way of relating the meanings of ON, IN, ACROSS? In this chapter I will show that the meanings of prepositions can be generated from a partial ordering based on global models as one would expect from the framework developed in chapter 3.

Finally, one can ask how prepositions pick out concrete entities in the world as their referents. Note that this problem is not trivial, since prepositions are abstract entities, while the world is concrete.

To summarize, an adequate account of spatial prepositions must be able to answer the following three questions:

(1) How can we model the multiple sources of meaning of prepositions?
(2) How can we generate the meaning of prepositions systematically?
(3) How can we map the abstract content of prepositions on to the real world?

In the next section, I propose a new theory of prepositional semantics that answers all of the above questions.

### 4.2 A Theory of prepositional semantics.

The semantics of prepositions can be understood in terms of *derivations* where the notion of derivation was defined in chapter 3. In the case of prepositions, I propose the following semantic hypothesis:

> **Hypothesis**: Let $\Gamma_1$,$\Gamma_2$, $\Gamma_3$, $\Gamma_4$ be categories (in the sense of chapter 2) and let $P_1$,$P_2$, $P_3$, $P_4$ be the partial orders generated by these categories. Let $C_1$,$C_2$, $C_3$, $C_4$ be concepts, i.e., nodes in the respective partial orders. The meaning of a preposition is a four step derivation of concepts $C_1$, $C_2$, $C_3$, $C_4$ where:
> (a) $C_1 \in \Gamma_1$ is the topological, core meaning of the preposition.
> (b) $C_2 \in \Gamma_2$ is an abstract augmentation of $C_1$.
> (c) $C_3 \in \Gamma_3$ is a concrete referent of $C_2$.
> (d) $C_1 \in \Gamma_4$ is a concrete augmentation of $C_3$.

In the rest of this chapter, I will flesh out the derivation theory of prepositional meaning. Before I go on to a detailed exposition of the "derivation" model of prepositional meaning, let me make a few general remarks about the semantics of prepositions.

**4.3. Generating the meaning of prepositions.**

The category theoretic approach to generating meanings depends on the existence of global models that are relevant to a particular domain. These models encode the representations that are functionally relevant to the tasks being performed by an agent who is embedded in that domain. What is the functional importance of spatial prepositions?

One important property of spatial prepositions is that they are locational, i.e., they are used to communicate the location of a target object (often smaller and possibly mobile) with respect to another reference object (often larger and static). We can see the locational nature of spatial prepositions when we look possible answers to the following question:

**Question**: Where is John?
**Answers**: (a) At school ; (b) In the house; (c) On the roof; (d) Across the park

In each of these cases, the location of John with respect to a reference object is being communicated. On the other hand, it is hard to imagine a question using "What" or "Why" that could get a plausible answer using a spatial preposition.

If spatial prepositions were purely locational, they could all be generated from one global model, a model that encodes the locations of one object with respect to another. Furthermore, since the location of an object in any coordinate frame is a purely geometric relation, one would expect that their meaning is completely determined by the geometrical relationship between the target and the reference objects. However, geometry alone cannot capture the meaning of prepositions as we saw in figures 5.1 above, which argues for the relevance of dynamic ideas of support. Does this mean that prepositions are not locational?

Despite the counterexamples, there are several reasons why prepositions should still be seen as being primarily locational. Let us look at some of them:

(1) The representation of the geometry of the figure and ground is asymmetric. Prepositions do not care about the shape of the figure – figure objects are conceptualized as points (Talmy 2000). The shape of the ground object is important – prepositions are sensitive to topological features of the ground such as its interior and boundary and to axial features of the ground as well. On the other hand, the motion of the figure relative to the ground does matter while prepositions do not let ground objects move with respect to figure objects at all. This asymmetry makes sense only if we think of prepositions as encoding the relative location of the figure with respect to the ground.

(2) Even simple non-locational relations are not encoded, even though they are as simple as locational relations that *are* encoded. Consider the relations illustrated in figure 5.3 along with a verb that describes the relation. HIT and BOUNCE are not more complex than CONTAINMENT or CONTACT, so why are they not encoded? One reason may be that HIT and BOUNCE are not locational.


**Figure 5.3**

(3) Even when non-locational relations are relevant, they are encoded only when they are consistent with a locational role. For example, while support is present in both figure 5.1 and figure 5.2, it is encoded only in figure 5.1, which plausibly communicates the typical location of a book with respect to a table.

For these reasons, I believe that prepositions are fundamentally locational.

However, the notion of location is more complex than thought before. Both figure and ground objects have degrees of freedom that need to be taken into account in a fully developed model of location. For example, figure objects have a tendency to move – under gravity or on their own, if they are autonomous – and also to change their state, say from containment within the ground object to being outside the ground object.

In terms of the technical vocabulary of derivations, I argue that the core meaning , $C_1$, (in my notation) is locational, but the augmented meanings, $C_2$, $C_3$ and$C_4$ need not be locational. In fact, the core meaning of prepositions has to be augmented to include dynamical representations for both the figure object as well as the ground object. It is not just the purely geometric relations that matter, geometry is a proper subset of dynamics. One way to include dynamics is to include figural motion and change in figure state in the figure models from the category $\Gamma_2$. Similarly, we can enrich the representation of the ground objects in $\Gamma_2$ by including physical characteristics of the ground object. Note that even if the figure is not moving, the models needs to encode the fact that the lack of motion is a dynamically maintained state (for example in figure 5.1 where the book and the table are in a dynamically stable relationship that does not involve motion). The category theory approach allows us to incorporate dynamic principles into the locational models rather easily. Let us see how each one of these steps is worked out, starting with the abstract, core meanings.

**Step 1. Generating the abstract, core meanings of prepositions.**

It seems reasonable to assume that spatial prepositions represent spatial relations relevant to human beings acting in a dynamic spatio-temporal environment. Then it also seems reasonable to assume that the abstract, core, concepts encoded by spatial prepositions are those that belong to the most universally valid, i.e., highly invariant sub-categories of our spatial environment. These common elements (borrowed from chapter 3, section 4) are:

(1) A gravitational environment with gravity pointing up.
(2) An autonomous agent (the perceiver), and possibly other agents.
(3) Solid objects of randomly distributed shapes, sizes and locations that satisfy the following physical laws:
    a. An object is at one place at one time
    b. An object has fixed shape and size and mass.
    c. An object moves in a continuous path.
(4) Most of the objects are motionless during the task, though some may move due to forces exerted by an agent or by some other object.
(5) All of the agents and objects are located in an environment suffused with light obeying the principles of geometric optics.

These constraints were modeled formally in chapter 3 as a model:

$E = GP \wedge T_{aff} \wedge O \wedge v_g \wedge S_i$.

Where, E is the spatial environment, $GP$ is the ground plane, are the $T_{aff}$ are the affine transformations in 3-D space, $O$ is the observer, $v_g$ is the vertical gravitational vector and $S_i$ represent the smooth surfaces of solid objects in 3-D. The partial ordering generated by the above model, $E$, has three broad sub-categories: Coordinate frames, Topological Structure, and Metric Structure. A short outline of these categories is given below:

(1) Coordinate Frame Representations: In order of decreasing invariance, we have three Coordinate frame representations- Gravitational Frame (GF), Blob-location (BL), Principal Axis-location (PA) and Minor Axis-location (MA).

(2) Topological Structure Representations: Blobs (B), Dimension (D), Generic spatial invariants (GI), Non-generic invariants, (NGI), Higher Topological Invariants (HTI).

(3) Metric Structure Representations: Global Properties (GP), The Part Structure Hierarchy (PSH), Convexity Hierarchy, (CH), Riemannian Structure (RS).

Metric Structure, Coordinate Frame and Topological Structure obey the following partial ordering relations:

(a) Metric Structure $\angle$ Coordinate Frame
(b) Metric Structure$\angle$ Topological Structure

Furthermore, as shown earlier, when classified according to invariance, Coordinate Frame, Topological Structure and Metric Structure can be decomposed into the partially ordered categories given below:

(1) Coordinate Frame: MA $\prec$ PA $\prec$ BL $\prec$ GF.
(2) Topological Structure: HTI $\prec$ NGI $\prec$ GI $\prec$ D $\prec$ B.
(3) Metric Structure: RS $\prec$ CH, PSH $\prec$ GP.

The partial ordering of *E* is the collection of all the partially ordered sequences given above. Thus, a classification of spatial categories based on stability leads to a partially ordered set of spatial categories.

The derivational hypothesis along with the inclusion relations in (a)-(b) and (1)-(3) above make the following prediction:

> *The core meaning of prepositions must either be a coordinate frame concept or a topological structure concept but not a metric structure concept.*

We can test this hypothesis rather easily. For simplicity, let us assume that each preposition has a default, geometric meaning. For example, the default meaning of *in* is "containment" which is a generic topological relation. A classification of geometric features associated with the default meaning of prepositions, based on an (informal) corpus analysis of The New York Times is given in the tables 1 and 2 below. While this classification does not claim to be exhaustive, it serves to demonstrate that metric features are not part of the meaning of prepositions. Even such seemingly metric prepositions such as "near" are ordinal and not truly metric. Ordinal relations are topological because they are invariant under all geometric transformations such as dilations, and stretches.

| Preposition | About | Above | Across | After | Against | Along | Alongside | Amid | Among | Around | At |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Meaning | N/A | GF | GI | GI | NGI | MA | MA | GI | GI | NGI | BL |

| Preposition | Atop | Behind | Below | Beneath | Beside | Between | Betwixt | Beyond | By | Down | From |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Meaning | BL | MA | GF | GF | MA | GI | GI | GI | N/A | GF | BL |

| Preposition | Forward | Here | Inward | Left | North | Outward | Right | Sideways | South | There | Together |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Meaning | MA | BL | GI | MA | MA | GI | MA | MA | MA | BL | NGI |

| Preposition | In | Inside | Into | Near | Nearby | Off | On | Onto | Opposite | Out | Outside |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Meaning | GI | GI | GI | GI | GI | GI | NGI | NGI | MA | GI | GI |

| Preposition | Over | Past | Through | Throughout | To | Toward | Under | Underneath | Up | Upon | Via |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Meaning | GF | GI | GI | GI | BL | BL | MA | MA | GF | BL | BL |

| Preposition | With | Within | Without | Afterward | Apart | Away | Back | Backward | Downstairs | Downward | East |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Meaning | N/A | GI | GI | GI | GI | GI | GI | MA | GF | GF | MA |

| Preposition | Upstairs | Upward | West |
|---|---|---|---|
| Central Meaning | GF | GF | MA |

Gravitational Frame, (GF), is the universal vertical frame defined by gravitation. Blob-location (BL), represents a Blob at a particular location. Principal Axis-Location (PA), represents a principal axis of an object at a given location while Minor Axis-Location (MA), represents a Minor Axis at a given location. Blobs (B), are undifferentiated spatial representations, with no internal structure. Generic spatial invariants (GI), are those spatial invariants that do not change when the stimulus is perturbed slightly. Non-generic invariants, (NGI) are those spatial invariants that change when perturbed along certain direction.

**Table 1: Core meaning of Prepositions**

| Partial Order Type | Gravitational Frame | Generic invariant | Blob Location | Principal Axis | OTHER |
|---|---|---|---|---|---|
| Frequency | 10 | 30 | 9 | 16 | 4 |

**Table 2: Prepositions by Spatial Type**

Furthermore, note that prepositions do not name any topological or geometric relations that have depth more than 2 in the partial ordering. For example, we do not have a preposition that encodes the relationship between the rectangle and the black dot given in figure 5.4 below.

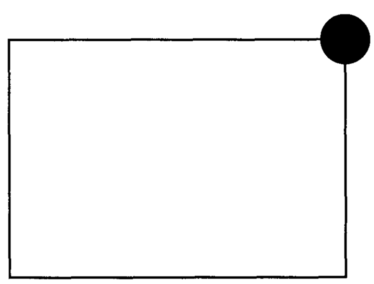


**Figure 5.4**

We can say, "The black dot is on the rectangle" which does not capture the exact geometrical relation or we can say "The black dot as it the upper right corner of the rectangle" which is far more precise, but requires the qualifiers "upper" and "right". There is no single preposition that says "at upper right corner". In other words, the core meaning of a preposition only encodes an abstract spatial relation that has low complexity, either 1 or 2.

### Step 2: Augmenting Abstract Models

In the previous section, we saw that the core meanings of a preposition are abstract locational predicates- either topological or coordinate frame predicates. In this section, I will show how these core meanings can be augmented to incorporate dynamical relations. Let FM stand for figure models and GM stand for ground models. Given the earlier arguments for the locationality of prepositions, the articulation of figure and ground models proceeds in different ways – increasing dynamic capacities for the figure and increasing geometric complexity for the ground.

The simplest model for both figure and ground is one where both figure and ground are static points. Formally

FM(1) : Figure is a static point.
GM(1) Ground is a static point.

The next step is to add complexity to the figure by adding motion or change.

FM(2): Figure motion is modeled as point moving along a curve.
FM(3): Figure change is modeled as a change C: S $\rightarrow$ S' where S and S' are two possible dynamically stable states. The simplest case of change, the one we will use for prepositions is when there are only two states let say **0** and **1**. In this case FM(3) consists of four choices: $\mathbf{0} \rightarrow \mathbf{0}$, $\mathbf{0} \rightarrow \mathbf{1}$, $\mathbf{1} \rightarrow \mathbf{0}$, $\mathbf{1} \rightarrow \mathbf{1}$.

Similarly, one can augment the ground models as follows:

GM(2): Ground topology is articulated. The partial ordering of topological states in the case of a 3D ground object is given below

$Vertex \prec Edge \prec Face \prec Interior, Exterior$

GM(3): Ground axial structure is articulated. The partial ordering of axial structure is given below:

$MinorAxes \prec MajorAxes$

Note that FM(1) is included (in the category theoretic sense as explained in chapter 2) in FM(2) as well as FM(3). Similarly GM(1) is included in GM(2) as well as GM(3). Now that we have these global models for the figure as well as the ground, we can state the basic semantic hypothesis:

**Prepositional Semantic Hypothesis**: Suppose we map the figure models FM(1), FM(2) and FM(3) on to the ground models GM(1), GM(2) and GM(3) using the mapping principle described in chapter 3. Then, prepositional meanings are generated by the partial ordering's that can be derived from each one of these mappings.

To see why the above hypothesis generates the appropriate meanings of prepositions, let us take a look at the partial orders that are generated when the figure and ground models are mapped on to each other and the prepositions that correspond to these partial orders. Remember that according to the scheme developed in chapter 3, the partial ordering comes from classifying the possible topological states of the combined figure-ground models.

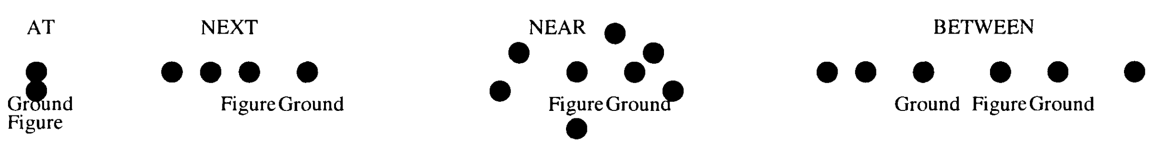



**Figure 5.5(a). FM1 * GM1**

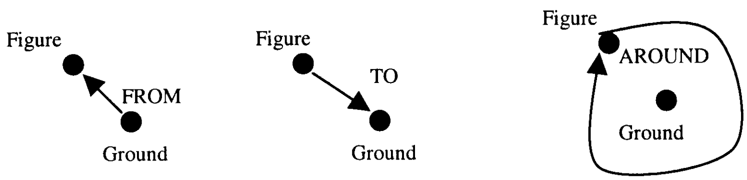



**Figure 5.5(b). FM2 * GM1**

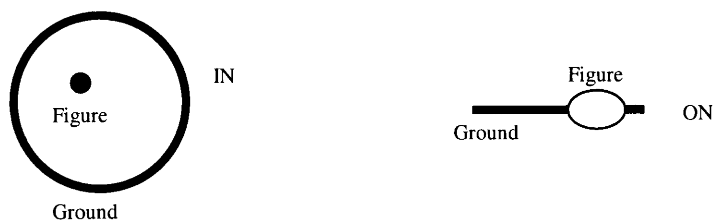



**Figure 5.5(c). FM1 * GM2**

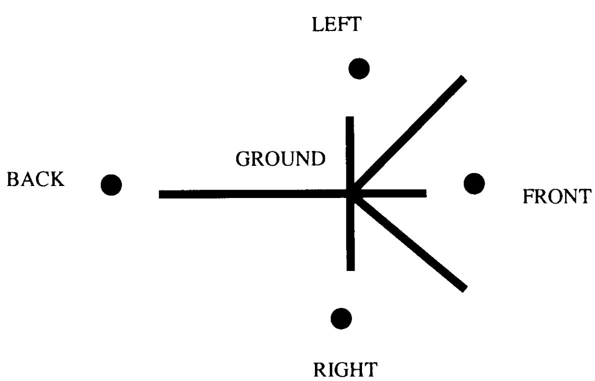



**Figure 5.5(d). FM1 * GM3**

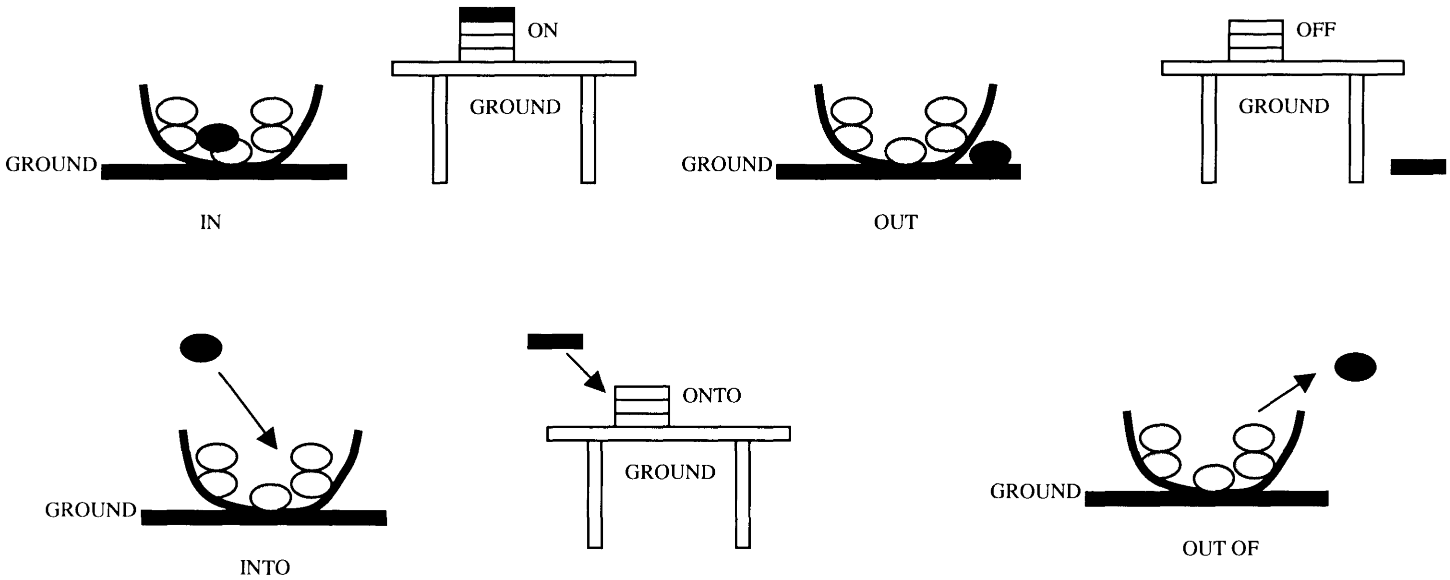



**Figure 5.5(e). FM3 * GM2**

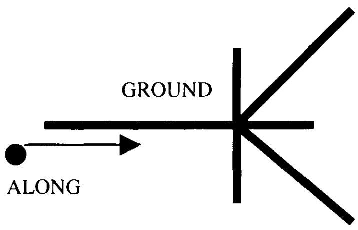



**Figure 5.5(f). FM2 * GM3**

**Step 3. Mapping Abstract Meanings of Prepositions on to the Concrete.**

In the previous section, I showed how the derivational approach can be used to generate the correct meaning of prepositions under various circumstances. I argued that the meaning of prepositions is determined by various models and that the semantics of prepositions is determined by looking at the mappings between these models. In this section, I show that the same approach can be used to understand the reference of prepositions, i.e., what properties of the world are picked out by prepositions.

As I argued in the introduction and chapter 2, the problem of picking out the worldly referent of a preposition is also a problem of mapping different models on to each other. While the meaning of a preposition is abstract, its referent has to be a concrete object in the world. How are these two mapped on to each other. My solution is to realize that both the abstract and the concrete are representations, though these representations make different aspects of the world explicit. Nevertheless, I will continue to assume that the two representational schemes are generated by the same principle of partially ordered categories described in chapter 3. In the previous section, I showed how the abstract meaning of prepositions can be generated by a combination of topological, axial and dynamic global models. In order to understand the mapping of prepositions on to the world, we need to answer two questions:

(1) What is the generative model for the representation of concrete "world" properties?
(2) Is the mapping from the abstract categories to the concrete categories given by the mapping principles developed in chapter 3?

In the next few paragraphs, I show how one can answer these questions using DERIVATIONAL framework. For simplicity, my discussion is restricted to the purely locational, topological meaning of prepositions in a 2D Euclidean world. For locational, topological meanings of prepositions in 2D, I show how one can model the concrete world as a "Metric" category.

To be more formal, let us assume that the world consists of smooth curves and their possible interiors in $R^2$ with the normal, Euclidean metric. One can then model this world using two different categories, the topological and the metric.

A single curve in $R^2$ can only have one of two possible topological states - it can be either open or closed. Similarly, given two smooth curves in 2D, their topological relations are the following

(1) containment/outside,
(2) contact
(3) number of intersections

At the metric level, each curve can be modeled as follows:

A curve is a vector $X(t) = [x(t), \partial x/\partial t]$, where x(t) is a parametrization of the curve and $\partial x/\partial t$ is its tangent vector field.

Then one can immediately define a few properties of the curve using the above parametrization. In order to classify the metric category according to its partial ordering, let us see how we can derive the properties of closure and intersection using the parametrization developed above. For closure, note that

$$\text{The winding number } = \frac{1}{2\pi}\oint \frac{dx}{\|x(t)\|} \text{ is an integer if and only if the curve is closed}$$

If a curve X(t) is closed, we can define its interior as follows:

The interior I(X) of a curve X(t) is the bounded component of $R^2 - X(t)$ (figure 5.6).

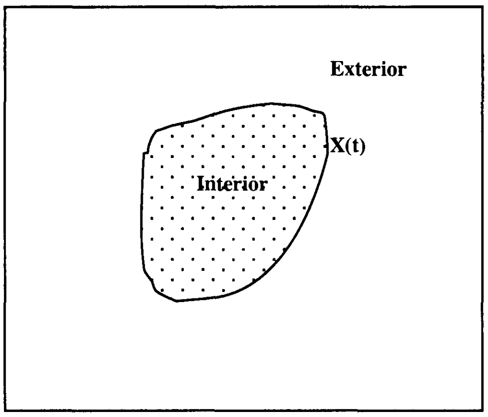



**Figure 5.6**

Similarly, note that intersections can be defined as follows:

> Two curves $X(t) = [x(t), \partial x/\partial t]$, and $Y(t) = [Y(t), \partial y/\partial t]$, intersect at a point $t_0$ if and only if $x(t_0) = y(t_0)$. Furthermore, the intersection is transversal if $\partial x/\partial t|_{t=t0} \neq \partial y/\partial t|_{t=t0}$ and it is non-transversal if $\partial x/\partial t|_{t=t0} = \partial y/\partial t|_{t=t0}$.

In this way, we can see how there is a systematic way of mapping topological categories to metric ones. Note that the mapping is an isomorphism of partial orders as well – the mapping is *codimension preserving*. For example, in the topological case, for curves in $R^2$, non-transversal intersections are codimension 1 intersections. Similarly, when two curves in $R^2$ intersect, generically, the two tangent vector fields are not equal. However, the set of possibilities when they *are equal* is of codimension 1. In other words, the notion of non-transversality in the metric case has the same codimension as non-transversality in the topological case. Therefore, the mapping from topological to metric categories (shown above for the case of $R^2$, but true more generally) is not just a mapping of models, it is also a mapping of partial orders.

**Step 4. Augmenting Concrete Models**

The final piece in trying to understand the semantics of prepositions is to understand how concrete models can be augmented. One may ask why this is necessary, given that the semantics of prepositions is mainly abstract. The reason is because concrete situations do make a difference to the meaning of a preposition. Consider the two pairs of pictures in figure 5.7 below

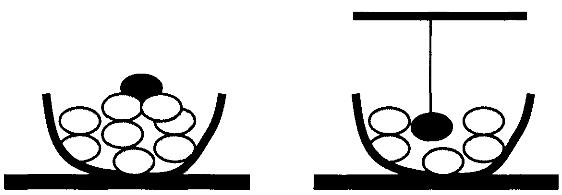
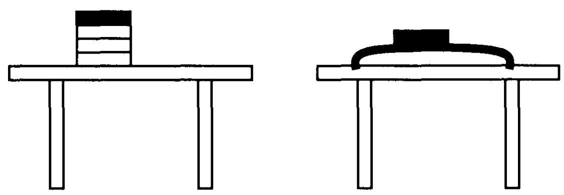


**Figure 5.7 (a). Is the black egg in the bowl?**



**Exterior, $E_1$**

**Figure 5.7 (b). Is the black book on the table?**

In both figure 5.7(a) and 5.7(b), the figure and the ground have the same abstract relationship, of containment in 5.7(a) and of support in 5.7(b). However they are judged differently. The difference between the two cases cannot be attributed to the difference in some abstractable feature. For example, one may think that support is the crucial feature that is missing in 5.7(a) but it is present in 5.7(b) and yet the two cases are judged differently. It is quite plausible that the semantic distinction between the two situations is because of the metric and physical conditions that differentiate the two circumstances. In other words, the meaning of a prepositions can reflect the concrete relations between the figure, the ground and other objects in the scene.

I claim that we can understand the difference between the two situations by looking at the way in which concrete spatial models can be augmented. As before, we start with the hypothesis that spatial prepositions are primarily locational. However, as we saw in the case of abstract models, the notion of location (in the abstract domain) has to be augmented to include motion and change on part of the figure, leading to a richer vocabulary for describing the semantics of prepositions. Therefore, it stands to reason that the notion of location can be expanded in the concrete domain as well, and that this expansion should lead to a richer vocabulary for describing prepositions in the concrete.

Some augmentations – such as incorporating the notions of change and motion into an expanded notion of location- are possible within the space of abstract models. The key question for augmenting concrete models is

> "Are there systematic ways of augmenting concrete models that are not reducible to instances of abstract augmentation mapped on to the concrete sphere?"

The answer to this question is "Yes". While the concrete world is too rich to permit a complete classification, there are two important ways, via the notions of are *neighborhood* and *causal locus*, in which concrete representations can be. Both of these notions are based on concrete metric relations between objects in the world, which are not reflected in the topological relations encoded in the abstract domain. Furthermore, both of these notions are ways of augmenting the notion of location in the concrete domain.

By the "neighborhood" of an object, I mean the set of objects that are "near" the object in some well defined metric space. The metric can vary, and depending on the metric used, we can get different "neighborhoods" for the same object. Nevertheless, the presence of some metric is crucial, because the notion of neighborhood is dependant on the existence of continuously varying parameters within a metric space. Some important choices for a metric are given below: Throughout this exposition, I will use the symbol "$\approx$" to denote nearness in some metric space.

(a) The Contact metric. Let A,B,C etc be a collection of objects in three dimensions. Then two objects A and B are similar, symbolically, $A \approx B$ if and only if A touches B. Given an object, A, the neighborhood of A, denoted N(A) is defined as :
$N_{contact}(A) = (B: \exists$ a sequence $A_1, A_2, \ldots A_n$ such that $B \approx A_n \approx A_{n-1} \approx \ldots A_1 \approx A)$
(b) The size metric. Let A,B,C etc be a collection of objects in three dimensions. Let the diameter Diam(A) of an object, A, be the largest distance between two points on that object. Then the neighborhood, $N_{size}(A) = (B: |Diam(B) - Diam(A)| \leq \varepsilon$ for some value of $\varepsilon$.
(c) The velocity metric. Let A,B,C etc be a collection of objects in three dimensions each moving with some velocity $V_A$, $V_B$, $V_C$ ... respectively. Let • be the scalar product between three dimensional vectors. Then the neighborhood $N_{velocity}(A) = (B: |V_B \bullet V_A - V_B \bullet V_B| \leq \varepsilon$ for some value of $\varepsilon$ (Intuitively, those objects whose velocity is close to the velocity of A).

By *causal locus*, I mean the set of objects that an object can effect physically. Going back to chapter 3, remember that causation was defined in two forms, one positive and the other negative, as follows:

(1) Positive Cause. An object causes the state S(B) of another object B, if $A \Rightarrow (S'(B) \rightarrow S(B))$, i.e., the presence of A leads to change of state of B from S' to S.
(2) Negative Cause. An object A causes the state S(B) of another object B, if $(S(B) \rightarrow S(B)) \Rightarrow A$ or counterfactually, $\neg A \Rightarrow (S(B) \rightarrow S'(B))$, i.e., the presence of A is essential to the maintenance of S(B).

In both cases, one can define the causal locus of A, CL(A), as follows:

**Definition**. The causal locus of A, CL(A) = (B: $\exists$ a sequence $A_1, A_2, \ldots A_n$ such that A causes $A_1$ causes $A_2$ causes … $A_{n-1}$ causes B)

The causal locus is a concrete set of objects because it depends on the actual set of objects in the world that an object can affect. Note that both the neighborhood of an object as well as its causal locus offer ways of augmenting concrete global models by replacing the model of an object by its neighborhood (in some metric) or its causal locus. Now let us see how we can apply these two means of augmentation to the semantics of prepositions.

As before, since prepositions are locational, we should expect that figure and ground models are augmented differently. The fundamentally, the idea is to include as many details about the location of the figure and as many details about the possible locations that the ground can generate. In the abstract case, the figure model was augmented to include dynamic concepts of figure location and the ground model was augmented to include subtler geometric aspects of its shape.

In the concrete case, in the figure model, we can augment the notion of figure location by including not only the location of the figure itself with respect to the ground, but also the location of the neighborhood of the figure with respect to the ground. Similarly, we can augment the ground model by including (in a coordinate frame with respect to the ground) not only those locations generated by the geometry of the ground alone, but also those locations that are part of the causal locus of the ground. While neighborhoods and causal loci are unique to the concrete domain, augmentation using neighborhoods and causal loci is *consistent* with the augmentation procedure in the abstract domain. This is because change and motion in figure objects is often the result of causal influences originating in the ground object and these influences are likely to be transmitted to the figure object via other objects that are part of the figure object's neighborhood (as happens in figures 5.7(a) and 5.7(b). Therefore, if needed, one can change the meaning postulates for spatial prepositions as follows:

> **Prepositional Meaning Postulate**: Let P(figure, ground) be a spatial preposition. Let A be its abstract meaning, i.e., A is a node in a partial order generated by an abstract global model $G_{abstract}$. Let $G_{concrete}$ be the concrete global model that $G_{abstract}$ maps on to and let A' = A'(figure, ground) be the concrete node in $G_{concrete}$ corresponding to A (in $G_{abstract}$). Then a concrete feature, $f_{concrete} \in$ Meaning(P) if and only if $f_{concrete} \in$ A'(neighborhood(figure), causal-locus(ground))).

Now, we can see how the prepositional meaning postulate accounts for the judgements in figures 5.7(a) and (b). In 5.7(a) while the left hand picture satisfies CONTAIN($N_{contact}$ (black egg), causal-locus(bowl)), the right hand picture does not. Similarly, in 5.7(b) the left hand picture satisfies CONTACT( $N_{size}$(black book), causal-locus(table)) while in the right hand picture it does not.

This concludes my presentation of the semantics of spatial prepositions. As I have shown here, the meaning of prepositions is rather intricate – it requires the understanding of abstract models, concrete models and the mapping between the two, as well means of augmenting abstract models and concrete models. Once all of these stages of modeling are available, we can generate a rich set of prepositional meanings –one that is adequate to capture meanings that have been hard, if not impossible, to generate systematically using other techniques. Nevertheless, the derivational approach offers a systematic method to understand and generate all of these stages in the semantics of prepositions suggesting that something category like must be quite central to semantics of spatial concepts in language.

## 4.4 A full derivation for ACROSS

So far in this chapter, we have seen how the semantics of prepositions can be modeled using the idea of a derivation. Now, I show how the notion of derivation can be used to understand the semantics of individual prepositions in greater detail. I will restrict my analysis to the semantics of the preposition ACROSS. Let us see how we can generate the meaning of ACROSS.

### Step 1. The abstract, core meaning of ACROSS.

First, I assume that the world is two dimensional and Euclidean. Let us also assume that the Figure objects are modeled as points moving along smooth curves and that the Ground objects are modeled as rigid, static regions that partition two dimensional space. More formally,

(1) SPACE = $R^2$.
(2) PATH(Figure) = Smooth curve in $R^2$.
(3) The compliment of the Ground, i.e., $R^2$—Ground is disconnected. For simplicity, we can assume that the compliment of the Ground has three components, one Interior Component, I, and two Exterior Components, $E_1$ and $E_2$ (figure 5.8).

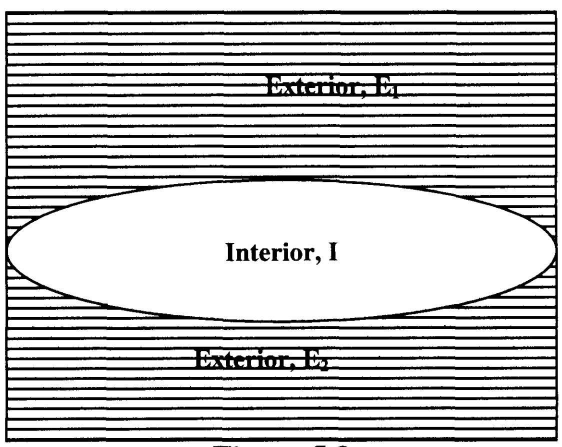



**Figure 5.8**

The product model for the figure and ground define the "world" that starts our derivation. Let this "world" be called **W**. Then, we can classify the topological states of **W** as follows:

> **Definition**: Let I, $E_1$ and $E_2$ be the connected components of $R^2$—Ground as given above. Then, let the Crossing Number, C(Fig, Gr) = Number of times PATH(Figure) goes from $E_1$ to $C_2$ or from $E_2$ to $E_1$.

Then, the set of distinct topological states of **W** is classified by C(Fig,Gr), i.e., each topological state of **W** has a unique crossing number. Therefore, the abstract meaning of ACROSS can be defined as follows:

> **Core Meaning Postulate**. ACROSS(Figure, Ground) ≡ C(Fig,Gr) =1.

Of course, there are an infinite number of possible configurations that are compatible with this meaning, some of which are given in path 1-3 in figure 5.9(a) below. At the same time, the above meaning postulate rules out an infinite number of paths as well, as illustrated in figure 5.6(b).

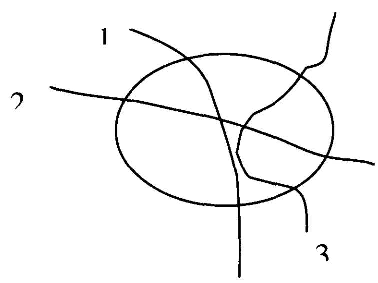



**Figure 5.9 (a)**

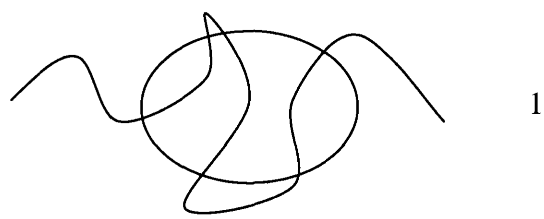



**Figure 5.9 (b)**

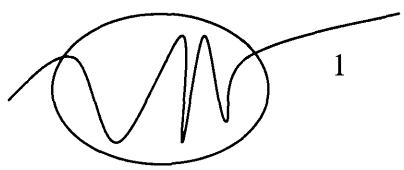



**Figure 5.9 (c)**

Therefore, the abstract meaning for ACROSS does not quite capture all possible nuances of its meaning. For example, the configuration in figure 5.9(c) above is not a case where we would normally license the use of ACROSS. In order to understand the semantics of ACROSS comprehensively, we have to augment the abstract meaning, which at this stage is purely topological.

**Step 2. Augmenting the abstract meaning.**

Remember that there are two highly invariant forms of representations – topology and coordinate frame representations. Therefore, coordinate frame representations can be used to augment topological representations. Let us see how we can do this in the case of ACROSS.

First, note that in the world model, **W**, illustrated in figure 5.8, there is a natural subdivision of the regions of the model. First, there is the bounded component, I, where, independent of the size of the region enclosed by the ground, we know it is finite. Then, there are the unbounded components $E_1$ and $E_2$ which extend to infinity (in the model). Therefore, one can define two natural coordinate frames:

(a) $\mathbf{CF_{int}}$: This is the coordinate frame based on the interior of the ground, which is bounded by the boundary of the ground.
(b) $\mathbf{CF_{ext}}$: This is the coordinate frame based on the exterior of the ground, which is unbounded.

Therefore, we can define augmentations of **W**,namely,

(a) $\mathbf{W^* \, CF_{int}}$.
(b) $\mathbf{W^* \, CF_{ext}}$.

Which in turn leads to augmented meanings of ACROSS, given as follows:

(c) $ACROSS_{int}$ (Figure, Ground) $\equiv$ (C(Fig,Gr) =1) $\wedge$ (PATH(Figure) $\subseteq$ $\mathbf{CF_{int}}$)
(d) $ACROSS_{ext}$ (Figure, Ground) $\equiv$ (C(Fig,Gr) =1) $\wedge$ (PATH(Figure) $\subseteq$ $\mathbf{CF_{ext}}$)

Already, note that the derivation, while based on the same principles, leads to two different "fleshed out" versions of ACROSS. As we start adding more concrete details, such as the geometry of the path of the figure and the shape of the ground, the meaning of a preposition becomes more interesting and far more dependent on its context. I argue that the abstract meaning of a preposition is fleshed out in different ways depending upon the context of the preposition.

**Step 3. Mapping abstract meaning to concrete representations.**

Let us see if we can understand the fleshing out procedure in the case of ACROSS, which we have modeled in terms of connected regions in two-dimensional space. In step 2, we saw that there are two natural ways to flesh out the meaning of ACROSS, depending on the use of an external versus internal coordinate frame. In order map the abstract meanings, $ACROSS_{int}$ and $ACROSS_{ext}$ on to concrete representations, we need to introduce some metric information in the form of a "scale" constraint.

> **Scale Constraint**. Let size(G) be the diameter of the ground. Let size(W) be the diameter of the world, i.e., the extent of the two dimensional space containing both the figure as well as the ground. Then, assume that size(W) >> size(G).

Given the scale constraint, there are two natural kinds of paths. Let size(PATH(figure)) be the length of the path of the figure. Then, we can define two classes of paths:

(a) size(PATH(figure)) >> size(G)

(b) size(PATH(figure)) $\approx$ size(G)

Now, let us see how these considerations lead to two different "fleshed out" meanings of ACROSS, the first of which is a coarse scale meaning, and the other meaning a fine scale meaning. To take a look at the large scale meaning first, note that if (a) above is valid, at a coarse scale the shape of the ground object and the detailed relationship between the figure and the ground can be ignored because their geometry does not show up at a coarse scale. The only thing that matters is the geometry of the path itself, since that is the only geometrical detail that is visible at that scale.

> **Definition.** Let C be a curve in 2D and let $\lambda$ be a scale parameter where. Let $C_\lambda$ be a smoothed ot version of C, let us say, $C_\lambda$ is C convoluted with a Gaussian, $G_\lambda$,of width $\lambda$ (more formally, $C_\lambda = C^* G_\lambda$). Then define the net curvature, $N_\lambda$,(C) of C at scale $\lambda$ as follows:
>
> $$N_\lambda(C) = \frac{1}{2\pi}\oint \|C_\lambda\|$$

Then, we can define the coarse scale meaning of ACROSS as follows:

> **Coarse Scale Meaning.** Let the coarse scale, $\lambda$, satisfy $\lambda >>$ size(Ground). Let $\varepsilon$ be a real number greater than 0.Then, for some $\varepsilon$,
> ACROSS(Figure, Ground) $\equiv$ (C(Fig,Gr) =1)$\wedge$( $N_\lambda$,(PATH(Figure)) $\leq \varepsilon$))

The above fleshing out takes care of the coarse scale meaning. What about the fine scale meaning? The fine scale meaning is a little more complex, since there are more geometric factors that need to be taken into account. These geometric factors are:

(1) The shape of the ground object.
(2) The curvature and the velocity of the path of the figure object.

How can we incorporate these geometric factors into the meaning of ACROSS? Let us first note that in determining the meaning of ACROSS, we need to take into account two factors:

(1) Whether the figure object crossed the ground object or not. In other words, we have to map the (topological) interior and exterior components of the world with respect to the ground, I, $E_1$ and $E_2$.on to some metric features of the ground object. In other words we have to determine the *sides* of the ground object. Then, we have to determine whether the path of the figure object went from one side to another or not. This determination is primary, for if the figure object did not go from one side to another, it does not matter how it moved.
(2) If we have determined that the figure has indeed gone from one side to another, then we can incorporate the manner in which it did so. In this case, we have to see how the path of the figure is located with respect to the ground as well as the curvature of the path.

Steps 1 and 2 above are accomplished as follows:

(1) The sides of the ground object are determined by looking at the lattice of features of the boundary of the ground object. That is to say, they are determined by the curvature extrema of the boundary. Positive extrema (+ in figure 5.10) demarcate side boundaries while negative extrema determine part boundaries (- in figure 5.10).
(2) Once the sides have been established and the validity of ACROSS has been determined, the *centroid* construction determines the relationship of the path of the figure to the ground. The centroid construction is given as follows: For any point, x, at the boundary of the ground, there is a unique point, y, such that a straight line from x to y divides the area of the ground equally. The path of the figure should be close to the centroid path. Finally, we should also expect the path of the figure to be not too curvaceous. The terms 'close to' and 'curvaceous' can e made precise using the notion of scale (see below).

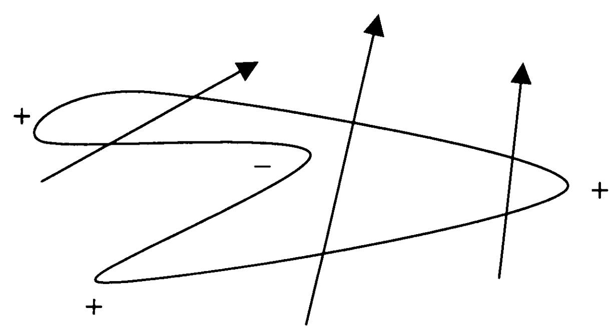



**Figure 5.10**

Combining these analyses, we should get the following model for the meaning of ACROSS in the fine scale case.

> **Fine Scale Meaning.** Let the fine scale, $\lambda$, satisfy $\lambda$ << size(Ground). Let $\varepsilon$ be a real number greater than 0.Then, for some $\varepsilon$,
> ACROSS(Figure, Ground) $\equiv$ (C(Fig,Gr) =1)$\wedge$( $N_\lambda$,(PATH(Figure)) $\leq \varepsilon$) $\wedge$ CONNECT(PATH(Figure), SIDE-A(Ground), SIDE-B(Ground)), where the edges of the sides are curvature extrema of the boundary of the ground.

The fleshing out of the "concrete" meaning of ACROSS also leads to specific empirical predictions. First, note that the concrete model is a combination of non-accidental metric features of the shape. Let us see what this means for two shapes – ellipses and squares.

(c) For an ellipse, the non-accidental features are the curvature maxima, therefore the non-accidental partitioning of the ellipse into sides should be given in terms of the major and minor axes of the ellipse. The major and minor axes determine the SIDE parameters in the fine scale meaning in the case of ellipses. Furthermore, the ratio, $\frac{Major}{Minor}$, of the major and minor axes should determine the acceptability of ACROSS.

(d) For a square, the non-accidental features are the edges and the equality of sides. Therefore, the SIDE parameters in the fine scale meaning should be determined by the edges. Also, since the sides are of equal length, the choice of sides should make no difference in the acceptability of ACROSS.

These hypotheses were tested in two experiments described below.

**Experiment 1. ACROSS an ellipse.**

**Stimuli**: Subjects sat in front of a computer screen. The computer screen contained an ellipse of axial ratio $\frac{Major}{Minor}$ =2 and of various sizes and at randomly chosen orientations.

**Task**: The subjects were presented with the sentence

> (A) John walked across the park.

After which they saw a series of paths connecting a randomly chosen point outside the ellipse to another randomly chosen point outside the ellipse like the dotted lines in figure 5.11 below. The total collection of paths for each subject covered the complete range of ellipse sizes as well as a 360 degree spectrum around each ellipse. The subjects were asked to rate the acceptability of the sentence (A) above on a scale of 1-5.

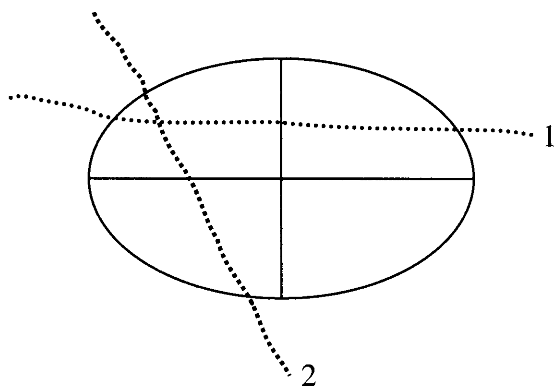



**Figure 5.11**

**Results**: The normalized results for 4 subjects are given in figure 5.12 below. Normalization was done by equalizing the average score of all the subjects.

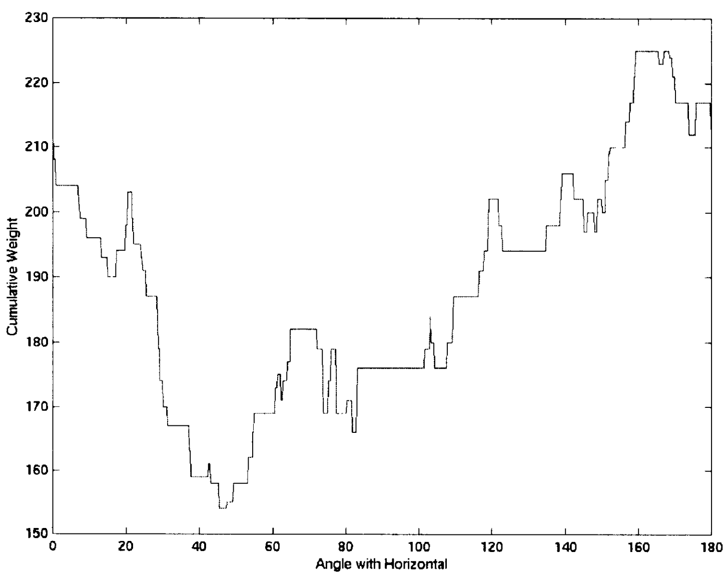



**Figure 5.12**

**Discussion**: The histogram in figure 5.12 shows how the subjects performed exactly as would be predicted by meaning hypothesis A above. The maximal acceptability of ACROSS is around the major axis of the ellipse (angle with respect to horizontal = 0,180 degrees) and then there is a monotonic decrease in acceptability that is well correlated with the minor axis (angle with respect to the horizontal = 90 degrees) This is in agreement with the larger picture, namely, that the semantics of ACROSS and other prepositions should take into account both abstract and concrete models. Furthermore, it also shows that not all features are represented in the concrete model, only the non-accidental features.

**Experiment 2. ACROSS a square.**

**Stimuli**: Subjects sat in front of a computer screen. The computer screen contained an square of various sizes at a randomly chosen orientation.

**Task**: The subjects were presented with the sentence

(B) John walked across the park.

After which they saw a series of paths connecting a randomly chosen point outside the square to another randomly chosen point outside the square like the dotted lines in figure 5.13 below. The total collection of paths for each subject covered the complete range of square sizes as well as a

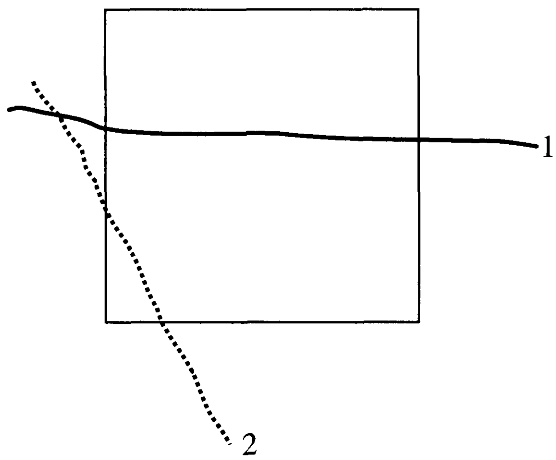



**Figure 5.13**

360 degree spectrum around each square. The subjects were asked to rate the acceptability of the sentence (A) above on a scale of 1-5.

**Results**: The normalized results for 4 subjects are given in figure 5.14 below. Normalization was done by equalizing the average score of all the subjects.

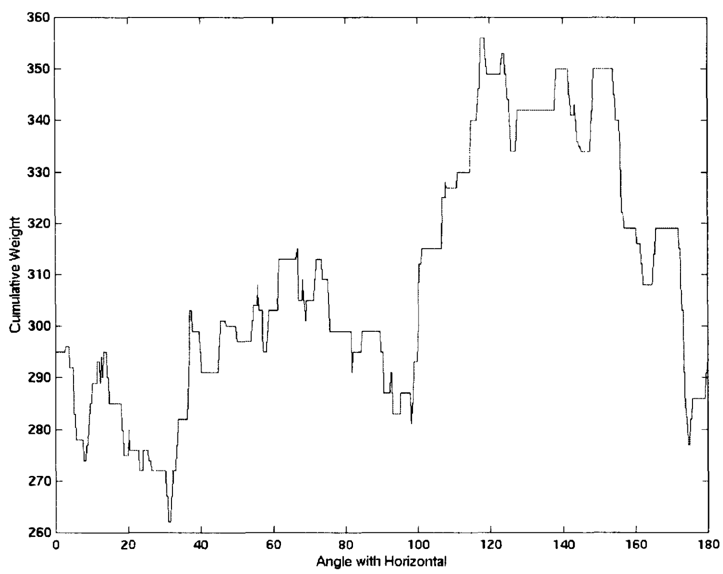



**Figure 5.14**

**Discussion**: The histogram in figure 5.14 shows that the results are partially predicted by the meaning hypothesis. The maximal acceptability of ACROSS is around the vertical axis of the square (angle with respect to horizontal = 0,180 degrees). There is another peak in the acceptability around the horizontal (90 degrees), though not quite as large. These two peaks are consistent with the SIDE parameters being determined by the edges of the square. However, the peaks are not equal. One reason for the inequality of the two peaks is the fact that from the subjects perspective, the vertical and the horizontal axes are not symmetric, therefore, crossing the vertical axis may have greater relevance than crossing the horizontal axis perhaps due to the association of the vertical axis with gravity. In other words, the subject may be imposing a coordinate frame that is not incorporated by the model at this level. However, note that this additional coordinate frame can easily be incorporated in step 4, namely, augmenting the concrete model, which I discuss next.

### Step 4. Augmenting the concrete model.

To complete the derivation of the meaning of ACROSS, we need to see how the concrete meaning can be augmented. The concrete models can be augmented in two ways:

(1) Imposing a metric coordinate frame on the representation of the geometry. Coordinate frames can either be observer or object centered. The non-accidental features represented in a coordinate frame would be the axes (for an object centered frame) and verticality, horizontalness and alignment with the body of the observer, for an observer centered coordinate frame. As I mentioned earlier in the discussion of the experimental results for squares, an observer centered frame may be implicitly involved in judging the acceptability of ACROSS even though the default coordinate frame is an object centered one.

(2) Augmenting geometry to include physical structure. For ACROSS, this can be done by using the idea of a *causal path locus*. By causal path locus, I mean the following:

**Definition**: Let G be the ground object, and let $O_1, O_2 \ldots O_n$ be objects located in the interior Int(G) of G. Let $\varphi$ be a parameter, where $\varphi \ll$ diameter(Ground). Let $N_\varphi(O_i)$ be a circle of radius $\varphi$ around the object $O_i$. Let R(G) be the complement of the circles in the interior of the ground.

$$R(G) = G - \bigcup_i N_\varphi(O_i)$$

Let F be the figure object. Then the causal path locus, CPL(F, G), of F with respect to the ground, G, is defined as

CPL(F, G) $\equiv$ The set of all paths, $P_\alpha(F)$ of the figure such that $P_\alpha(F) \subseteq R(G)$.

Then, we can define the augmented concrete meaning of ACROSS as follows:

**Augmented Concrete Meaning.** Let the fine scale, $\lambda$, satisfy $\lambda \ll$ size(Ground). Let $\varepsilon$ be a real number greater than 0.Then, for some $\varepsilon$,
ACROSS(Figure, Ground) $\equiv$ (C(Fig,Gr) =1)$\wedge$( $N_\lambda$,(PATH(Figure)) $\leq \varepsilon$) $\wedge$ CONNECT(PATH(Figure), SIDE-A(Ground), SIDE-B(Ground)) $\wedge$ (PATH(Figure) $\in$ CPL(F, G))

The augmented concrete meaning of ACROSS allows us to explain why the path in figure 5.15 is acceptable.

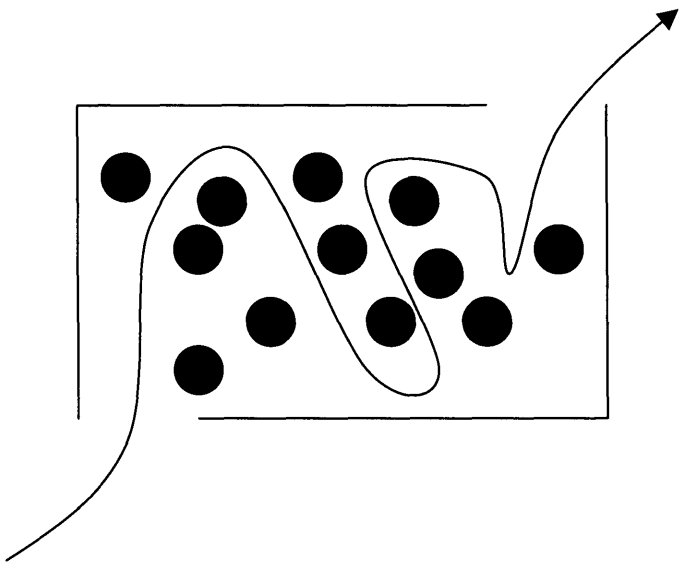


**Figure 5.15: John walked across the park**

This ends my presentation of the semantics of ACROSS. I have shown how the derivational account of the meaning of prepositions leads to a detailed analysis of the meaning of ACROSS, accounting for all shades of meaning, from the core, abstract meaning based on the world model illustrated in figure 5.8, to the concrete, augmented meaning illustrated in figure 5.15.

### 4.5 Summary

In this chapter, I argued that the meaning of prepositions can be understood in great detail using the derivational theory of meaning, involving four steps – the abstract core meaning, the augmented abstract meaning, the concrete meaning and the augmented concrete meaning. I showed how the abstract meaning of prepositions can be seen as the most invariant topological and coordinate frame features in a three-dimensional world model. I then showed how one can augment these features by adding dynamical elements which in turn get mapped on to concrete, metric features of the world. Finally I showed how these metric representations can be augmented to reflect the geometric notion of neighborhood and the physical notion of cause.

These four steps were illustrated in the case of one preposition, ACROSS. Here, I showed how the derivational account explains a wide range of semantic phenomena for ACROSS, from the abstract meaning to the involvement of the details of figure and ground geometry.

I believe that the derivational theory of prepositional semantics shows how the meaning of prepositions can be derived systematically from category theoretic principles. Furthermore, by showing how prepositions start from a core, context-insensitive, abstract meaning but "flesh out" to include rich, geometrical and physical details, the derivational theory of prepositional semantics shows how prepositions manage to be highly versatile, context-sensitive concepts. The lessons learned about prepositions can be applied to concepts in general – because concepts have an abstract core, they can be combined easily and since they can be fleshed out to include rich details of world configurations, they say something substantive about the world. This combination of abstraction and concreteness, both within the unified framework of "derivations" may offer some important clues as to why human cognition is so versatile.

## Chapter 5: Tangrams: A Test Bed for Spatial Cognition

### 5.1 Introduction.

In the previous chapter, I argued that linguistic expression of concepts, such as prepositions, can be modeled using the idea of *derivations*. The derivational method offers a principled way in which various levels of a concept, the abstract as well as the concrete, can be mapped on to each other. The treatment of prepositions in the previous chapter was focused on "static" regularities, i.e., the regularities that determine the structure of the various levels – topological, metric etc. In this chapter, I want to extend the derivational approach to cover dynamical regularities as well. I develop my ideas about dynamical regularities within the toy domain of 'tangrams', ancient Chinese jigsaw puzzles. Before I go on to delineate the details of puzzle solving, let me summarize the positions developed in the previous chapters.

### 5.2 Orientations

The fundamental assumptions behind the modeling of spatial concepts are as follows:

(a) Concepts cannot be divorced from context, where each context is a *global model* of the world. Each global model is a category (in the precise sense of chapter 2). Global models generate a set of concepts by means of the natural partial ordering defined on categories.
(b) Concepts are represented in multiple contexts, some abstract, others concrete. Within a given context, a concept is a node in a partial order. The representation of a concept, C, in a context, W, is its *projection* $C_W$.
(c) Across contexts, the various projections of a concept can be mapped on to one another and then combined in a *derivation*.

Roughly speaking, dynamical regularities come from the interaction of static regularities over time. Not surprisingly, these interactions can quickly become rather complex. Let us try to build a link from the semantics of prepositions and jigsaw puzzles. We start with the situation illustrated in figure 5.1 (a)-(c) below.

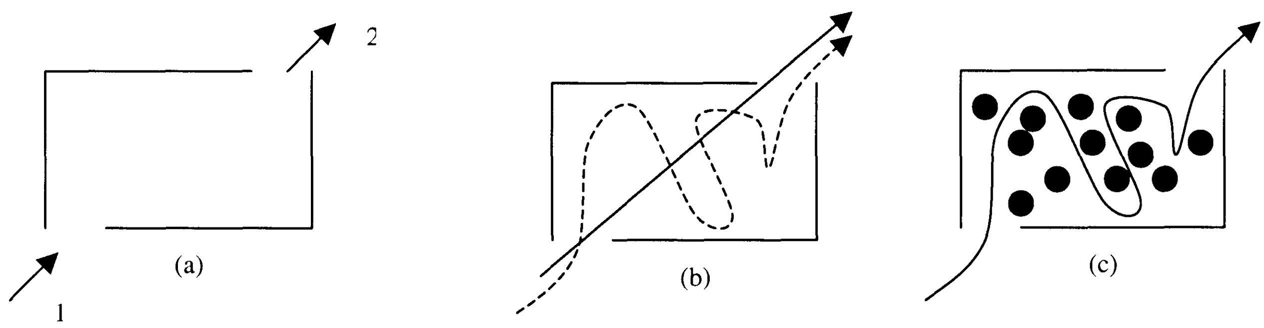



**Figure 5.1: Assume that a person is standing at the location marked 1 and he needs to walk across a rectangular "Park" to the location marked 2. Draw a curve illustrating a representative path taken by the person**

The task of crossing the park in figure 5.1 can be performed in an infinite number of ways. However, some ways are better than others. For example, in the absence of any obstacles, as illustrated in 5.1 (b), the straight path is definitely more typical than the twisted path (the dashed line). However, if we add some obstacles, as in 5.1 (c), the same twisted path seems natural. How can we capture the regularities inherent in the dynamics of the paths?

First, note that in the above example, there are two interacting structures:

(a) Geometry: The formal, geometric structure of the task, which mandates a path while constraining the beginning and the end of the path.
(b) Dynamics: The context - in this case physical – that interacts with the formal structure of the path to produce new, dynamical regularities, i.e, regularities in the way the path progresses in time.

Depending on the context, we can imagine that there are different dynamical regularities. How can we study these regularities? Is there a way to formalize these regularities within the derivational approach?

## 5.3 Complex Spatial Tasks

The main problem in modeling the relationship between prepositional concepts in language and the perceptual world was in understanding how abstract conceptual features

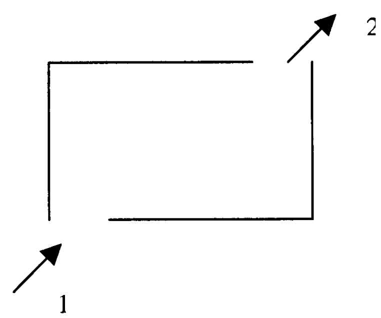
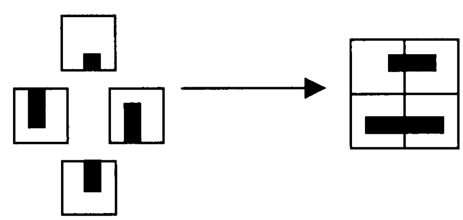
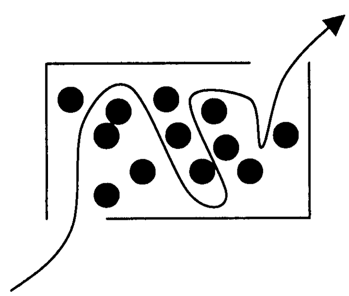
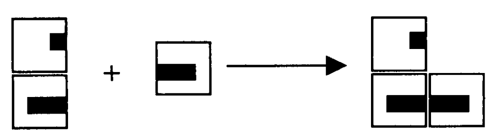
| | Across the Park | Jigsaw Puzzle |
|---|---|---|
| Geometry | Initial State: One side of the Park<br>Final State: Other side of the Park<br> | Initial State: Unarranged pieces<br>Final State: Complete Puzzle<br> |
| Dynamics | Dynamical Regularities: Path constrained by obstacles and shape of the interior of the park.<br> | Dynamical Regularities: Path constrained by edge relations between puzzle pieces as well as the shapes of the figures in the puzzle<br> |

**Table 5.1: Mapping from "Across the park" to Puzzle Solving**

can map on to concrete perceptual features. In the case of prepositions, the mapping problem is relatively easy, since there is only one map in the transition from abstract to concrete. However, many spatial task are far more complex, requiring several steps before the "abstract" and the "concrete" states are mapped, typically in the form of a goal state. For example, think of navigating a city neighborhood, searching for your friend's house. Abstractly, you might know that your friends lives somewhere within a square block. Also you might know something concrete, like his address. These two attributes come together at a particular location. In other words, the problem of navigating to your friends house can be understood as mapping the abstract knowledge of the general area in which the house is located on to a concrete datum namely, the address of the house. However, in order to do the mapping, you may have to go through several steps, make many turns, perhaps get lost before you get to the house. Also, you might need an aid, such as a map or written directions, that allows you to transition from the current location you are in to the final location you want to go to.

Another example of a complex spatial task is that of solving a jigsaw puzzle, such as a tangram. A tangram is a jigsaw puzzle consisting of seven pieces, five triangles, a square and a parallelogram (figure 5.2). The typical task is to combine these pieces to form a square (figure 5.2). I have deliberately left space between the pieces in the depiction of the square so that the non-accidental relations between the pieces becomes more explicit. Jigsaw Puzzles are an interesting domain for studying spatial cognition for they can be modeled explicitly. The relationship between the analysis of prepositions in the previous chapter and jigsaw puzzles is highlighted in Table 5.1, showing the mapping between the domain of jigsaw puzzles and the problem of drawing paths "across a park".

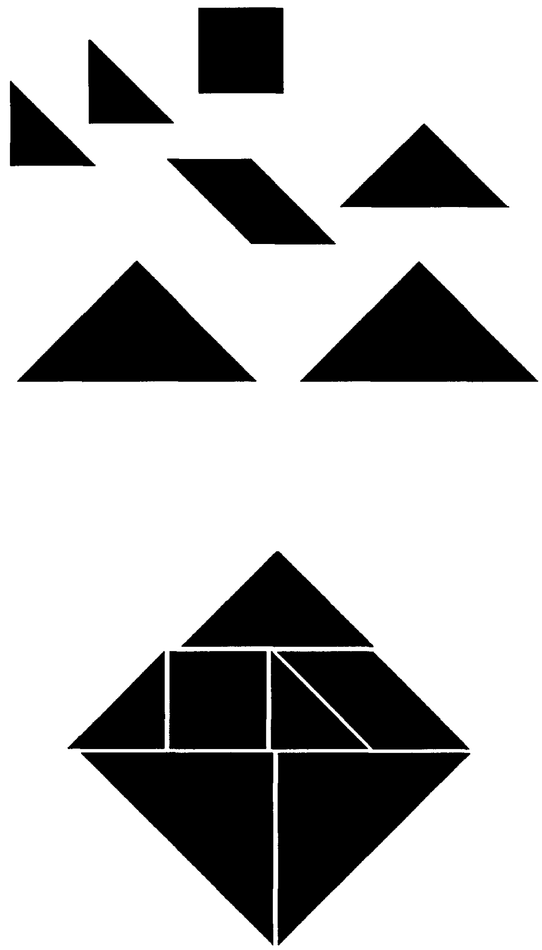


**Figure 5.2: Tangrams**

Note that in solving all jigsaw puzzles, including the tangram example, one needs to balance the demands of the abstract goal, namely, that the final shape is a square, and the concrete data that is available to you, which is the current arrangement of pieces.

In the rest of this chapter, I will explore how tangram puzzles are a useful toy domain in which one can understand the dynamical problem of mapping abstract states to concrete states, where the mapping spans several steps. The reason for using a toy domain is because the complexities can be formalized explicitly and then analyzed for deeper principles.

Furthermore, most of us have some experience in trying to solve a jigsaw puzzle, whether it be large wood-cut animal shapes children play with or the more complex 1,000 piece cut-ups of great works of art. At the same time, in the jigsaw puzzle domain, we can make the dynamical regularities transparent and formalize them explicitly. In other words, the jigsaw puzzle domain has all the features of the unconstrained domain of drawing paths across a park, while permitting formalization in a computational theory. My goal is not to invent a new way to solve tangrams, or even to model human or computer strategies for solving puzzles (Nagpal, 2001). Instead, I want to show, using tangram puzzles, the complexity underlying a domain that we can agree upon as having uncontroversial data. Therefore, the goal of developing a toy model is to achieve the following aims:

(a) To show the derivational structure of a complex domain in which several spatial concepts interact with each other.
(b) To illustrate how the interactions between spatial concepts lead to emergent properties. These properties lead to computational problems that cannot be addressed by using the concepts developed in the original mapped context.
(c) To show how one can solve the problems raised by the emergent phenomena can be addressed within the derivational approach by adding new models to the models that define the original spatial domain. In other words, to argue that global models are situated within a meta-context of other global models, some of which are necessary to address problems raised *within* a given global model.

## 5.4 Tangrams: Computational theory

**Definition**. Tangrams are seven piece jigsaw puzzles consisting of solid shapes made put of wood. To be more precise, jigsaw world has the following constraints:

(1) All figures are polygons – a square, a parallelogram and five triangles, with the triangles having three possible sizes.

(2) The task, for the solver is to assemble the pieces to form a square (figure 5.2).

Tangrams can be represented quite easily within the derivational approach. First, let us define the notion of *equivalent derivations* as follows:

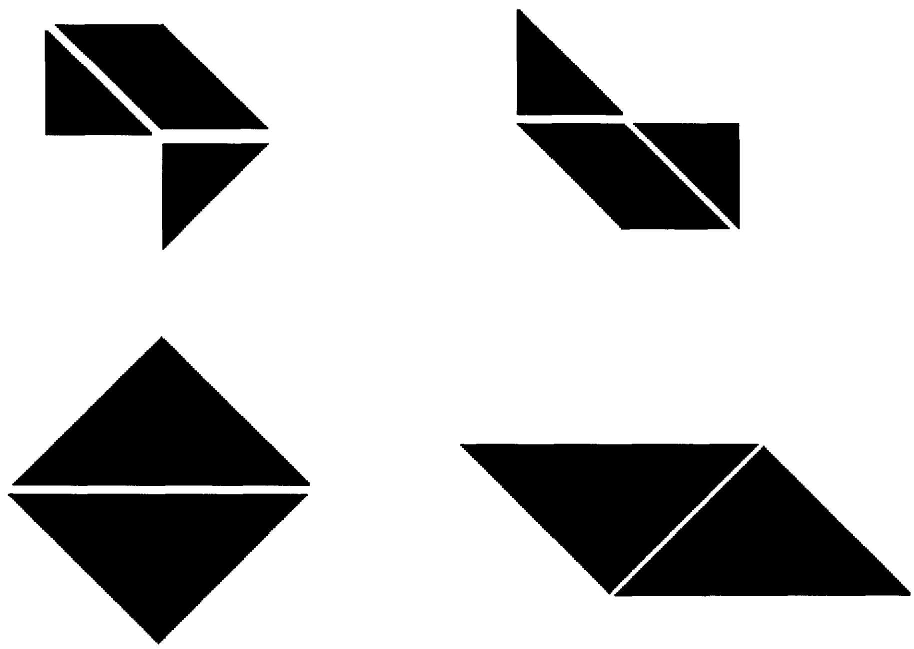
**Definition**. Let $D_1, D_2 \ldots\ldots D_n$ be derivations. For each derivation, $D_i$, let $C^i_\alpha$ be its starting concept and let $C^i_\omega$ be its ending concept. Then, $D_1, D_2 \ldots\ldots D_n$ are equivalent derivations if

$$C^1_\alpha = C^2_\alpha = \ldots. C^n_\alpha \textit{ and } C^1_\omega = C^2_\omega = \ldots. C^n_\omega$$

i.e., their beginning and ending concepts are the same. We can denote a class of equivalent derivations with common initial concept $C_\alpha$ and common final concept $C_\omega$ with the following notation: $C_\alpha \mapsto .. \mapsto C_\omega$

We can view each tangram as a class of equivalent derivations. A tangram can be seen as a mapping problem with two distinct abstract-concrete pairings. These two pairing can be understood in terms of two models, 2D LOCAL GEOMETRY and PUZZLE that have to be jointly represented in order to define jigsaw world.

"2D LOCAL GEOMETRY" uses simple, topological features to represent subsets (henceforth, "clusters") of a puzzle. The basic unit of the local geometry is a non-accidental relationship between adjacent pieces (figure 5.3). The non-accidental relationship takes the form of an exact match between the boundary of one piece with the boundary of another piece (figure 5.3). Formally, the category 2D LOCAL GEOMETRY is the category whose objects are connected clusters of 2D puzzle pieces and whose arrows are edge matches between clusters. In each case, the *minimal mapping* of two pieces or clusters gives rise to a cluster that arranges the pieces non-accidentally with respect to each other.



**(a)** **(b)**

**Figure 5.3: Non-accidental alignments between tangram pieces. The right pair and the left pair consist of the same individual pieces but in different arrangements**

2D LOCAL GEOMETRY was crucial in the understanding of prepositions as well, where I argued that the meaning of prepositions can be understood in terms of simple (i.e., of depth less than 2) non-accidental relations between objects in space, such as contact and containment. In this sense, local geometry in the guise of non-accidental features is basic in most representations of space. However, local geometry may represent the simplest model of spatial representation. More complex spatial tasks like jigsaw puzzles show that spatial representations deviate from the purely local non-accidental features encoded in 2D LOCAL GEOMETRY.

To take a more technical perspective, one can ask, 'what kind of constraints do we need to map spatial concepts of large depth?' In particular, the tangram puzzle illustrated in figure 5.2 is of depth 6. By understanding the mapping problem for tangrams, we can explicitly see some of the ways in which purely local cues are not enough.

For example, note that there are many possible minimal mappings of the same pieces, since they can enter into several non-accidental relations with each other such as the two pairs in figure 5.3 (a) and 5.3 (b). This poses problems to the solver, for he has to choose between equally valid minimal mappings. In other words, there are several local minimal mappings that are equally plausible in the absence of other data. What should the solver do? The problem is further complicated when the solver learns that an entire sequence of minimal maps may lead him nowhere, i.e., while these maps are all non-accidental locally, they do not cohere to give rise to a square at the end (figure 5.4). In other words, while local minimal mappings give rise to many possible derivations, only a few of these derivations actually get to the final, desired state.

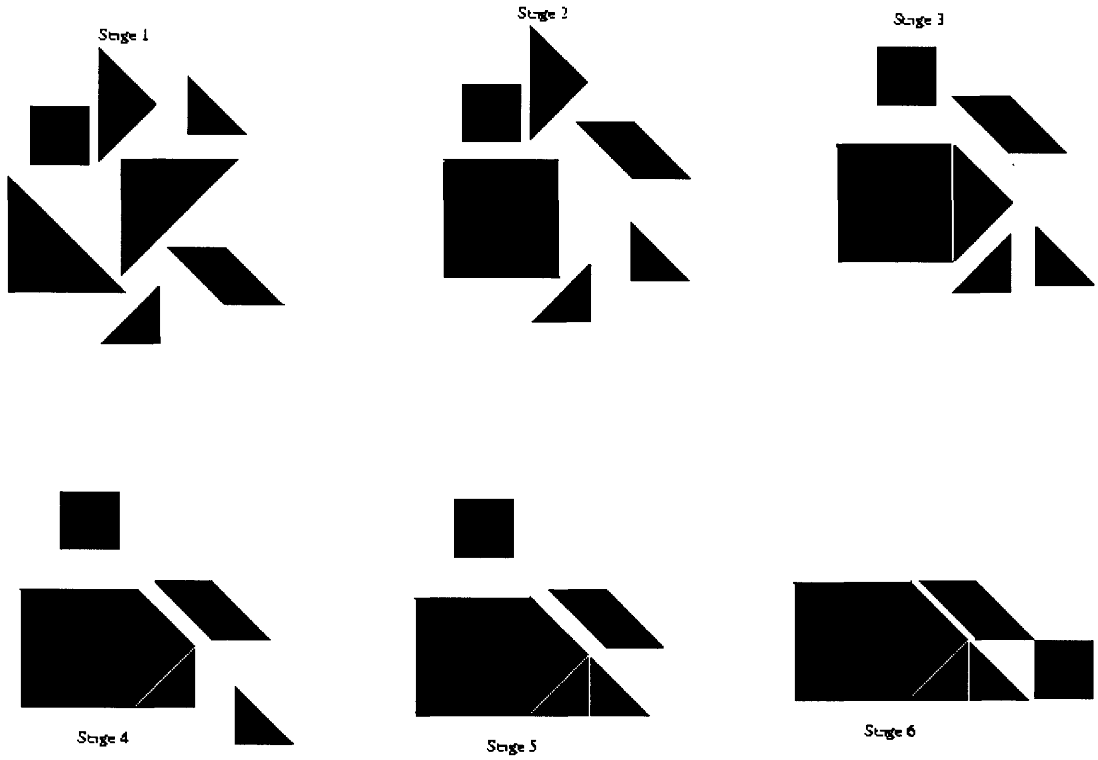



**Figure 5.4. A sequence of minimal, non-accidental mappings that fails to solve the Tangram**

The fundamental reason for the complexity of a tangram is that the (abstract) goal is always in the future, since the puzzle has to be solved sequentially, rather than in parallel, as one can imagine a computer being able to do. In order to solve the tangram sequentially, the solver has to combine attributes of the goal state (that it is a square) with the current state of the puzzle, which may or may not be goping in the right direction. Without some "Top-down" feedback from the goal state, it would be too hard to pick exactly the right set of minimal mappings that lead the solver to the solution. Therefore, we need additional non-local constraints in order to pick out a right path from the initial state. These non-local constraints are embodied in another model, that I call PUZZLE.

"PUZZLE", embodies a set of global constraints without which the tangram cannot be completed. One can think of this global model as the knowledge about the large-scale structure of the puzzle available to the solver at all times. This knowledge consists of three abstract elements.

(1) Completeness: All of the pieces are used in the final puzzle.
(2) The shape of the puzzle is a square.
(3) Size Constraint: An approximate representation of the size of the square.

The global model for the tangram is neither 2D LOCAL GEOMETRY by itself, or PUZZLE by itself, but rather the minimal mapping the two. Formally,

TANGRAM = 2D LOCAL GEOMETRY * PUZZLE

In other words, one needs both local information (the non-accidental features from 2D LOCAL GEOMETRY) and non-local information (coming from PUZZLE) in order to solve the tangram. To take one particular example of the use of a non-local constraint combined with a local non-accidental relationship, let us see how the size constraint can be used to select the appropriate local alignment of pieces. One way to do so (an intelligent entrypoint in my experience) is to begin by asking, even before attempting to solve the puzzle:

What is the length of a side of the square?

Now, we know that the length has to be implementable in terms of one of the pieces of the tangram. Therefore, given an approximate idea of the size of the square, a reasonable assumption is that the best "side-piece" is one of the big triangles. Once we use the non-local "size constraint" and put the big triangular pieces together, the tangram becomes much easier (figure 5.5)

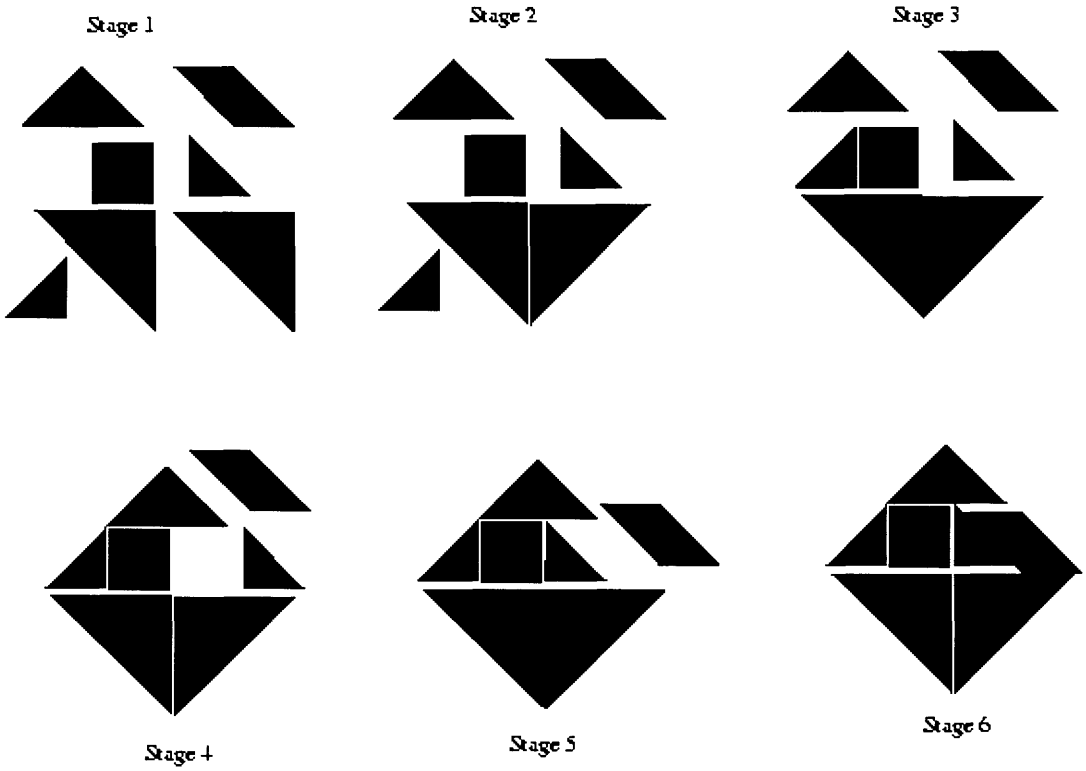



**Figure 5.5. A correct sequence of minimal, non-accidental mappings using a non-local constraint**

The use of the size constraint is an example of non-local information being used to solve the problem of selecting between equally valid local minimal mappings. Once we have non-local information, the problem of sequencing local mappings becomes far more tractable, as we will see in sections 5.7-5.9. To conclude the exposition of tangrams note that:

(1) Each tangram is equivalent to a partial ordering whose maximal element is the complete tangram (figure 5.5), in the shape of a square.
(2) The lowest elements of the partial order are individual pieces and in between the minimal and maximal elements, we have clusters of pieces with cluster sizes varying from 2 to 6 (figure 5.5).
(3) Furthermore, this partial order is generated by the minimal mapping of two global models, 2D LOCAL GEOMETRY and PUZZLE.

(4) The task for the solver is to ascend this partial ordering from the minimal elements to the maximal element while making sure that he does not end up with a false maximum (as in figure 5.4).

By its very nature the solution of the tangram requires multiple steps with each step involving a minimal mapping. I argue that the combination of local and non-local cues is crucial to the solution of spatial tasks involving multiple steps. In the remainder of this chapter, I give a detailed analysis of how non-local cues can reduce the complexity of spatial tasks immensely.

## 5.5 Emergent Properties.

In the previous section, I showed that local non-accidental relations, while essential, are not enough to solve the tangram. The local nature of the nonaccidental features derived from simple topology should alert us to the source of emergent properties, namely, *scale*. The local cues are all at the scale of the individual piece of the puzzle and its immediate neighbors. On the other hand, the emergent complexity of the puzzle is measured by the depth of the partial ordering of the puzzle, which is the ratio – size(PUZZLE)/size(2D LOCAL GEOMETRY) = N, where N is the total number of puzzle pieces. Each cluster of pieces has a scale between 1 and N.

A quick calculation shows why the puzzle can get difficult as the size of the puzzle increases. First define the depth of a subcategory in a partial order as follows:

**Definition**: If M is a maximal element in a partial order and B is an element in the partial order such that $B \prec M$, then define the depth of B as
$\text{The smallest number d : There exist elements } C_i \text{ with } B \prec C_1 \prec C_2 \prec .. \prec C_{d-1} \prec M$

In a jigsaw puzzle with N elements, a cluster of size j has depth N-j, because there are clusters of size j+1, j+2 ... N-1 between the cluster and the maximal element, which is the puzzle itself. Solving the jigsaw puzzle is equivalent to climbing the partial ordering tree generated by JIGSAW. However, in the absence of other information, the only way to solve the puzzle is to use *local* ascent, i.e to solve the puzzle by climbing the partial order tree one depth level at a time. While this is guaranteed to give a solution (in the absence of errors) it is time consuming because there is a fundamental problem, namely, at each stage the solver has to grow the set of clusters he has by picking new pieces. In the absence of extra information, the clusters have to grown painstakingly, with each new piece being tested out. At each step, selecting a new piece takes on the order of N-j comparisons where N is the total number of pieces and j is the total number of pieces already clustered. To complete the puzzle, the solver will take N+N-1+ N-2 +.... comparisons, which is of the order of $N^2$. To see why this can be overwhelming, let N = 1000 (thousand piece puzzles are universal in most gaming stores). Then $N^2 = 1{,}000{,}000$. If it takes 1 second to perform a comparison, it will take about eleven and a half days just to perform the comparisons, with no breaks. In other words, while all equivalent derivations are equal in principle, humans must have access to regularities in the way the derivations unfold dynamically if they are to have a shot at solving the puzzle.

**Definition**. Let **D** = $\{D_1, D_2 \ldots\ldots D_n\}$be equivalent derivations with common initial point $C_\alpha$ and common final point $C_\omega$. Then each derivation is a sequence connecting $C_\alpha$ and $C_\omega$. Let $\mathbf{D_{small}} = \{D_\alpha, D_\beta \ldots D_\psi\}$of cardinality φ be a *small* subset of **D**, i.e., a subset whose cardinality is much smaller than that of **D** (formally $\varphi << n$). A dynamical regularity is a probability distribution on **D** concentrated around a small subset $\mathbf{D_{small}}$ of **D**.

At this point one can ask, "What are sources of dynamical regularities?" One natural answer to this question, within the framework of category theory, is that the dynamics of an equivalent derivations should be guided by models correlated with, but independent of the models defining the equivalent derivations. I call these regularities *path regularities*.

> **Definition**: Let $\mathbf{D} = \{D_1, D_2 \ldots\ldots D_n\}$be a class of equivalent derivations with common initial point $C_\alpha$ and common final point $C_\omega$. For each derivation, $D_i$, let $l_i$ be its length, i.e., the derivation $D_i$ consists of $l_i$ steps. Then a path regularity is a probability distribution on the set **D** peaked around the derivations of minimal length, i.e., $D_j$: $l_j \leq l_i$ for all i.

Path regularities are important for understanding jigsaw puzzles because the solver would like to solve the puzzle with as few moves as possible. Therefore, the natural question is: what is a good source of path regularities in a jigsaw puzzle? The answer to this question is implicit in the analysis of complexity in the jigsaw puzzle, namely, scale. Given that the source of complexity is the discrepancy between the scale of the puzzle as a whole (quantified by the total number of pieces) and the scale of the individual piece, the dynamical regularities should be based on *correlations between clusters at different scales*. For a tangram, two plausible ways to correlate clusters at different scales are

(1) Use non-local cues to choose the "right" local cluster.
(2) Circumvent local ascent by jumping ahead to a new cluster that contains the current set of clusters. From the point of view of partial ordering, this is equivalent to leaping several depth levels to a node (of smaller depth) that dominates the current set of nodes.

In tangrams, we can see the influence of both of these strategies in correlating clusters at different scales. I already showed how the "size constraint" can be used to choose the correct set of pieces to form the side of the square. Local ascent can also be circumvented in tangrams by using large-scale non-accidental features rather than local non-accidental features.

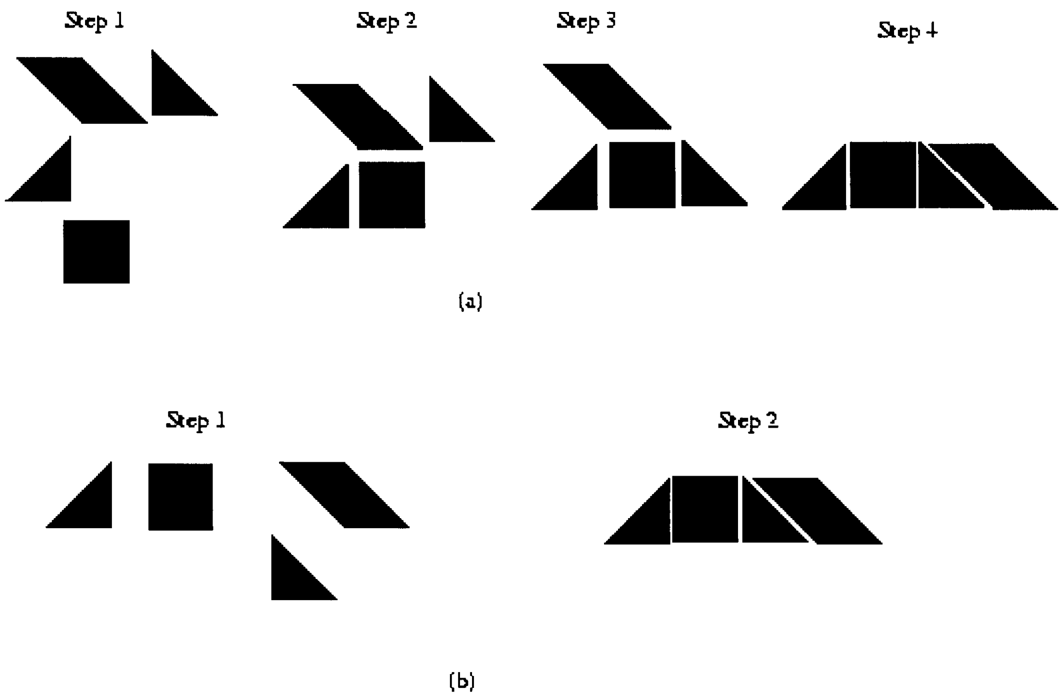



**Figure 5.6. An example of a local combination of non-accidental features in 5.6 (a), compared with a non-local combination of non-accidental features in 5.6 (b)**

For example, a relatively large number of non-accidental relations may come together at once. Therefore, instead of an assembly that takes place one step at a time, as in figure 5.6(a), there might be a single step assembly of a large cluster, as in figure 5.6(b). The guiding intuition is that a large-scale group of non-accidental features may come together at the same time. One way that might happen is if by chance, the relevant pieces are aligned together as shown in figure 5.6(b). In general, one can expect large scale non-accidental arrangements to cohere every once in a while, far more infrequently than local non-accidental features, but often enough that they make a difference. In the next section, I show how even a small number of large-scale non-accidental features can lead to a sharp decrease in the complexity of the task.

## 5.6 Algorithmic Analysis.

I have argued that the tangram, while solvable in principle on the basis of intrinsic cues alone, is hard to solve in practice without including external constraints. Informally, the complexity of the original puzzle is related to the issue of *scale* and that external cues lead to correlations between different scales. In this section, I make these intuitions precise using tools from graph theory.

From this point forward, I am going to think of the partial ordering tree as a graph, albeit a special one. All of the techniques I am using are general to all graphs though, not just to tree. First, let us see what the problem of solving the puzzle consists of:

(1) The puzzle is generated by a global model whose top node is the completed puzzle. The diameter of the partial ordering when seen as a graph, (the diameter is the maximum value of the minimum pair-wise distance between points in the graph) is at most twice the depth of the tree, i.e., the distance between the maximal node and the lowermost leaf.

(2) Solving the puzzle is equivalent to climbing the partially ordered tree to the top. In order to solve the puzzle, the solver can perform one of three kinds of actions:

   (a) In the absence of other cues, he can use local edge connectivity to bring together clusters that share edges. Bringing pieces or clusters together using local connectivity is a rather laborious process since the solver has to check a large number pieces one by one to see if their edges match with one of the edges of the cluster at hand. In any case, climbing the tree using local connectivity moves up the ladder only one depth level at a time.

(b) He can use shading based cues or shape based cues to bring together a large number of pieces simultaneously. In this case, the solver climbs orders of depth up the partial order tree in one go.
(c) The solver may merge two clusters. If this is done, the solver suddenly moves several steps up the partial order tree. Since two clusters can be merged if any one of their pieces share an edge, this can be a particularly profitable way to move quickly up the ladder.

Now note that while step (a) above is guaranteed to solve the puzzle, the actions in steps (b) and (c) occur with lower probability. Their occurrence depends on the solver noticing the existence of a non-local cue or if he finds himself looking at two clusters that share an edge, all of which are events that are inherently random. In other words, in trying to model the tree climbing behavior of the solver, we have to acknowledge two kinds of movements by the solver: one that is slow but sure, and the other that is faster, but less likely. Combining these considerations, we can postulate the following model for climbing up the tree (Note that this model is similar in structure to Shimon Ullman's sequence seeking model (Ullman, 1996)).

(i) The solver climbs up the partial order tree by jumping from a higher depth node $\alpha$ to a lower depth node $\beta$ where $\beta$ dominates $\alpha$.
(ii) At any given stage, with high probability $1-p$ where $p \ll 1$, the solver climbs up the tree using local cues, ascending the tree one depth level.
(iii) At any given stage, with low probability, $p$, the solver climbs the tree many steps at a time using a non-local cue from one of the sources outlined in (a) and (b) above.

Given these constraints, how does the solver perform on a given puzzle? The notion of multiple-scale graphs gives the right formal tools to understand this question (Kasturirangan, 1999).

### 5.7 Multiple Scale Graphs and the Multiple Length Scale Hypothesis

**Multiple scale graphs:** Let $N$ be a graph with $n$ vertices. Let the distance between two vertices, $v$ and $w$ in $N$ be denoted by $\text{dist}_N(v, w)$. Let $\omega$ new edges be added to $N$, forming a new graph $N'$. Call the set of new edges $\Omega$, so that cardinality( $\Omega$ ) = $\omega$. For each new edge $e^{\omega}$, $\text{dist}_N(e)$ is the distance- $\text{dist}_N(v^e, w^e)$ - in $N$ between the endpoints $v^e$ and $w^e$ of $e$. The distribution of length scales in $\Omega$ is measured by the function-

$$D: \Omega \rightarrow (1, \infty), \; D(e) = \text{dist}_N(e). \tag{1}$$

Then we say $N'$ is multiple scale with respect to $N$, denoted $N' \angle N$, if –

There are many length scales $l_i$, $i = 1, 2, 3, \ldots r$ ; $r >> 0$; $0 < l_1 << l_2 << l_3 << \ldots . l_r \leq n$

and $\quad \forall i : i \leq r, l_i \in D(\Omega)$. (2)

Finally, we say that a graph $N$ is a *multiple scale graph* if it has a subgraph $M$ with the same number of vertices such that $N \angle M$. In other words, $M$ contains all the vertices but not all the edges. In the case of the jigsaw puzzle, one can translate these definitions in the following form: Let the Partial Ordering Tree for TANGRAM be called $J_1$ and the partial ordering for the augmented puzzle –including shading, shape and other non-local cues- be called $J_2$.

Then, $J_1 \angle J_2$, i.e., the augmented partial ordering is a multiple scale graph with respect to the partial ordering from the unaugmented original puzzle. Furthermore, since:

(a) The complexity of the puzzle is proportional to the depth of the partial order tree and,

(b) The depth of the partial order tree is regulated by the diameter of the tree;

There is a direct connection between the complexity of the jigsaw puzzle (both in the original case and in the augmented case) and the reduction in diameter of a graph when new, non-local edges are added.What is the relationship between the existence of multiple scales in a graph and its behavior? One answer to this question is given in the multiple length scale hypothesis below.

> **The Multiple length scale hypothesis**: The reduction in average path length when adding new edges to a graph is proportional to the number of length scales present in the new connections and the number of connections at each length scale. In other words, *any* graph, random or deterministic, that is a result of adding sufficiently many connections at many different length scales to a graph, will exhibit a rapid decrease in diameter. Moreover, the decrease in diameter is not dependent on whether the new connections are long range or short range. What matters is the *number of scales* represented in the new connections.

In the next section, I show analytically how the multiple length scale hypothesis leads to a rapid decrease in graph diameter.

### 5.8 Deriving Path Length Behavior from the Multiple Length Scale Hypothesis

**Tight covering**: A length scale $l$ *covers* the graph $R_{n,k,\omega}$ if for each vertex $x$ in $R_{n,k,\omega}$, there is a new edge $e$, $\text{dist}_{\text{reg}}(e) \cong l$ such that $x$ lies between the endpoints $v$ and $w$ in the regular graph $R_{n,k}$. In other words, the new edges of scale $l$ wrap around the graph. Furthermore, in order to reduce average path length, it is not enough that a length scale $l$

cover the graph. Suppose there are two vertices $v$ and $w$ such that $\text{dist}_{\text{reg}}(v,w) \cong l$. Ideally, we want the shortest path between $v$ and $w$ in $R_{n,k,\omega}$ to contain edges of length $\cong l$ only. In order to ensure this property of the shortest path, we have to make sure that the end point of an edge of length $\cong l$ is close to the starting point of another edge of length $\cong l$, i.e., the new edges of length $\cong l$ are tightly packed. In such a situation, we say that a covering of the graph by edges of the length scale $l$ is *tight*.

> Let $R_{n,k,\omega}$ be a regular graph that has been rewired by adding $\omega$ new connections. Furthermore assume that there is a scaling factor $s$ and length scales $s^i$, $i \le \log_s n$ such the new edges are chosen randomly and that their distribution is uniform with respect to the length scales, i.e.,
>
> $$P\{ e \in \Omega : s^k \le \text{dist}_{\text{init}}(e) \le s^{k+1} \} = P\{ e \in \Omega : s^l \le \text{dist}_{\text{init}}(e) \le s^{l+1} \} \quad 0 \le k,l \le \log_s n \tag{3}$$
>
> Let k $\cong \log_s n$ and let $\omega \cong n << nk$. (4) implies that the number of edges per length scale is approximately $\frac{n}{\log_s n}$.

If $s^i > \log_s n$, the edges of that length scale almost certainly cover $R_{n,k,\omega}$ (in the sense of IV.1.3). The average distance between successive edges of length $s^i$ is about $\log_s n \cong k$, which means that in $R_{n,k,\omega}$ successive edges are separated by about an edge since $k \cong \log_s n$. Therefore, all the length scales $s^i \ge \log_s n$ *tightly cover* the graph.

Let $v$ and $w$ be two vertices such that $\log_s n \le s^i \le \text{dist}_{\text{reg}}(v,w) \le s^{i+1}$, $i \le \log_s n$. The following inequality is a consequence of the fact that the scales tightly cover the graph-

$$\text{dist}_{\text{final}}(v,w) \le 2s + \text{dist}_{\text{final}}(x,w) \tag{4}$$

Where $x$ is an intermediate point such that $s^{i-1} \le \text{dist}_{\text{reg}}(x,w) \le s^i$. Inequality (4) gives rise to

$$\mathrm{dist}_{\mathrm{final}}(v,w) \leq 2s.\log_s(\mathrm{dist}_{\mathrm{init}}(x,w)) \quad (5)$$

showing that the distance in the final graph is uniformly logarithmic for all pairs $v$ and $w$.

In other words, if there are enough non-local connections in a graph, the average distance between two points is logarithmic in the size of the graph. Now, note that in the case of the jigsaw puzzle, the size of the puzzle was N, where N is the number of pieces and is also the depth of the partial order tree. An earlier calculation showed that the puzzle complexity was approximately of the order $N^2$. Now, with the new non-local connections, the complexity of the puzzle can be reduced to the order of $(\log N)^2$, a much smaller number.

Of course, there are other non-computational factors that will mitigate the effect of the non-local links. In particular, if the puzzle has many pieces a lot of time is spent moving pieces from one location to another, making sure that the pieces fit together and stay connected. Each one of these tasks is purely physical, they are computationally easy but each one them takes a fixed amount of time that adds up in the long run. Nevertheless, factoring out these physical concerns, one can see how the augmentation of the original global model by adding other models can lead to a drastic decrease in the computational complexity of the puzzle.

### 5.9 Simulations

In the previous section, I formulated the jigsaw puzzle problem in terms of multiple scale graphs and demonstrated how extrinsic cues can give rise to correlations between scales, thereby make it easier to solve the jigsaw puzzle. The “multiple scale graph” model for jigsaw world makes two predictions:

(1) The average path length in a multiple scale graph is inversely proportional to the number of scales in the graph.
(2) A multiple scale graph is much more robust to "modal" noise than a graph with only one scale (where modal noise is noise that is concentrated at some length scale).

The first of the two predictions follows directly from the analysis in equations (4)-(6) above. The second prediction can also be understood as follows:

> If "modal" noise is added to a multiple scale graph, then the only length scale at which connections are disrupted is the length scale of the noise. All other length scales are still available. Therefore, the average path length will only decrease linearly at that length scale. On the other hand, if the graph has only one scale and the modal noise is concentrated at the critical length scale, then the connectivity of the graph will be severely damaged. To get a sense for this intuition, think of a social network, say a company. Suppose a person moves up the corporate ladder depending on the number of superiors who recommend him for the promotion. If there are two workers, one with a single backer (however powerful) and the other with multiple backers, who is more likely to keep getting promoted, given that everybody keeps changing jobs. The person with a single backer is likely to get stalled sooner or later because his mentor may move somewhere else.

The second hypothesis was tested in a computer simulation using the **maximum rule** for graph rewiring.

> Let $N_t$, $t = 0,1,2\ldots$be a graph that is evolving in time. At each time step $t$, a new connection is drawn between the two points that are most distant from each other in $N_{t-1}$. In the randomized version of the maximum rule, $m$ pairs of points are selected randomly at each time step $t$ and an edge is drawn between the most distant of the $m$ pairs (in $N_{t-1}$). The results of a computer simulation of the maximum rule are displayed in table 1 along with the results for a randomly rewired graph. In the randomly rewired graph, the rewiring is done assuming a uniform distribution of the new connections.

From an intuitive point of view, it is quite surprising that the random rewiring rule performs as well (if not better!) than the maximum rule that creates short cuts between far off points. However, in the light of the multiple length scale hypothesis, the results do not seem surprising. Table 5.2 shows the comparison between the maximum rule and the random rewiring rule. For $\omega$ = 5, 10 and 25 we got $r_{\text{max rule}}$ = 3, 5 and 9 respectively while $r_{\text{random rule}}$ = 5, 8 and 13 respectively. Therefore the number of length scales in the randomly rewired graph is more than in the maximum rule graph if the number of new connections is the same. Only when the number of new connections is relatively large does the maximum rule do better than the randomly rewired graph and by this stage, both graphs have very short average path length.

The comparison of the maximum rule and the random rewiring rule suggests that the strongest determinant of small-world behavior is the presence or absence of a variety of length scales (and not whether the new connections are short range or long range). This hypothesis was tested in a computer simulation by selectively removing length scales from a randomly rewired graph.

The initial graph was a regular graph with $n$ = 275 and $k$ = 6, randomly rewired by adding 50 new connections. Then, (approximately) half of the connections were removed in three different ways- The shortest 50% of the connections, the middle 50% of the connections and the longest 50% of the connections. The average shortest path length was computed for each of the three graphs. The results are shown in Table 2 below. Table 5.3 shows that for $\omega$ = 27 there is no difference in the performance of the three graphs, suggesting that the absolute length of the new connections is not as important as their distribution.

**Table 5.2: Comparison of the maximum rule and random rewiring rule, $n$ = 275 and $k$ = 6, for 1000 randomly chosen pairs of vertices.**

| | Maximum rule | | | | | Random rewiring | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of connections in initial graph** | 825 | 825 | 825 | 825 | 825 | 825 | 825 | 825 | 825 | 825 |
| **Number of new connections ( $\omega$ )** | 5 | 10 | 25 | 50 | 100 | 5 | 10 | 25 | 50 | 100 |
| **Number of scales, $r$, represented by the new connections ( maximum $r$ = 20)** | 3 | 5 | 9 | 15 | 18 | 5 | 8 | 13 | 19 | 20 |
| **Average shortest path length, $L$, in initial graph** | 25.14 | 24.63 | 25.00 | 24.25 | 24.83 | 24.81 | 25.16 | 24.90 | 25.42 | 25.14 |
| **Average shortest path length in final graph** | 18.05 | 14.12 | 9.60 | 7.12 | 5.42 | 16.42 | 12.73 | 9.97 | 8.11 | 5.97 |

**Table 5.3: Selective removal of length scales in a randomly rewired graph with $n$ = 275 vertices. 1000 random pairs of vertices were chosen in each of these graphs**

| **Length scales removed** | 70 - 138 | 4 - 70 | 30 - 100 |
|---|---|---|---|
| **Initial number of connections** | 825 | 825 | 825 |
| **Number of new connections** | 27 | 27 | 28 |
| **Average shortest path in initial graph** | 24.99 | 24.91 | 24.81 |
| **Average shortest path in rewired graph** | 10.08 | 10.04 | 10.01 |

As we can see from the above table, the removal of edges makes a big difference when the original graph has only a few scales, while it does not do so when the graph has multiple scales.

## 5.10 Conclusions

To summarize the findings of this chapter, the derivational approach offers a systematic, formally describable answer to the problem of learning regularities during the performance of a task. What the solver learns are the dynamical regularities that are not contained within the original statement of the problem itself. These dynamical regularities are often path regularities, i.e., privileged ways of moving from the starting point to the end point of the derivation. The reason why external representations are crucial is because they are the source of dynamical regularities. It is this shift from a "static" to a "dynamic" perspective that allows us to quantify the information being learnt by the solver. In particular, the analysis of puzzles shows that the notion of scale plays an important role, which can be analyzed formally using the notion of multiple scale graphs. In the appendix, I analyze another puzzle domain and present experimental evidence that supports the points made in this chapter.

# Chapter 6: Generics and Extensions

## 6.1 Introduction.

In the previous two chapters, I showed how the interplay between abstract and concrete aspects of representation can be understood within the same framework by invoking the notion of *derivation*. In chapter 4, I applied the derivational method to model the static aspects of the semantics of spatial concepts (in the form of spatial prepositions). In chapter 5, I outlined the derivational structure of the dynamical aspects of the semantics of spatial concepts, in the form of jigsaw puzzles. Both of these domains were directly correlated with perceptual aspects of space, that is to say, the concepts modeled in chapters three and four were clearly grounded in the external world. In this chapter, I want to extend the derivational method to two important linguistic phenomena – generics and quantifiers- that are not grounded in the external world in any obvious manner. Nevertheless, the semantics of these two phenomena exhibit the same interplay between abstract and concrete that we saw in the previous chapters. The derivational structure of these cognitive structures ostensibly removed from perceptual experience lends support to the argument, made by linguists such as Talmy (Talmy, 2000), that even the most abstract semantic phenomena are related to spatio-temporal, embodied cognitive structures.

## 6.2 An introduction to Generics and Quantifiers.

Generics and quantifiers are among the most intriguing linguistic constructs from the point of view of semantics. They share the three properties – abstractness, groundedness and flexibility- that I highlighted as being paradigmatic of spatial cognition. Let me first start with the linguistic notion of generic classes and statements. Note that the linguistic notion of genericity is related to but distinct from the topological notion of genericity that were important earlier.

Generics are linguistic concepts as well as statements connecting linguistic classes that have a *free variable* (Carlson & Pelletier, 1995; Prasada, 2000). That is to say, a generic concept denotes a collection of entities without specifying exactly which member of that collection is being denoted by the generic. Examples of generics are the usual linguistic categories such as “Dog” “Table” and others, where we assume, unless otherwise stated, that the utterance denotes a whole class of entities, not any particular one. Similarly, a generic statement is a statement that is a *free statement*, i.e., it is a relation between two classes of entities that does not specify a concrete “truth value”. For example, the statement (A) below is not necessarily true of a given dog.

(A) A dog has four legs.

Generic statements such as (A) above, are linguistic utterances about a concept that are “central” to that concept (Prasada, 2000). Generic statements have many interesting properties. As I said earlier, they are abstract. The class “Dog” extends to a potentially infinite number of real world instances and furthermore, statements such as (A) can be acceptable, without being empirically true. At the same time, to use linguistic terminology, they have an *extension*, i.e., each generic concept denotes a set of real world entities and each generic statement is true or false of some real world configurations as well. In other words, generics are grounded in concrete real world objects, events and situations.

Finally, generics are extremely flexible. In fact, they are the usual way in which we couch our understanding of a given domain in natural language. Like other spatial representations, the semantics of generic concepts and statements is quite complex.

On the one hand, generics almost seem definitional; For example, statement (A) above seems like a good answer to the question:

What is a dog?

However, generic statements are not *really* definitions, since they admit exceptions. A dog is still a dog if it is lame and has three legs. Similarly, generic statements are statements about a whole class of objects; the sentence "Dogs have four legs" seems to say something about dogs as a whole. At the same time, a generic statement need not *literally* cover the entire class that it relates; "Dogs have four legs" does not necessarily imply "Fido has four legs" (Fido could have lost a leg).

Generic statements are the standard way to make statements about a class of objects - they are our implicit theoretical statements about the state of affairs in a given domain. For example, if we are asked the question "Describe the United States", we might answer as follows:

(1) It is a democracy.
(2) It is wealthy
(3) It is in the Western Hemisphere

and so on. Each one of these statements is a generic statement, i.e., it is a statement that is essential to the concept "United States".

How can we model the semantics of generic statements, given the issues raised above? Generics have usually been seen as *defaults*, where a default is a value of a feature or property that is assumed to be valid unless there is an explicit assumption to the contrary. However, it is not clear whether calling a generic statement a statement about the default value of some property helps matters much, for we are left with the problem of clarifying the meaning of defaults. How does a default value capture a feature that is central to a concept?

Before going on to understand the meaning of generics, let us first realize that generics cannot be understood probabilistically. That is to say, a default value of a feature is not determined by the frequency of values. For example, even if all the dogs in Cambridge were lame, the statement "dogs have four legs" would have some significance – if anything, it might lead us to ask what happened to all the dogs in Cambridge.

From a category theoretic perspective, generic concepts and statements can be understood quite easily. Remember that each concept is a node in a partial order generated by the *geometric* notion of genericity. Furthermore, note that according to the maximal rule (see Chapter 3), an exemplar is classified as belonging to the lowest concept that is consistent with that exemplar. One can generalize this to properties as well. In other words, just like exemplars, one can attribute properties as follows:

**Definition**: For each pair of nodes in a partial ordering
$A, B$ that satisfy $A \prec B$, define their relative depth (as in chapter 5) as
The smallest number $d$ : There exist elements $C_i$ with $A \prec C_1 \prec C_2 \prec .. \prec C_{d-1} \prec B$.

**Definition**: Let $\Omega$ be a category (in the technical sense of chapter 2) and let $P_\Omega$ be the partial ordering generated by that category. Let C be a node that is generated by that partial ordering with exemplars {c1, c2...}. Similarly, let $P_\Phi$ be another partially ordered set and let D be a node in $P_\Phi$. Then we can model generic concepts and statements as follows:
(a) A generic concept is a node, C, in the partial ordering generated by a category.
(b) A generic statement (about a generic concept C), is a derivation $F: D \rightarrow C$ where D is the property that is being asserted.

Now we can understand the problems that were raised earlier. Why do generic statements look like definitions? The answer is that they assert the property D for a concept C, at a relative depth of 0 in the partial ordering. All the counterexamples are true only of *higher relative depth* nodes. Therefore, violation of the generic statement is *non-accidental*. For example, a dog that has three legs is so because something happened to the dog – having three legs cannot be asserted as being a property of the depth 0 node in the partial ordering for the concept "Dog".

We can also understand why generic statements seem to cover an entire class of objects - since the partial ordering is an ordering within an abstract category, an property that is attached to a 0 depth node will be true of almost all (except for a measure 0 subset) exemplars of that category.

**6.3 Extensions**. Now that we have modeled generic statements using derivations, we can also understand how one can get deviations from genericity. Note that generic concepts and statements are nodes in an *abstract* partial ordering. Deviations from genericity can happen in two ways

> (1) The default model for a generic concept or statement is changed because of the presence of additional *abstract* models. In other words, if $\Omega$ is the default model for a concept, C, and $\Psi$ is a new model that extends $\Omega$ in some fashion, then C has to be extended to a new concept C' in the mapped category $\Omega * \Psi$.
> (2) The abstract category, $\Omega$, is mapped on to a *concrete* category, $\Gamma$ that *binds* the *free* variables in $\Omega$. As we will see below, this mapping of abstract to concrete models can capture aspects of the semantics of explicit operators such as quantifiers.

Let us look at these two deviations from genericity one by one. For the first consider the two pairs of statements given below:

> (a) The book is on the table.
> (a') The waiter was on the floor serving the guests.
>
> (b) The light flashed.

(b') The light flashed until dawn.

In sentence (a) the preposition ON refers to the relationship of contact between the book and table, a relationship that does not have an explicit temporal component. This is consistent with the usual meaning of ON, which is purely spatial, without any temporal component. In (a') however, ON refers to an extended period of contact between the waiter and the floor, i.e., ON explicitly invokes a temporal extension. Similarly, in (b) the verb "FLASH" invokes a temporally limited event, which is consistent with its default meaning, while in (b') "FLASH" refers to a temporally extended event. What is there in (a') and (b') that invokes a deviation from the default meanings of ON and FLASH respectively?

One can use the category theoretic methods to show how deviations from defaults happen. First, note that in both (a') and (b'), the context provides cues that indicate an extended temporal interval (the presence of multiple guests in (a') and the phrase "until dawn" in (b')). In other words, the context induces an additional model that contains temporal information.

Now, let us see how one can understand the extension of the default spatial meaning by a temporal model.

> Let **S** be a spatial category and let $\mathbf{C_S}$ be a concept generated by the category, using its partial ordering. Let **T** be a temporal category, i.e., a model of some interval $(p,q) \subseteq R^1$. Since **S** and **T** are independent of each other, according to the mapping rule from chapter 3, the mapped category that contains the two is isomorphic to $\mathbf{S} * \mathbf{T}$, the product category. The spatial category, **S**, is naturally included in $\mathbf{S} * \mathbf{T}$, i.e.,
>
> $$\mathbf{S} \mapsto \mathbf{S} * \mathbf{T}$$

Furthermore, according to the mapping rule if $\mathbf{C_S}$ is a concept in **S**, then

$\mathbf{C_S}$ is mapped to the maximal concept in **S * T** that is compatible with $\mathbf{C_S}$, which is nothing but the concept $\mathbf{C_S}$ * (p,q), i.e., the concept whose spatial projection, $\mathbf{C_S}$, take up the maximal possible temporal duration.

Therefore, if ON is defined as "contact" in (a), it should be defined as "contact* (p,q)" where p and q are the beginning and ending times for the waiter being on the floor. Similarly if FLASH is defined as (**e**, $\mathbf{t_0}$) in (b), where **e** is the event of flashing at some location at some time $\mathbf{t_0}$, then FLASH would be defined as **e** * (p, dawn) where e is the flashing at some location, p is the start time of the flashing and the end point is dawn. In this way, we can see how the mapping rule and the fact that categories naturally admit extentions allow us to extend the meaning of concepts to those cases that do not obey the default because of the presence of additional models.

### 6.4 Quantifiers.

To see how the derivational method can be used to model quantifiers, let us see the different kinds of quantifiers that are possible. Suppose we start again with the statement

Dogs have four legs.

We can quantify this statement in many ways:

(a) All/No dogs have four legs.
(b) A few/some/many/most dogs have four legs.
(c) Two/seven dogs have four legs.
(d) This dog has three legs.
(e) There is a dog with three legs in the house.

To understand these different kinds of quantification from a category theoretic perspective (Link, 1998), note that the first three quantificational forms are *scalar*, i.e., they can put on a 1-dimensional scale in which they can be compared with each other. The fourth quantifier, "This," introduces a concrete *coordinate frame* that distinguishes one location from another. The fifth introduces a concrete *object*. Let us look at the quantifiers in (a-c) above first.

We can think of all of these quantifiers as *models* of a 1D interval. There are three possible models of the 1D interval - topological, qualitative metric and numeric.

The topological model $T_I$ of an interval I is the model that has three states (lower boundary, interior, upper boundary).


**Figure 6.1**

The qualitative metric model $Q_I$ of an interval I is the model which "covers" I with intervals. A simple qualitative model is one with four covering intervals (lower boundary, lower boundary $+\alpha$), (midpoint - $\alpha$, midpoint), (midpoint, midpoint + $\alpha$), (upper boundary - $\alpha$, upper boundary).

**Figure 6.2**

The numeric model $N_I$ is a model that divides the interval I into some specified number of sub intervals.

**Figure 6.3**

Note that $N_I$ is the finest and $T_I$ is the coarsest model, or in more formal terms,

(B) $N_I$ # $Q_I$ # $T_I$.

Let us call this ladder (B) above the Interval Hierarchy, IH for short. Then, we can define a quantifier as follows.

> **Definition**: Let $\Omega$ be a category and let C be a concept generated by the partial ordering of the category $\Omega$. Then each quantifier in (a-c) above is a derivation F: IH $\rightarrow \Omega$.

A diagram showing the distribution of quantifiers is given below in figure 6.4.

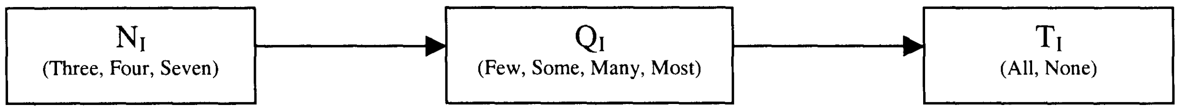



**Figure 6.4**

Similarly, to understand the quantifiers in (d-e) above, let us introduce two more concrete categories:

Let $\Theta$ be a three dimensional coordinate frame, with designated locations { $\theta_1, \theta_2, \theta_3 \ldots$ } and let $\Sigma$ be an object model (perhaps something like an object file - Pylyshyn (2001)) with a collection of objects $\{\sigma_1, \sigma_2, \sigma_3 \ldots..\}$. Then,

**Definition**. Let $\Omega$ be a category and let C be a concept generated by the partial ordering of the category $\Omega$.

(1) A definite description (This, That) is a derivation F: $\Theta \rightarrow \Omega$, which maps a location, $\theta_1$, to a generic concept C.
(2) An existential quantifier (There is, There are) is a derivation F: $\Sigma \rightarrow \Omega$, which maps an object, $\sigma_1$ , to a generic concept C.

In this way, one can see how quantifiers can be modeled as maps from abstract categories, i.e., categories of generic concepts, to concrete categories such as scalar models, locational coordinate frames and object models.

### 6.5 Discussion.

In this chapter, I showed how the derivational approach can be used to operationalize and model generic concepts, their extensions as well as quantifiers. These phenomena are not grounded in the external world in a straightforward manner. Yet they share the same derivational structure of abstraction and concreteness mapped on to each other. I argue that these results indicate a common computational structure for both the external grounding of spatial concepts as well as the grounding of non-perceptual aspects of language in spatio-temporal structures.

The general computational framework is as follows: both abstract concepts as well as concrete concepts can be modeled as nodes in a partial ordering. The way to combine linguistic concepts flexibly, i.e., to understand the *compositionality* of language, is to look at conceptual combinations as derivations, mapping one kind of category and its partial ordering to another category and its partial ordering. If the two categories are abstract, we get the phenomena of *concept extension*, which was explained in section 6.3.

If one of the categories is abstract, and the other is concrete, we get *quantification*, which was explained in section 6.4. Nevertheless, both of these phenomena can be understood within the same framework of derivations. To conclude, I argue that the derivational method gives us a *universal* tool to model concepts and their relationship to the world, because it can handle abstract and concrete phenomena at the same time.

# Chapter 7. Endgame

## 7.1 General Observations.

In this thesis, I have shown how cognitive systems are amenable to a theoretical, deductive approach, based on computational theory. In a computational theory, we can introduce formal frameworks as well as abstract constraints that would not be available in a normal empirical investigation, or for that matter, in a computer model that did not first look at the computational theory.

The logic of a computational theory is somewhat different from other cognitive theories. Normally, one might be satisfied with a model that predicts the phenomena in question. For a computational theory, the goal is different. We can think of a computational investigation as involving four essential steps:

(a) Identify key computational issues that constitute the 'problem' that has to be solved by the cognitive or perceptual system.
(b) Postulate computational constraints that solve the problem.
(c) Explain the relevant set of data using the computational constraints contained in the 'solution'.
(d) Generalize the theory to a new domain that was not part of the original problem statement.

If we can perform all four of these steps, then we would have a theory that cannot easily be replaced by a purely empirical investigation or by a computer model that only tries to 'fit' data. In the process, the computational theory may invoke constraints that seem quite abstract. The abstractness of a computational theory is not a drawback but an advantage, if it can explain a range of phenomena – for then, the success of the theory suggests that the actual principles that govern the phenomena are rather abstract – in which case a usual empirical observation or a model that merely fits data could not possibly capture the relevant principles.

In this thesis, I have followed the procedure that I outlined in the previous two paragraphs. I outlined three computational problems – abstractness, groundedness and flexibility – as being central to the understanding of spatial relations. Then I showed how these problems can be addressed within the framework of category theory. In order to do so, I invoked several abstract constraints: non-accidental features, minimal mapping and derivations. None of these constraints seem obvious from a purely phenomenal perspective. Nevertheless, I showed how the derivational approach offers a rather complete analysis of two domains of data, namely, the semantics of prepositions and the jigsaw world. Finally, I extended the theory to a new domain – the semantics of generics and quantifiers. In this way, I showed that the computational theory extends to phenomena that are ostensibly outside its domain of definition.

The main achievement of this thesis is to show that a computational approach to problems of semantics, i.e., the meaning of concepts and their relationship to the perceptual world, can lead to insights, such as the minimal mapping principle, that are not easily captured by other means. The computational approach of this thesis is complementary to the approach taken by linguists (Landau and Jackendoff, 1993; Jackendoff, 1996; Lakoff, 1987; Regier, 1996) and by AI researchers (Siskind, 1991; 1995) that offer either a phenomenological approach or a computer implementation. While these works are valuable, I believe that by formulating a computational theory, we can focus attention on deeper, more abstract factors that simply cannot be studied using descriptive or data-driven models.

To give one illustration, the meaning of a preposition encodes relations that are *at most depth 1 in the partial order* as discussed in chapter 4. From this it follows that languages do not have prepositional terms for the spatial relations illustrated in figure 7.1 below (in these figures, the spatial relations are of depth 2).

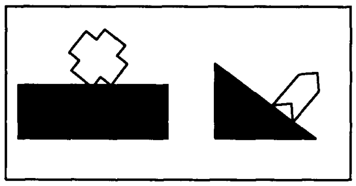


**Figure 7.1**

At the same time, the meaning of prepositions sometimes depends quite crucially on the detailed geometry of the figure and ground objects as illustrated in figure 7.2.

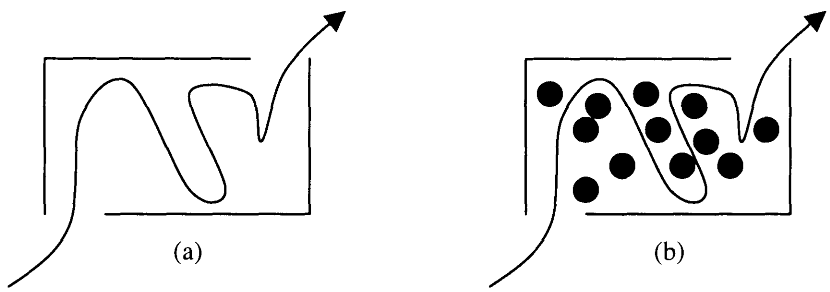



**Figure 7.2: Assume that a person, John, is standing at the location marked 1 and he needs to walk across a rectangular "Park" to the location marked 2. Is the curve in 7.2(a) a valid case of "John walked across the park? What about the path in figure 7.2(b)?**

How can the neglect of topological features and the incorporation of geometrical details coexist at the same time? Without a deeper computational analysis, there is no means by which we can understand the lack of complex features in language. The derivational approach, on the other hand, gives us precise tools that can model the lack of spatial relations such as the ones in figure 7.1, while at the same time acknowledging the role of geometrical details in figure 7.2. For these reasons, I believe that future investigations of semantics must incorporate a computational theory from the beginning.

That said, computational theory is not divorced from other levels of investigation. While the investigations in my thesis, have, for the most part been computational, in the next section, I make a few remarks about how these computational constraints can be implemented in a psychologically plausible manner.

## 7.2 Psychological Implementations

The key idea for transitioning from a computational account of dynamical mapping to a psychological account is the idea of routines (Ullman, 1987)– visual routines if the input is visual, but more generally, we can think of routines that are abstract as well. Routines can be given a computational as well as a psychological interpretation – though the two are closely tied to each other. For the sake of exposition, let me restrict myself to purely geometric routines. We can define a geometric routine as follows:

> **Definition**. A geometric routine is a transformation $T: S_1 \rightarrow S_2$ where $S_1$ and $S_2$ are geometric states.

We can assume that a geometric routine maps a geometric state $S_1$ into a geometric state $S_2$ that is at the same level in the structure lattice. Then, the space of routines acquires a structure lattice that parallels the structure lattice of the geometric state space. A list of prominent geometric routines is given below (figure 7.3):

(1) Coordinate Frame routines such as Move focus to a Blob location or a Principal axis location or a Minor axis location.
(2) Topological Structure Routines such as (a) Trace a curve/surface and (b) Fill in a curve/surface.

(3) Metric Structure routines such as: Draw a generalized cylinder around a given point/curve/surface.

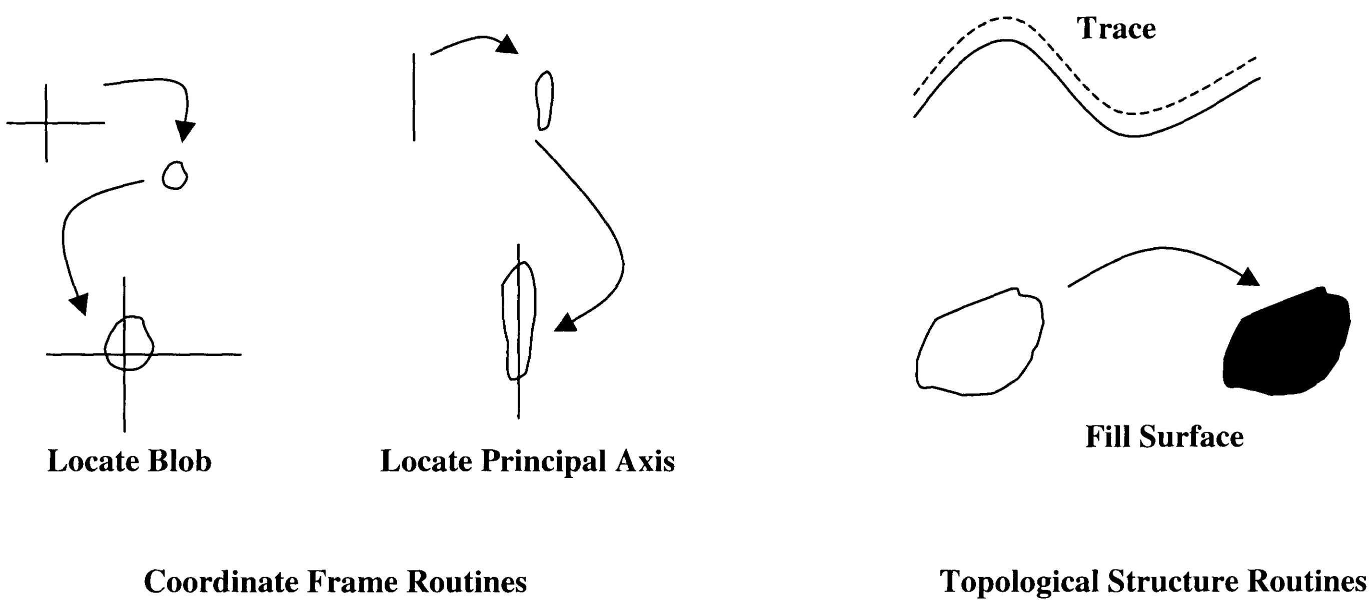



**Figure 7.3. Examples of Routines**

At the same time, we can define routines at a perceptual level as follows (quoted from Ullman, 1997):

> The perception of shape properties and spatial relations is achieved by the application of so-called "visual routines" to the early visual representations. These visual routines are efficient sequences of basic operations that are "wired into" the visual system. Routines for different properties are then composed from the same set of basic operations, using different sequences. (Ullman, 1997).

Experiments on animals with relatively primitive visual systems - such as ants and frogs- have demonstrated their ability to use coordinate frames and topological structures (Gallistel, 1990). Frogs snap at all moving blobs irrespective of the detailed shape of the blob (Lettvin et. al., 1959). Frogs can also detect if a curve is closed or open, suggesting that topological information is of importance to frogs. The primacy of topological structure and coordinate frames is not restricted to simpler visual systems. In humans, the detection of topological and coordinate frame information is effortless and immediate (Jolicoeur et. al., 1986; Ullman, 1997).

As Ullman and others have argued, visual routines are also related to computations performed by attentional processes. I argue that object based and spotlight based notions of attention (Scholl, 2001; Treisman, 1993) compute the basic building blocks of routines. Let us see how.

In any spatio-temporal situation, we have natural topological and metric constraints that are, in principle, available for restricting the set of objects worthy of the subjects attention at a given location and time. Let us first take a look at topological constraints. Often, the objects relevant to a particular task are all located within a bounded region. Objects outside the region are irrelevant to the task. A person cutting vegetables on a table may want to restrict his attention to objects on the table only (Rensink, et. al., 1997; Wolfe et. al., 1989; Wolfe et. al., 1997). Therefore, the topological constraint

ATTEND(SURFACE(TABLE))

is an appropriate constraint for the task of cutting vegetables. In general, many tasks involve restricting attention to a 0,1,2 or 2 dimensional topological object – a point, curve, surface or a volume - forming a natural boundary. Our world is organized in terms of topological constraints – computer screens, table tops, rooms, houses and so on.

In other situations, there may not be any pre-determined topological boundary demarcating the region of interest. In such a situation, it may be important to restrict attention to only those objects that are nearby (where nearby is determined according to context). While crossing a street the cars close by matter the most, not the cars that are far away. Restriction of attention to nearby objects is an example of a metric constraint. The three constraints, SET, TOPOLOGY, METRIC, noted above, are illustrated in figure 7.4.

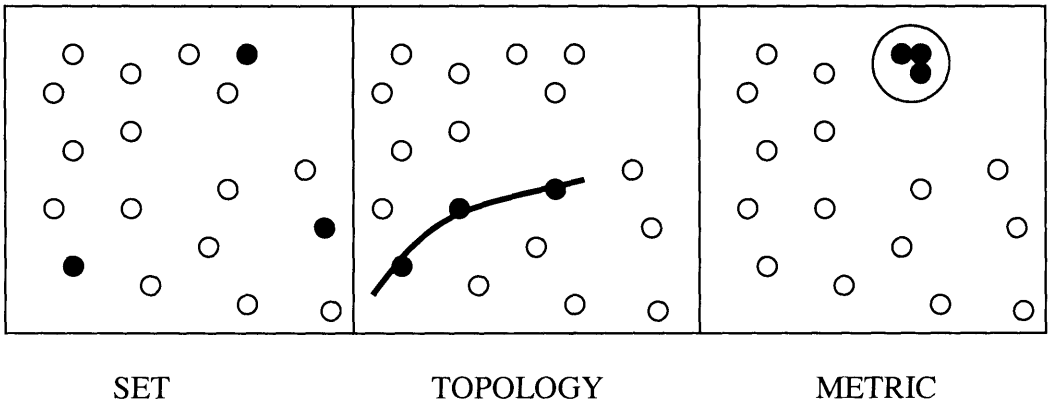



**Figure 7.4. Three kinds of Loci. The three black dots form a locus**

The SET-TOPOLOGY-METRIC constraint levels unify object and spotlight based forms of attention. Set theoretic constraints are often studied in the context of subitization (Pylyshyn, 2001). Topological constraints map on to object-based models of attention (Scholl, 2001). Metric constraints can clearly be seen as spotlights (Treisman, 1993). Within the psychological literature on attention, numerical, object based and spotlight based attention have been thought of as separate forms of attention (Scholl, 2001). However, from a formal point of view, they are all different aspects of the same hierarchy of computational constraints, a formal consequence of imposing a graded hierarchy of geometric restrictions on the locus of attention.

Therefore, there is a natural link from the formal representations of the computational theory and the units of selection that were central to the psychological models of attention (O'Regan, 1992). While these links add plausibility to the computational theory, they do not constitute a proof of the computational theory. Nevertheless, the theory can be tested more directly using psychophysical tests. For example, if the topological features of spatial relations are computed by an object-based model of attention, then we should be able to disrupt the ability to detect a particular spatial relation by asking subjects to attend to objects that do not have the spatial relation being tested. In any case, this is work for the future. The goal of this section was not to present a computational theory of attention, but to show that our understanding of psychological models can only increase if we augment them with computational concerns.

**7.3 Concluding remarks**

The fundamental problem motivating the investigations in this thesis, is "how are concepts grounded in the world?". In the introduction, the grounding problem was raised in the form of a dialogue between Socrates and a slave boy. At that time, I asked three questions:

(1) What is the knowledge shared by Socrates and the slave boy?
(2) What is the slave boy learning as a consequence of the exchange?
(3) Why are external representations needed in order to prove the Pythagorean theorem?

I believe that the *derivational approach* offers a systematic, formally describable answer to the grounding problem. Prepositions and the 'Jigsaw World' environment are not the same as concepts as the concepts encoded in geometry of triangles. In their case, the grounding problem is not of mapping concepts from one agent (Socrates) to another agent (The Slave Boy), but from abstract, 'concept-like' representations to concrete, 'world-like' representations. A corresponding version of these three questions for the mapping of abstract representations to concrete representations *has* been answered in this thesis, namely:

(4) Both concrete (world) as well as abstract (mind) representations share a global model. The global model generates a partial order in which both the abstract and the concrete are included.
(5) The abstract and concrete representations are mapped on to each other using derivations. What the solver of the puzzle (or the child learning the meaning of a preposition) learns are the dynamical regularities that are not contained within the goal state of the puzzle. These dynamical regularities are often path regularities, i.e., privileged ways of moving from the starting point to the end point of the derivation.
(6) The reason why external representations are crucial is because they are the source of dynamical regularities. It is this shift from a "static" to a "dynamic" perspective that allows us to quantify the information being learnt by the solver. In particular, the notion of scale plays an important role, which can be analyzed formally using the notion of multiple scale graphs.

Although, in this thesis I applied the derivational method to prepositions and jigsaw puzzles, the import of these results goes beyond these domains. A computational theory of grounding allows us to address some of the foundational questions for cognitive science. A few decades ago, John Searle in his famous arguments about Chinese rooms, (Searle, 1980) made the claim that purely syntactic procedure cannot possibly have meaning, since syntax does not give rise to semantics. One can respond to Searle's arguments in several ways, such as:

(1) While a syntactic procedure may not be grounded, a physical agent with causal interactions with the world would be grounded.
(2) While a syntactic procedure may not have semantics, a system containing the syntactic procedure might.
(3) Classical, symbolic systems would not have meaning, but meaning would emerge arise in large scale connectionist networks.

There are several other rejoinders to Searle, but they all have the problem that they do not address the issue of what the problem of meaning is in the first place. While Searle's Chinese room example shows us the inadequacy of our usual models of cognition, they do not present a computational formulation of the problem. In order to answer Searle's question, we need to ask ourselves "What is the computational structure of the relationship between Mind and the World?"

In my thesis, I have isolated three crucial aspects of the relationship between mental states and world states, namely, abstractness, groundedness and flexibility. The derivational method integrates these three constraints into a coherent and formalizable account of the relationship between syntax and semantics. Along the way, I showed that the same generative principles, based on categories, non-accidental features and partial orders, operate on both sides of the syntax-semantics divide, in two domains, prepositions and jigsaw puzzles. Furthermore, by generalizing the derivational framework to a more abstract setting, namely, that of generics and quantifiers, we see that quantifiers and generics has the same structures as prepositions and jigsaw puzzles, suggesting that the same computational mechanisms might be at play.

To conclude, the derivational method argues for a unified approach to the abstract, internal world and the concrete external world, and that concepts are spread out over this divide. In other words, to answer Searle's question, we have to study the common "forms" of syntax and semantics. There is no systematic reason to believe that there is an impermeable boundary between "mental" entities and "world" entities at the computational level.

## Appendix: Testing the Multiple Scale Hypothesis

My experimental goal was to test the multiple-scale hypothesis in the domain of puzzles. Since the tangram puzzles are relatively hard (it takes most people more than ten minutes to assemble the square), for experimental purposes, I replaced the tangrams by more easy puzzles consisting of rectangles subdivided into square pieces. In figure A.1 below, a simple example of such a puzzle along with its partial ordering is given. Figure A.2 shows an example of a puzzle actually used in the experiment.

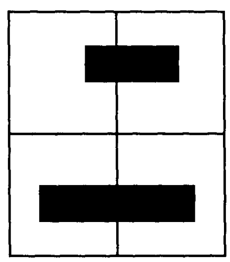


**Figure A.1 (a): The Puzzle**

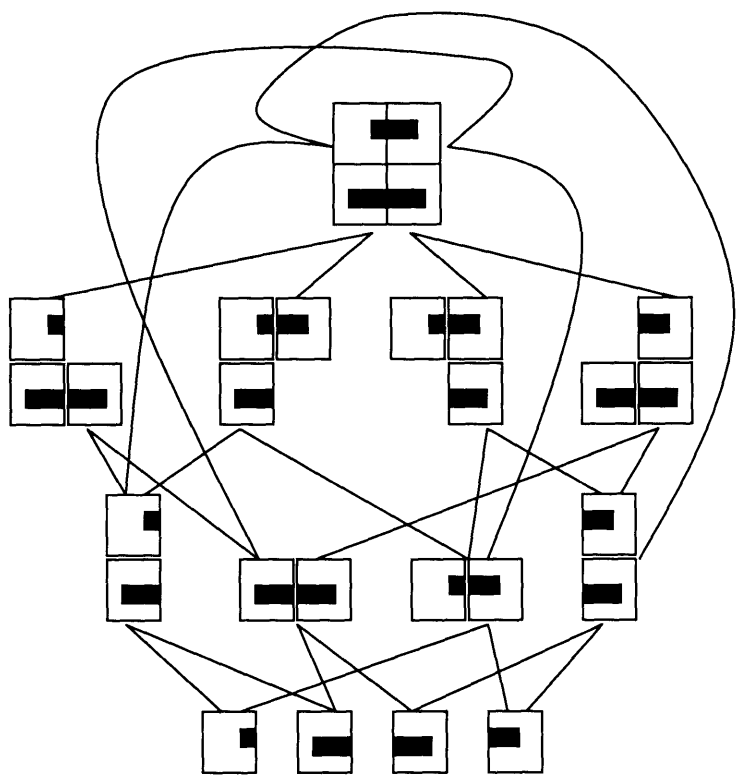


**Figure A.1 (b): The Partial Order**

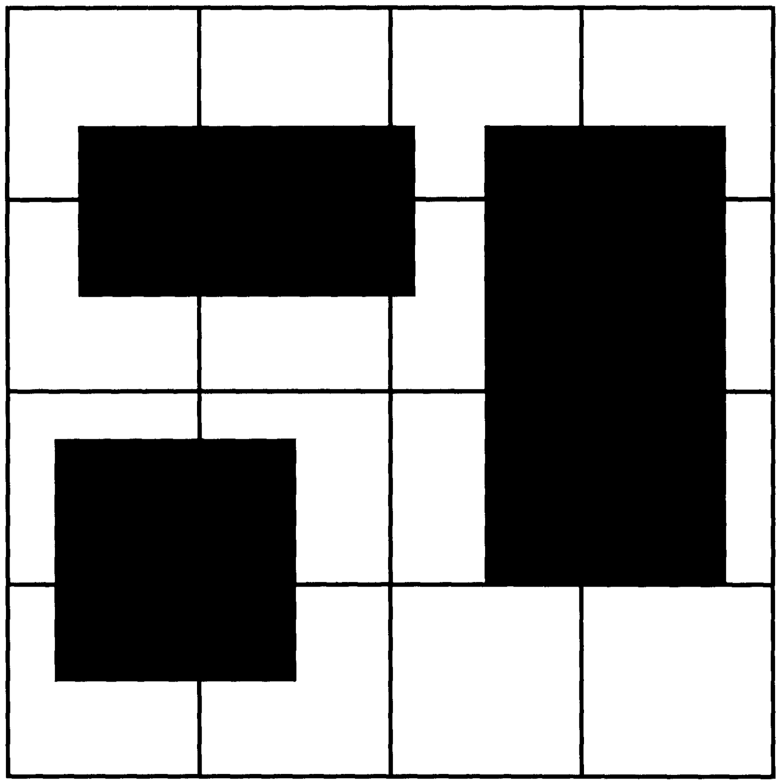


**Figure A.2: An example of a 4 × 4 puzzle used in the experiment**

Note the following constraints on the puzzles:

(1) All figures are black rectangles, at arbitrary positions with axes aligned with the sides of the puzzle.
(2) The puzzle pieces are all squares cut using vertical and horizontal cuts running across the puzzle
(3) The task, for a human (or for a computer) is to assemble the jigsaw puzzle, given some information about the puzzle, say, a statement like "the puzzle consists of three rectangles against a plain background (figure A.1)".

In the case of the 4 × 4 puzzles illustrated in figure A.2, solving the puzzle can be seen as a mapping problem whose goal is to ascend a partial ordering which is 15 levels deep, from the individual pieces (depth 15) to the puzzle as a whole (depth 0). The multiple scale hypothesis predicts that links between clusters of different depth will help solve the puzzle as well as making the solution more robust. I tested these hypotheses in an experiment involving relatively simple puzzles. The description of the experiments is given below.

**Task**: Each subject was given an individual 16 piece (4 by 4) jigsaw puzzle, like the one illustrated in figure A.2. The "figure" part of the puzzle consisted of three rectangles oriented either horizontally or vertically. The "ground" of the puzzle had four possible additional constraints (see figure A.3 below):

(a) The background was either shaded or unshaded.
(b) The edges of the puzzle were either left as is or they were covered with a black grid.

The subjects were told that the puzzle consisted of three vertically or horizontally oriented rectangles and that their job was to reconstruct the puzzle.

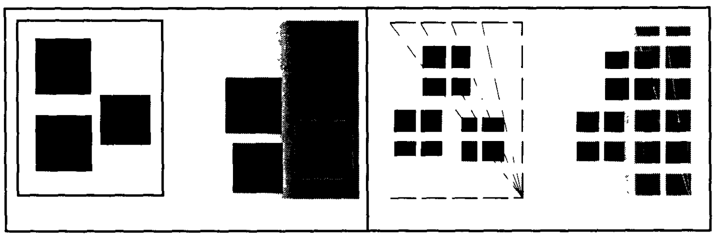


**Figure A.3**

**Predictions**: Given the multiple-scale hypothesis, we should make the following predictions:

(1) Shading , i.e., the addition of non-local cues, should make the puzzles easier significantly.
(2) Removal of local cues should make the unshaded puzzles much more difficult than the shaded ones.

According to the multiple scale graph model, the original puzzle (with three rectangles) should be hard without any shading, because the only clues that are available are local cues. On the other hand, when shading is added, the puzzle should become much easier, because coordinate frame information allows the solver to correlate pieces. In the third manipulation, the blackening of the edges should make the unshaded puzzle much harder. This follows because the blackening obscures the local connections, while in the shaded case, the effect of noise should not be as severe. To use the language of non-accidental features developed in chapter 3, shading should lead to the perception of non-local minimal maps (the bottom figure in figure A.4) while in the absence of shading, the solver can only use local minimal maps (top figure in figure A.4).

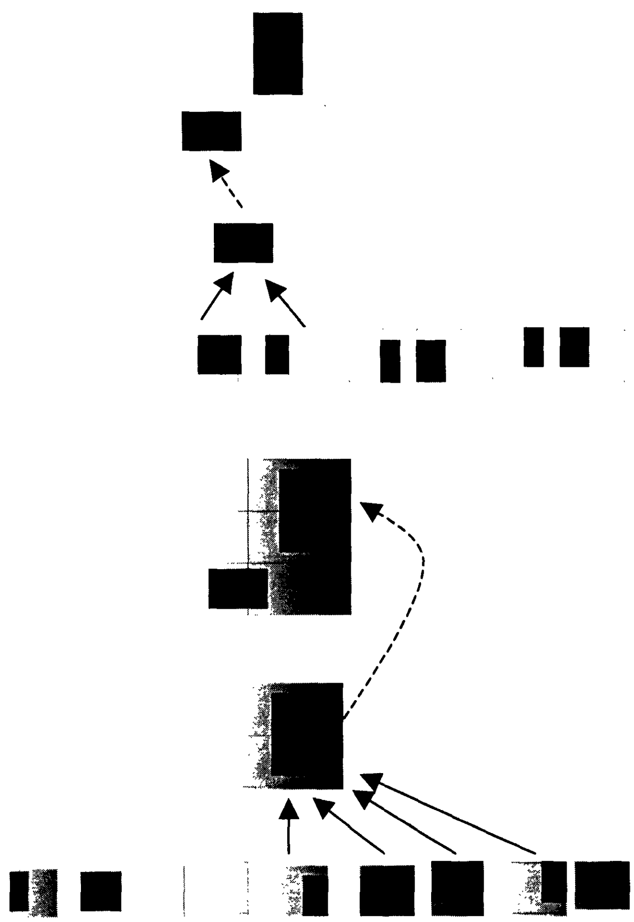


**Figure A.4. Local combinations in unshaded puzzles (top) contrasted with non-local combinations (bottom) in unshaded puzzles.**

**Test**: I tested these hypotheses on a group of 3 subjects. Each subject sat near a flat table upon which the puzzle pieces were arranged. The puzzles were made out of white cardboard upon which a puzzle pattern was printed and then cut up into 16 pieces. The subject was then given a randomly chosen puzzle out of the four possible choices. Each subject was timed as they did the four possible puzzles.

**Results**: The normalized results from the three subjects are shown in figure A.5 below. Normalization was done by dividing each subjects' timing by the lowest timing of the four possible puzzles.

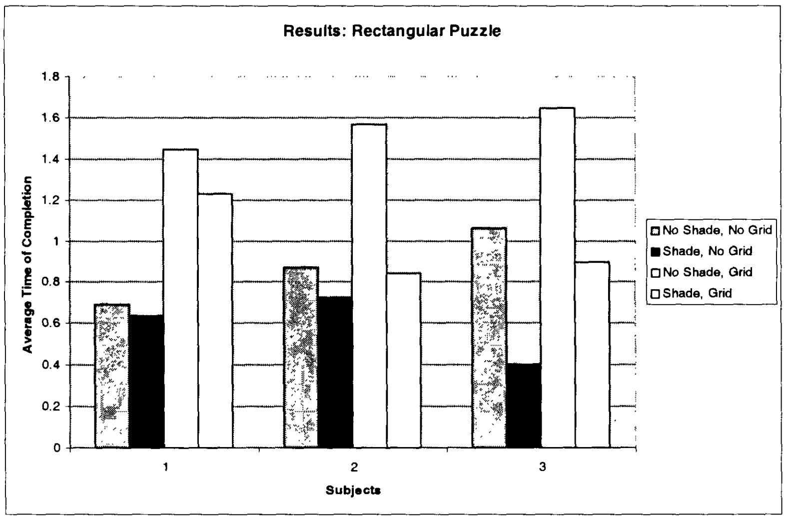



**Figure A.4**

As we can see from figure A.5, the results for the puzzles are consistent with the multiple scale model. While the addition of shading does not make a big difference to the puzzle timing, because the rectangles themselves provide enough non-local cues, the removal of edges does make a big difference, as one would expect. A further prediction would be that if the figure objects were amorphous shapes, the effect of shading and edge occlusion would be much greater.

## References


Agre, P. (1997). *Computation and Human Experience*. Cambridge, England: Cambridge University Press.

Ballard, D.H., Hayhoe, M.M., Pook, P.K., and Rao, P.N., (1996). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences.*

Beer, R.D., (2000). Dynamic approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), 2000.

Carey, S. (1985). *Conceptual Change in Childhood.* Cambridge, MA: MIT Press.

Carey, S. (1999). Sources of conceptual change. In E. K. Scholnick, K. Nelson, S. A. Gelman & P. Miller (Eds.), *Conceptual Development: Piaget's Legacy.* Hillsdale, NJ: Erlbaum, 293-326.

Carey, S. & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science, 63(4),* 515-533.

Carey, S. & Xu, F. (2001). Infants' knowledge of objects: beyond object files and object tracking. *Cognition* 80 (1-2): 179-213.

Carlson, G & Pelletier, F. (1995). *The Generic Book.* Chicago: University of Chicago Press.

Chomsky, N. (1957). *Syntactic Structures.* s-Gravenhage: Mouton.

Cohn, A. & Hazarika, S. (1996). Qualitative Spatial Representation and Reasoning: An Overview. *Fundamenta Informaticae*, 46.

Cooper, J., & Hutchinson, D. (Eds) (1997). *Plato: Complete Works.* Indianapolis: Hackett.

Coventry, K. & Garrod, S. (2002). Towards the Development of a Psychologically-Plausible Model for Spatial Language Comprehension Embodying Geometric and Functional Relations. To appear in L. A. Carlson & E. van der Zee (Eds.), Functional features in language and space: Insights from perception, categorization and development. Oxford: Oxford University Press.

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? *AI Magazine* 14:17-33.

Dennnett, D. (1998). *Brainchildren.* Cambridge, MA: MIT Press.

Fauconnier, G. & Turner M. (1998). Conceptual integration networks. *Cognitive Science* 22 (2): 133-187.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge, England: Cambridge University Press.

Fauconnier, G. & Turner, M. (2002). *The way we think*. New York: Basic Books.

Feldman, J. (2001). An Algebra of Human Concept Learning. *To Appear*.

Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology,* **41,** 145-170.

Fodor, J. A. (1975). *The Language of Thought.* Sussex: Harvester Press.

Gallistel, C.R. (1990). *Organization of Learning*. Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7:155-170.

Gentner, D., and A. B. Markman. (1997). Structure-mapping in analogy and similarity. *American Psychologist* 52:45-56.

Gibson, J.J. (1979). *The ecological approach to visual perception.* Boston, Houghton Mifflin.

Garrod, S., Ferrier, G. & Campbell, S. (1999). In and on: investigating the functional geometry of spatial prepositions. *Cognition.* 72 (2): 167-189 SEP 30 1999

Gopnik, A & Meltzoff, A. (1997). *Words, Thoughts and Theories*. Cambridge, MA: MIT Press.

Herskovits, A. (1985). Semantics and Pragmatics of Locative Expressions. *Cognitive Science*, 9 (3): 341-378.

Herskovits, A. (1986). *Language and Spatial Cognition*. Cambridge, England: Cambridge University Press.

Hoffman, D.D. & Richards, W. (1984). Parts of Recognition. *Cognition*, 18(1-3).

Jackendoff, R. (1991). Parts and boundaries. *Cognition* 41 (1-3): 9-45

Jackendoff, R. (1996). The proper treatment of measuring out, telicity, and perhaps even quantification in English. *Natural Language and Linguistic Theory* 14 (2): 305-354.

Jackendoff, R. (1996). The Architecture of the Linguistic-Spatial Interface. In P.Bloom et.al. (Eds.) *Language and Space*. Cambridge, MA: MIT Press.

Jepson, A. & Mann, R. (1999). Qualitative probabilities for image interpretation, Proceedings of IEEE International Conference on Computer Vision, Greece.

Jolicoeur, P. , Ullman, S. & Mackay, M. (1986). Curve Tracing - A Possible Basic Operations in the Perception of Spatial Relations. *Memory & Cognition* 14 (2): 129-140

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press; Cambridge, MA: Harvard University Press.

Kasturirangan, R. (1999). Multiple Scales in Small World Networks. AIM 1663, MIT AI-Laboratory Technical Report.

Keil, F. (1989). *Concepts, Kinds, and Cognitive Development.* Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, Fire and Dangerous Things.* Chicago: Chicago University Press.

Landau, B. & Jackendoff, R. (1993). What and where in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16(2): 217-238.

Lawvere, F. (1999). Kinship and Mathematical Categories in Jackendoff, R., Bloom, P. & Wynn, K. (eds.). *Language, Logic, and Concepts. Essays in Memory of John Macnamara*. Cambridge, MA: MIT Press.

Lawvere, F. & Schanuel, S. (2002). *Conceptual Mathematics: A First Introduction to Categories*, Cambridge: Cambridge University Press.

Le Palma Reyes, M., Macnamara, J., Reyes, G. & Zolfaghari, H. (1999). Count Nouns, Mass Nouns, and Their Transformations: A Unified Category-Theoretic Semantics In Jackendoff, R., Bloom, P. & Wynn, K. (eds.). *Language, Logic, and Concepts. Essays in Memory of John Macnamara.* Cambridge, MA: MIT Press.

Lettvin, J.Y., Maturana, H.R., McCollough, W.S., & Pitts, W.H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers*, 47, 1940-1951.

Link, G. (1998). *Algebraic semantics in language and philosophy.* Stanford, Calif. : CSLI Publications.

Lowe, D. (1985). *Perceptual organization and visual recognition.* Boston: Kluwer Academic Publishers.

Maclane, S. (1998). *Categories for the Working Mathematician.* New York, NY: Springer.

Macnamara, J.: 1999, *Through the Rearview Mirror. Historical Reflections on Psychology*, Cambridge: The MIT Press

Margolis, E. & Laurence, S. (eds)(1999). *Concepts: Core readings.* Cambridge, MA: MIT Press.

Marr,D.(1982). Vision. San Fransisco: W.H.Freeman.

Miller, G. & Johnson-Laird, P. (1976). *Language and perception.* Cambridge, MA: Harvard University Press

Minsky, M. (1974). A Framework for Representing Knowledge. AI memo 306, MIT AI Lab.

Minsky, M. (1986). *The Society of Mind.* New York, NY: Simon and Schuster.

Nagpal, R., (2001). Programmable Self-Assembly: Constructing Global Shape Using Biologically-Inspired Local Interactions and Origami Mathematics, PhD Thesis, *MIT Artificial Intelligence Laboratory Technical Memo 2001-008.*

Newcombe, N.S. & Huttenlocher, J., (2000). *Making Space.* Cambridge, MA: MIT Press.

O'Regan, J. K. (1992). Solving the real mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology*, 46, 461-488.

Poston,T. & Stewart, I. (1978). Catastrophe theory and its applications. London: Pitman Publishing Limited.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, 4, 66-72.

Putnam, H. (1975). *Mind, Language and Reality: Philosophical Papers,* vol. 2. Cambridge: Cambridge University Press.

Pylyshyn, Z. (2001). Visual Indexes, preconceptual objects and situated vision. *Cognition*, 80, 127-158.

Regier, T. (1996). *The Human Semantic Potential.* Cambridge, MA: MIT Press.

Regier, T. & Carlson, L. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology – General*, 130 (2): 273-298

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, 8 (5), 368-373.

Richards, W. (ed) (1988). *Natural Computation*. Cambridge, MA: MIT Press.

Richards, W. Jepson, A. & Feldman, J. (1996) Priors, preferences, and categorical percepts. In D. Knill & W. Richards (eds.) Perception as Bayesian Inference. Cambridge, England: Cambridge University Press.

Rosch, E., C. B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. (1976). Basic objects in natural categories. *Cognitive Psychology* 8:573-605.

Scholl, B. (2001). Objects and attention: the state of the art. *Cognition*, 80, 1-46.

Schank, R. C., and R. P. Abelson. (1977). *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Erlbaum.

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3, 417-424.

Siskind, J. (1995). Grounding Language in Perception. *Artificial Intelligence Review*, volume 8, pp. 371-391.

Siskind, J. (1991). Naive Physics, Event Perception, Lexical Semantics and Language Acquisition, in Proceedings of the AAAI Spring Symposium Workshop on Machine Learning of Natural Language and Ontology, pp. 165-168.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science* 12(1):49-100.

Talmy, L. (2000). *Towards a Cognitive Semantics*. Cambridge, MA: MIT Press.

Thelen, E., and L. B. Smith. (1993). *A Dynamics Systems Approach to the Development of Cognition and Action.* Cambridge, MA: MIT Press.

Treisman, A. (1993). The perception of features and objects. In A. Baddeley, & L. Weiskrantz (Eds.), *Attention: Selection, Awareness and Control*. Oxford: Oxford University Press.

Ullman, S. (1996). *High Level Vision*. Cambridge, MA: MIT Press.

Varela, F., Thompson, E. & Rosch, E. (1991) *The Embodied Mind.* Cambridge, MA: MIT Press.

Watanabe, S. (1985) *Pattern recognition: human and mechanical.* New York : Wiley.

Witkin, A. & Tenenbaum, J. (1985).On perceptual organization. In A. Pentland (ed), *From Pixels to Predicates.* Norwood, NJ: Ablex.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object Æles: shapeless bundles of basic features. *Vision Research*, 37 (1), 25-43.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*,15 (3), 419-433.