The Cognitive Map:
Could It Have Been Any Other Way?*

Benjamin Kuipers

1983

Abstract

Could the human cognitive map, with all of its peculiarities, be structured in any other way and still perform the useful functions it does? We can perform a thought-experiment by imagining that we must design the cognitive map for a robot, operating under limited cognitive resources, which must assimilate and use knowledge about its large-scale environment acquired from observations during travel. By taking this design perspective, we can determine that some familiar proposals for the structure of the cognitive map are inadequate, by themselves, to meet the constraints of the task. We develop a design consisting of separate representations for relative-position information, topological connections, and knowledge of routes, each rich in states of partial knowledge. Each step of the derivation is motivated by the pragmatic needs of the task, and the result bears a strong resemblance to the human cognitive map. We also discuss the relationship between this type of argument from design and the more usual approach of psychological explanation.

The Cognitive Map

When we observe the peculiarities of the human cognitive map, we are tempted to ask, "Could it have been any other way?" We can conduct a thought-experiment to study this question, by viewing it from the point of view of the computer scientist and system designer rather than that of the psychologist.

---

*In H. L. Pick, Jr. and L. P. Acredolo (Eds.), *Spatial Orientation: Theory, Research, and Application.* New York: Plenum Press, 1983, pages 345–359.

The structure of this argument from design is a familiar one to the engineer, computer programmer, or system designer, but may appear strange to the psychologist, accustomed to drawing heavily on empirical observations. The designer begins by specifying the task to be performed and the constraints that a solution must satisfy. He then proposes the simplest mechanism that he believes can achieve the desired level of performance, and observes it in (perhaps simulated) action. Failures to accomplish the task within the constraints motivate extensions to the proposed mechanism, repeating this process until the design problem is solved.

In this essay, we will begin with a simple "straw man" proposal for the cognitive map, and by progressively revealing its inadequacies and refining its design, we will arrive at a robust and somewhat more complex model. The traditional process of empirical science can answer certain "What" and "How" questions by confirming hypotheses about the observable behavior of the world. A design approach, by contrast, can answer certain "Why" questions by showing how task and performance constraints motivate certain design decisions in the construction of any mechanism to achieve the given behavior.

By the *cognitive map* we mean that body of knowledge of a large-scale environment that is acquired by integrating observations gathered over time, and is used to find routes and determine the relative positions of places. A *large-scale* environment is one whose structure is revealed by integrating local observations over time, rather than being perceived from a single vantage point. Notice that a graphical map of a city uses the (quite different) cognitive mechanisms of visual space to convey information about the large-scale space.

Let us imagine, for the purposes of this essay, that we are part of a large team collaborating on the design of an intelligent robot. Other members of the team are responsible for providing the robot with perceptual and motor abilities. Our task is to design the cognitive mechanism by which it acquires and stores knowledge of its large-scale environment. and the processes by which it finds routes and determines relative positions.[1] We assume that our robot will travel in an urban environment of streets and intersections, though the design carries over to less structured environments. We take a computational approach to solving this problem because only computational

---

[1] Becker and Merriam [Merriam, 1977], under a contract from NASA, actually attempted to design the cognitive structure for a robot travelling in a simulated "Martian" environment. Most of their effort, however, was devoted to the environmental simulation, and to their design for the robot's sensory, attentional, and motor sub-systems.

processes provide the expressive power to achieve very complex behavior while being grounded in a working device. This parallels the psychological assumption that the mind can be described, at some level of detail, as a computational process.

The cognitive map must operate under considerable resource limitations, since the robot must be simultaneously planning its activities, solving problems, and dealing with interruptions and hazards during its travels (Kuipers, 1979a). The cognitive map must be constructed using the computational resources left over after more pressing needs are satisfied. It will turn out that many important design decisions will be forced by the constraints of learning from a sequence of isolated observations while under serious resource limitations.

Almost by definition, the cognitive map must contain information about the relative positions of places in two dimensions. Our first approach, then, might be to endow our robot with a two-dimensional structure (a "Map in the Head") intended to be an analog of the environment, and give it the strategy of drawing new locations into the map as it travelled, along with connections between known places. Over time, we hope, the robot will add more places to its store of knowledge, drawing a progressively more complete map of its environment, using observations gathered during travel.

That is, we can define the 2-D analog as a data structure with certain access functions: the ability to insert a place into the structure with a given location, and the ability to compare two locations to determine the relative-position vector between them. This 2-D analog could conceivably be implemented by a physical analog, such as an internal blackboard with a TV camera serving as the "mind's eye." More plausibly, it could be implemented as a list of coordinates associated with known places, along with trigonometric computations to discover relative-position vectors when needed. There are other possible implementations, each *functionally equivalent* to the "Map in the Head," in that the access functions are the same. Shepard and Chipman (1970) call this a "second-order isomorphism."

The 2-D analog "Map in the Head" has a single frame of reference for the coordinates of locations and for the orientation of relative-position vectors, so any place drawn onto the map has a relative position defined with respect to any other place on the map. This is a faithful reflection of the real world, but it is not a useful feature in a cognitive representation designed for learning the structure of space from observations. An observer's knowledge of the world is necessarily partial.

If, for example, our robot were to travel from a familiar area to another area it then learned in some detail, interruptions, the limits of its perceptual abilities, and other cognitive resource limitations (Norman & Bobrow, 1975) could prevent it from maintaining its orientation accurately between the two areas. Limited to a 2-D analog representation, the robot is compelled to draw the two areas in some well-defined though probably incorrect relation to each other.

This has several problems for the unfortunate robot. First, such errors can propagate to become very serious, resulting in the same place drawn in two positions on the map. Second, with only a single 2-D analog structure, if an error is detected there is no "scratch space" for working out possible corrections without destructive modifications to the original. Third, if the correct orientation were learned between the two areas, it could require large amounts of processing to correct the analog map place by place. Furthermore, during the correction process, the map could be left in an inconsistent state.

When we look at the human cognitive map as an example of good design, we see that people frequently lack any belief, right or wrong, about the relative orientation of remote regions. While people occasionally do jump to premature conclusions about relative orientation, causing distortions in their cognitive maps, this is not their only mode of behavior, as it must be if the single 2-D analog representation is used. On the other hand, we do find many of these problems when people are asked to draw sketch-maps (i.e., 2-D analogs of the environment) from their cognitive maps (Beck & Wood, 1976; Lynch, 1960). This argument shows that the sketch-map representation cannot *by itself* serve as an adequate representation for the cognitive map.

We conclude that the 2-D analog is not sufficiently powerful to serve as the cognitive map because it has too few *states of partial knowledge.* That is, an observation often cannot be stored without making strong and perhaps unwarranted assumptions about the global structure of space. These assumptions are required by the single frame of reference implicit in the 2-D analog map. We must return to the drawing board to design a cognitive map that is able to store spatial relationships between certain places while tolerating ignorance of the relationships between others.

The only solution to this problem is to allow multiple frames of reference for storing locations and relative-position vectors. Locations and vectors within the same frame of reference can be compared easily; those within different frames of reference can be compared only with difficulty, if at all. A place may be represented in more than one frame of reference, which makes it possible for a problem-solving method to deduce the relationship between two of these frames. Metaphorically, we have replaced the "Map in the Head" with an "Atlas in the Head," consisting of many maps on separate sheets.

We can determine what types of maps should appear in the "Atlas" by anticipating the types of partial knowledge of relative position the robot may need to represent. As our robot explores a particular region, and has sufficient processing resources available, it can build up a detailed map of a particular area. Additional relative-position vectors can be acquired from observations within the same frame of reference. The robot may later travel to another region or under conditions of more limited resources for observation or processing (e.g., by air), and learn the second region in a disconnected frame of reference. The relation between two regions might well be known to less precision, and with respect to yet another frame of reference distinct from those within either region. Thus, we must give our robot the capability to construct frames of reference (or maps in the Atlas) at different scales, so that a large-scale map might represent the relative positions of regions whose detailed structures are represented on maps of their own (Kuipers, 1978; McDermott, 1980).

In order for our design to accommodate many states of partial knowledge of position, we have provided the possibility for a highly fragmented, disconnected cognitive map. Indeed, we observe such cognitive maps in humans under some circumstances (Appleyard, 1970; Lynch, 1960). However, the continual tendency of the learning process is to create links among places and regions, making it easier to infer the relations between different frames of reference, knitting the separate maps of the "Atlas" into a more unified structure. We must look more closely at our design, however, to determine the fabric of that structure.

Topological Relations

The cognitive map, as we have designed it so far, represents only metrical relations between places: relative-position vectors. It does not indicate whether there is a path allowing direct travel between two places. By contrast, topological relations such as connection and order play an important role in the human cognitive map (Lynch, 1960), so we wonder whether they are necessary components of the robot's cognitive map. In the following city-block diagram, we would certainly expect our robot's

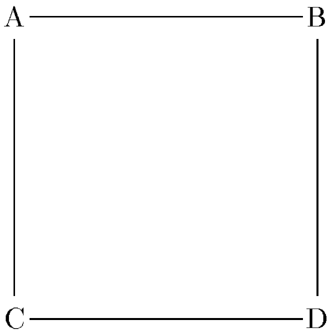

Figure 1: Even though the relative positions of the diagonals may be known in the cognitive map, their lack of "travellabiIity" must be stored.

cognitive map to be able to store all six relative-position vectors between pairs of places. However, it must also store information that distinguishes the four travellable paths around the periphery from the two diagonals, whose relative-position vectors are known but which do not correspond to travellable paths. This new feature of the relation between two places—the travellability of a connection—is our first glimpse of a new type of spatial relation in the cognitive map: the purely topological assertion of connectivity.

In the case of the city block, we can imagine a "travellability" indicator attached to each stored relative-position vector. However, in the limiting case, it may be necessary to indicate the existence of a travellable connection between two places without any relative position information at all. For example, if our robot began a particular leg of its journey in a disoriented condition (as we assume it might due to resource limitations), when it reaches the next recognizable place it creates a relative-position vector whose orientation is completely arbitrary! Since the new vector is the only one in a new frame of reference, there is no reason to assign it one direction rather than any other. If additional resource limitations have kept the distance, as well, from being observed or stored, then the information about the path between these two endpoints is purely a topological assertion about connectivity.

The topological and metrical relations must be considered as two distinct types of spatial knowledge, represented separately in the cognitive map, since cases can easily arise where two places are related by one or the other type of relation, or both, or neither. Metaphorically, we can think of the metrical knowledge of relative position functioning as multiple 2-D analog maps, while the topological connections function as flexible strings joining related places.

We are forced to represent topological relations separately from metrical ones by the need to store a type of partial knowledge that the cognitive map could encounter under conditions of resource limitations. However, we find that this design then pays off handsomely in increased problem-solving power. Route-finding is greatly simplified by representing the environment as a network consisting of zero-dimensional places connected by one-dimensional paths of travel. Route-finding based solely on relative-position knowledge is vulnerable to blind alleys and dead ends, whereas a network describing how places are connected by paths can be searched by simple, powerful algorithms to find routes.

As the number of places and paths in our robot's cognitive map becomes very large, however, simple graph-search algorithms become inadequate. The time required to search a large graph to find a route can be intolerably large. Fortunately, having a design for the cognitive map with two types of knowledge comes to our rescue here. Where the map contains information about the relative positions of the endpoints of the desired route, it is easy to add a heuristic to the graph-searching algorithm to look first at paths going in approximately the desired direction.

While the direction-oriented heuristic provides some improvement to the route-finding process, much more can be obtained by grouping places into regions. Here again, we need not add a new feature to the cognitive map, but only take advantage of one that was already present for a different reason. We were previously compelled to organize the metrical knowledge into multiple frames of reference, even allowing frames of reference at different scales, so that what appeared as regions full of places from one point of view could appear as a single place from another. By treating these frames of reference as regions organized into a hierarchy by containment, it is not difficult to devise a route-finding algorithm that operates first at an abstract level, finding connections among regions, and then fills in the details at the lower levels. This is a powerful technique for reducing search in large spaces.

Since this design for the cognitive map is very rich in states of partial knowledge, observations can be assimilated in a way that is very tolerant of resource limitations. When observations can be made carefully and processing resources are plentiful, the topological and metrical maps can be created together. When the available information or resources are less bountiful, the topological relations alone can be stored and used for route-finding. As further resources become available, the topological connections can be strengthened into relative-position vectors by assimilating observations of distance and of orientation with respect to the current frame of reference. By the same process, topologically connected regions with well-specified but disjoint metrical maps can be knit together by subsequently-learned relative-position information. If a connection between two familiar areas is learned incorrectly, better information acquired subsequently can correct the problem by changing the relationship between the two area maps, not by modifying the stored location of every known place.

Although this is a very good design for our robot's cognitive map, and illustrates the design requirements underlying several features of the human cognitive map (cf. Kuipers, 1978, 1979b), it is not yet sufficiently well-specified. We have implicitly assumed that our creature is somehow provided with knowledge of *places*, and that observations provide relations between them. This is not the case. Rather, the robot's sensory world consists of a temporal sequence of sensory images (*views*) and motor operations (*actions*). To be complete, our design must specify how places, and the relations among them, can be obtained from this kind of sensory input.

Learning Routes From Observations

In designing a cognitive map for our robot, we assume that its perceptual and motor abilities have already been provided. In particular, we assume that its perceptual system produces datastructures we call *Views* that describe the sensory image obtained from a particular viewpoint. Without specifying their detailed structure, we assume that two Views can be matched quite reliably (though perhaps not perfectly) to determine whether they represent the same place, and that Views can be used as keys for retrieval of information from database storage. A View need not be visual. The robot could distinguish places by radar or by sense of hearing without affecting the underlying structure of the cognitive map. All we require is that different places, and different orientations at a given place, be distinguishable (most of the time) by the sensory information available there. We will assume that motor operations are described by datastructures we call *Actions*, which are of two types: a *rotate* records the angle of rotation in a turn, and a *travel* records the distance travelled. An Action may be used to record an operation taking place or to command the motor system to perform it.

The cognitive map must include two processes to bridge the gap between passive observation of the temporal succession of sensory images and active navigation using a description of the fixed environment. The first process constructs descriptions of routes travelled so that they can be followed without guidance, or mentally reviewed in the absence of the environment. The second must construct descriptions of the fixed features of the environment (places and paths) from the succession of Views and Actions.

There are two capabilities we would expect to correspond with knowing a route: being able to travel the route without guidance, and being able to describe or review the route without physically moving along it. These are different capabilities, since we occasionally experience the anomalous state of being able to travel a route, but not describe it without travelling it. Travelling the route must include at a minimum, knowing which Action to take at any given moment, that is, when faced with a particular View. Thus, knowledge of a particular route must include a set of associations (View → Action) between descriptions of sensory images and the corresponding actions to take. An action like *travel* would terminate at the next decision point, when the current View should activate an association providing the next Action. The ability to rehearse the route in the absence of the environment requires, in addition, a different set of associations (View → Next-View) from the current View to the View resulting at the next decision point.

We can represent both associations in a three-part schema for a step of the route:

$$(\text{Context: } \langle \mathit{View} \rangle;\ \text{Action: } \langle \mathit{Action} \rangle;\ \text{Result: } \langle \mathit{View} \rangle)$$

The route description consists of a set of these schemas, each indexed for retrieval under the View in its Context part. As a new route is learned from experience, new schemas are created, and their Action and Result components are filled in. A route description consisting of a set of partially filled-in schemas constitutes partial knowledge of the route, and has some interesting properties (Kuipers, 1979b). When only the Context parts of the schemas are filled, the route-description supports recognition of landmarks, but not self-guided travel; when the Action parts are also present, the route description supports self-guided travel but not mental rehearsal of the route apart from the environment; finally when all three components are present, knowledge of the route is complete. These states of partial knowledge allow the route to be learned incrementally from observations when processing resources are scarce.

The other major process at the foundation of the cognitive map is the creation of descriptions of the fixed features of the environment—places and paths—given the temporal sequence of sensory images experienced during travel. We must augment the cognitive map to include a "place" description that has explicit associations with the Views obtainable at that place, and is the object of the topological and metrical relations. We then ask how the many Views available in our robot's sensory world come to be grouped according to place. At a minimum, we expect that two views will be associated with the same place description if they are linked by an Action consisting of a *rotate* with no *travel.* In fact, this relation is an equivalence relation[2] on the set of Views, and the equivalence classes turn out to correspond exactly with the places in the environment. Furthermore, a simple incremental algorithm can link Views to place descriptions as they are encountered during travel.

Precisely the same technique can be used to define path descriptions as equivalence classes of Views joined by *travel* without *rotate.* A path, by this definition, is more than an arc in a network. Rather it collects the Views occurring along a street. Since those Views may also correspond to places, a topological connection between a place and a path is made whenever the current View has associations to both a place and a path description. The spatial order of places along a path can be obtained, incrementally, from the temporal order of Views during travel.

Since neither of these two processes depends on the result of the other, they can proceed independently, constructing knowledge of routes and of places and paths out of the temporal sequence of sensorimotor images that constitute the robot's experience during travel. There are links between the two processes, however. For example, when a route description consists of Views that have not yet been aggregated into place descriptions, the route is irreversible, because there is no chain of associations from the View facing in one direction to the View in the reverse direction.

As before, we have selected these knowledge representations by taking the minimal features required to provide a certain capability, and encoding them to allow the most possible states of partial knowledge, so as to support incremental assimilation of observations under resource limitations. The result turns out to bear a strong resemblance to the human cognitive map.

---

[2] An *equivalence relation* is a reflexive, symmetric, transitive relation on a set of objects. It partitions the set into disjoint subsets, called *equivalence classes*, of objects that are mutually equivalent under that relation.

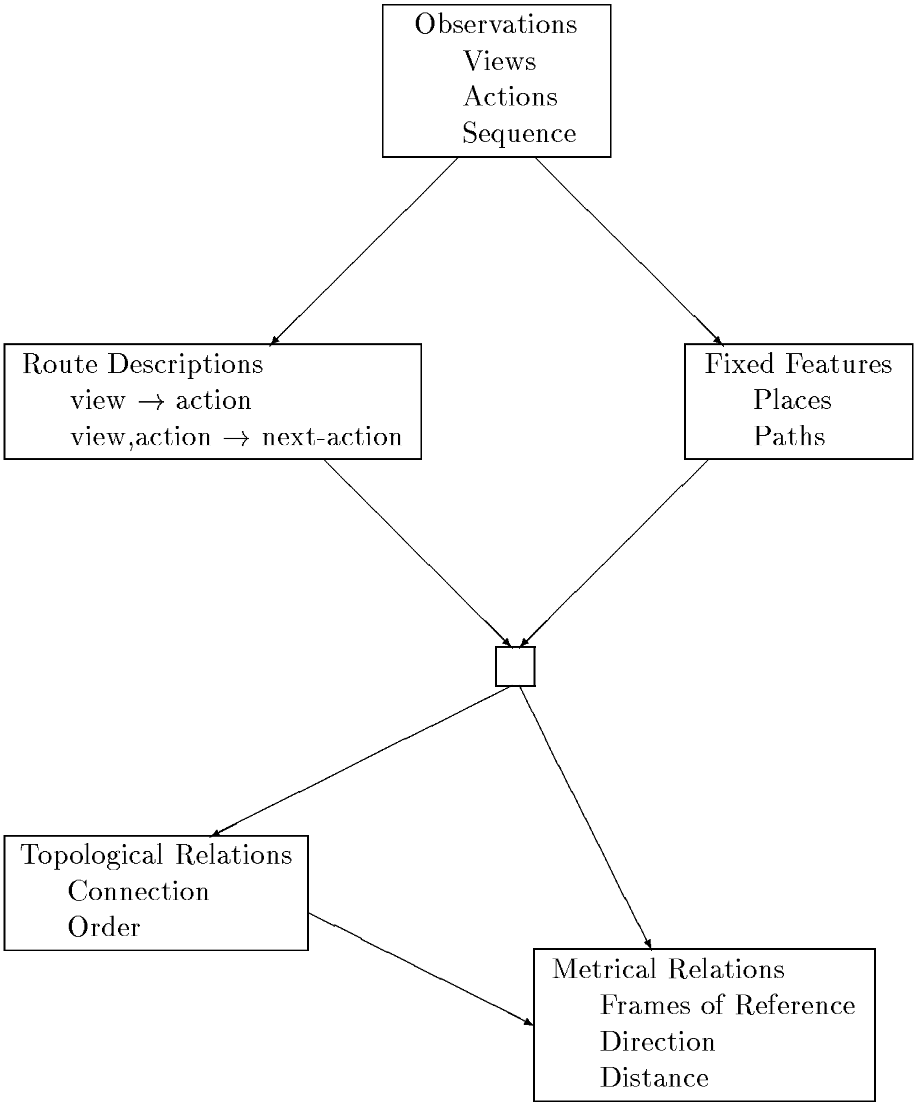

Figure 2: The cognitive map consists of five different types of information, each with its own representation. The arrows indicate how each representation depends on the prior acquisition of knowledge of a different type. The knowledge in the upper half of the diagram could be called the "route map," while the lower half could correspond to the "survey map."

Why Design The Cognitive Map?

We look at the peculiarities of the human cognitive map and wonder "Could it have been some other way?" Tentatively, our answer is No. The constraints of learning a large-scale environment from local observations while operating under interruptions and resource limitations are strong enough that any process that achieves the required level of performance must resemble the human cognitive map, at least at a gross level.

The framework of the design process lets us make an argument that cannot be made using psychological methods. by attempting to design the cognitive map to provide certain capabilities, we can infer the reasons *why* a knowledge representation is the way it is.

1. Metrical knowledge is represented with respect to multiple frames of reference *because* a single frame of reference would require strong and often erroneous assumptions to be made during the assimilation of observations.

2. Topological relations are represented separately from metrical relations *because* observations under resource limitations may capture the fact of a connection between two places without the distance or direction of their relative positions or orientations. Furthermore, route-finding is much more efficient based on a combination of topological and metrical information than on metrical information alone.

3. Knowledge of routes has its own representation *because* it is valuable for an organism to be able to follow an experienced route before creating a coordinated representation of the fixed features of the environment, or even before being able to rehearse the route without travelling it.

4. There is a separate representation for sensory images *because* observational input consists of a temporal sequence of images rather than direct perception of fixed features of the environment. Places, paths, and their relationships must be inferred from the sequence of images.

The key to this argument is the need for the cognitive map to provide adequate performance under resource limitations of various kinds: interruptions, failure or imprecision of perceptual input, and lack of processing resources. Adequate performance can be achieved if the knowledge representations in the cognitive map allow incremental assimilation of observations as resources are available: that is, if the cognitive map provides many states of partial knowledge. The need to provide more states of partial knowledge motivates most of our choices of particular knowledge representations over their alternatives.

This is the approach of artificial intelligence research: given a type of intelligent behavior, design a solution, or a range of solutions, that exhibit that behavior under given constraints. Potential solutions are tested for *sufficiency* to satisfy the constraints and accomplish the task. A design is evaluated solely according to its ability to accomplish the intended task within the constraints of the environment and cognitive resource limits. The existence and behavior of the human cognitive map provides inspiration for possible designs, but does not determine the choices, since it might represent one of many possible solutions to the same problem.

The nature of the argument from design is almost entirely different from the argument from empirical evidence in descriptive psychology. The interface with descriptive psychology comes when there are several ways to accomplish the given task. A design problem may have several distinct solutions. The alternate designs can then be considered as alternate hypotheses by the experimental psychologist, who focuses on the critical differences that make it possible to determine which design is used by humans. Our conclusion in this essay is that such design alternatives *do not exist* at the level of the gross structure of the cognitive map. While there may be alternate ways of representing each type of spatial knowledge—of relative position, of topological order, or of the relations between sensory image and place—any theory of the cognitive map sufficient to provide adequate performance must accommodate all three types of knowledge and the relations among them. The internal structure of these three knowledge representations in the cognitive map is a fruitful area for future research.

References

Appleyard, D. Styles and methods of structuring a city. *Environment and Behavior*, 1970, 2, 100-117.

Beck, R., & Wood, D. Comparative developmental analysis of individual and aggregated cognitive maps of London. In G. T. Moore & R. G. Golledge (Eds.), *Environmental knowing: Theories, research, and methods.* Stroudsburg, PA: Dowden, Hutchinson and Ross, 1976.

Kuipers, B. J. Modelling spatial knowledge. *Cognitive Science*, 1978, 2, 129-153.

Kuipers, B. J. On representing commonsense knowledge. In N. V. Findler (Ed.), *Associative networks: The representation of use of knowledge by computers.* New York: Academic Press, 1979. (a)

Kuipers, B. J. Commonsense knowledge of space: Learning from experience. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence.* Stanford, CA: Stanford Computer Science Department, 1979. (b)

Lynch, K. *The Image of the City.* Cambridge, MA: MIT Press, 1960.

McDermott, D. V. Spatial inferences with ground, metric formulas on simple objects. Department of Computer Science Research Report 173, Yale University, 1980.

Merriam, E. W. "ROBOT" Computer Program Solving System. Cambridge, MA: Bolt Beranek and Newman, Inc. Report No. 3516. January 18, 1977.

Norman, D. A., & Bobrow, D. G. On data-limited and resource limited processes. *Cognitive Psychology,* 1975, 7, 44-64.

Shepard, R. N., & Chipman, S. Second-order isomorphism internal representations: Shapes of states. *Cognitive Psychology,* 1970, 1, 1-17.