BEHAVIORAL AND BRAIN SCIENCES (1993) 16, 217–265
Printed in the United States of America

# "What" and "where" in spatial language and spatial cognition

**Barbara Landau**
*Department of Cognitive Science, University of California, Irvine, CA 92717*
**Electronic mail:** *blandau@orion.uci.edu*

**Ray Jackendoff**
*Linguistics Program, Brandeis University, Waltham, MA 02254*
**Electronic mail:** *jackendoff@brandeis.bitnet*

**Abstract:** Fundamental to spatial knowledge in all species are the representations underlying object recognition, object search, and navigation through space. But what sets humans apart from other species is our ability to express spatial experience through language. This target article explores the language of *objects* and *places*, asking what geometric properties are preserved in the representations underlying object nouns and spatial prepositions in English. Evidence from these two aspects of language suggests there are significant differences in the geometric richness with which objects and places are encoded. When an object is named (i.e., with count nouns), detailed geometric properties – principally the object's shape (axes, solid and hollow volumes, surfaces, and parts) – are represented. In contrast, when an object plays the role of either "figure" (located object) or "ground" (reference object) in a locational expression, only very coarse geometric object properties are represented, primarily the main axes. In addition, the spatial functions encoded by spatial prepositions tend to be nonmetric and relatively coarse, for example, "containment," "contact," "relative distance," and "relative direction." These properties are representative of other languages as well. The striking differences in the way language encodes objects versus places lead us to suggest two explanations: First, there is a tendency for languages to level out geometric detail from both object and place representations. Second, a nonlinguistic disparity between the representations of "what" and "where" underlies how language represents objects and places. The language of objects and places converges with and enriches our understanding of corresponding spatial representations.

**Keywords:** count nouns; identification; language; localization; object recognition; object structure; prepositions; spatial language; spatial representation; what/where system

The representations underlying object recognition, object search, and navigation through space are fundamental to spatial knowledge in all species. What sets humans apart from other species is our ability to use these representations to express our spatial experience, talking about what things are and where they are located. Clearly, language and spatial understanding map onto each other. In this target article we ask what that mapping might be – how language draws on our spatial representations such that we can manage to talk about what we perceive.

Our focus will be the language of objects and places in English. Our specific goal will be to describe these linguistic domains in a way that is compatible with constraints on nonlinguistic spatial cognition. At the same time, however, we will use evidence from language to provide boundary conditions on a satisfactory theory of spatial cognition. Our premise is that any aspect of space that can be expressed in language must also be present in nonlinguistic spatial representations. Simply put, whatever we can talk about we can also represent.[1]

In the main tradition of research on spatial cognition, psychological and neuroscientific techniques have been used to study our ability to visually perceive objects and their locations and motions in space. At the same time, there has been a substantial tradition in linguistics and psycholinguistics of studying the spatial expressions in human language. Such analyses have been important not only because of the richness and complexity of spatial language itself, but also because the organization of spatial language extends readily to many abstract domains such as time, status, possession, and social organization (Gruber 1976; Jackendoff 1976; Lakoff & Turner 1980). The burden of this target article is to show that the latter concerns can be brought to bear as a new source of evidence on the nature of human spatial cognition. Spatial language, properly analyzed, can shed light on spatial thinking.

We should begin by clarifying what we mean by spatial representation and what we mean by the language of objects and places. By spatial representation, we intend a level of mental representation devoted to encoding the geometric properties of objects in the world and the spatial relationships among them. Because spatial information can be derived from vision, audition, and the haptic (touch) faculty, the format must be either heteromodal or amodal. That is, this representation is not exclusively visual or haptic or aural, but *spatial;* understanding spatial configurations in the world involves either translating modality-specific information into a common format or providing interfaces between modalities. We also assume that spatial representations must be translatable into a form of representation specific to the

© 1993 Cambridge University Press 0140-525X/93 $5.00+.00

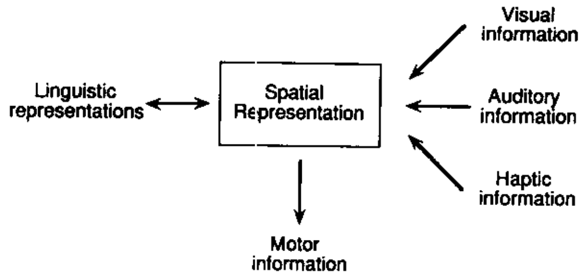



Figure 1. Spatial representations take as input information from vision, audition, and the haptic system, and provide information to the motor system and language.

motor system, used to initiate and guide behavior: We can touch what we see, look at what we hear, and avoid obstacles as we navigate through space[2] (see Figure 1).

To account for language about space, there must be a translation between spatial representations and language. For our purposes, the language of space will concern those words and simple phrases that encode objects and places. In English, objects are represented by count nouns and places are represented canonically by prepositions or prepositional phrases (see below for some qualifications). Although people can obviously construct syntactically complex descriptions of objects (e.g., "the thing over there with the two pointed tops") or places (e.g., "not less than 3,000 miles away"), we will restrict our discussion to count nouns and prepositions because these single words are likely to correspond to the simplest well-formed formulas in our conceptual system (Fodor 1983; Jackendoff 1983; 1990).

Our discussion will be in three parts. First, we will discuss language pertaining to object representation – in particular, that required for identification – and its bearing on some current theories of the encoding of object shape. Second, we will discuss language pertaining to places, the locations of objects and the spatial relationships holding between different objects. We will highlight the differences between these two aspects of language, in particular, the fact that object naming draws on rich shape-based representations of objects, whereas place naming draws on quite sparse elements of object shape.

These differences will lead us to the third issue of why the identification and location of objects draws on such disparate representations. We will consider two hypotheses, one focusing on the design of language and one focusing on the design of spatial representations. We will conclude with some predictions derived from these hypotheses.

## 1. Talking about objects

In the average adult vocabulary, there are roughly 10,000 names for things – count nouns that label different kinds of objects. For a large proportion of object categories, shape is among the most important criteria for identification, and in particular for judgments of what a thing should be called: Categories of things with the same shape, including natural kind objects and artifacts, often share the same name. The importance of shape in object naming is particularly dramatic in the case of representational art. As one example, the 60-foot metal sculpture by Claes Oldenburg that graces downtown Philadelphia is universally recognized as and labeled "the clothespin," although it clearly violates most of the critical properties of true clothespins: Its shape is the dominating criterion in choosing its name.[3]

Aside from such informal observations, experimental evidence shows that two-dimensional representations of object shape can support object identification and naming. Adults can easily recognize and label familiar common objects – both artifacts and natural kinds – on the basis of line drawings or silhouettes (Biederman 1987; Rosch et al. 1976). Such representations elicit recognition immediately, automatically, and irresistibly.[4] This recognition ability is present in infancy (DeLoache et al. 1979) and may be innate: Even a child who has never learned name-picture correspondences can easily identify familiar common objects from simple black-and-white line drawings (Hochberg & Brooks 1962). (In contrast, imagine how difficult it would be to name most objects based only on a patch of their color or texture.)

Object shape also seems to have a privileged status in learning names for *novel* objects. A number of studies have shown that young children rely specifically on shape when learning labels for novel objects (Au & Markman 1987; Bornstein 1985; Heibeck & Markman 1987; Landau et al. 1988). For example, Landau et al. found that when young children or adults were shown a novel object labeled as a count noun (e.g., "This is a dax"), they tended to generalize that label to objects of the same shape as the original, even in the face of rather large differences in texture or size. In addition, they found that this "shape bias" becomes stronger over age, with 2-year-olds showing a rather weak but stable preference, 3-year-olds showing a stronger one, and adults showing the bias in an extreme form, consistently rejecting even quite small deformations in object shape. Finally, they found that the shape bias does not strictly mirror perceptual preferences and can even occur in the context of highly salient competing properties (Jones et al. 1991; Landau et al. 1992; Smith et al. 1992). The shape bias appears most consistently and strongly in the context of the word-learning task, suggesting a developmentally early link between names for things and representations of object shape.

The importance of shape imposes a basic constraint on the relation between spatial representations and language: The spatial representations that are linked to object names must provide enough different shape descriptions, configured in the proper way, to be able to distinguish all the kinds of objects we categorize (or partially categorize) linguistically on the basis of shape. We propose that just as the number of possible shapes must meet or exceed the number of object names in a language (Biederman 1987), any spatial distinctions we can encode linguistically must be capable of corresponding to spatial representations. Such a correspondence is necessary if talking about objects and places is linked to thinking about or acting on them. This section begins by reviewing some current approaches to the spatial representation of object shape. We then proceed to show that linguistic evidence motivates interesting augmentations of the theory of spatial descriptions.

### *1.1. Whole objects*

From the nonlinguistic side, a traditional approach to the problem of shape description takes object shapes to be represented componentially by simple three-dimensional components such as cylinders. This approach dates back at least to Leonardo da Vinci, who described the human figure as a combination of units fitting a particular set of proportions. Twentieth-century work in computational vision and psychology has also drawn on componential analyses to describe object shape (Biederman 1987; Binford 1971; Lowe 1985; Marr 1982). That is, a limited number of shape components are taken to be the units used by perceivers in recognizing object shapes.

What corresponds most closely to our notion of spatial representation is dubbed by Marr the "3-D model level," in which objects are encoded in an object-centered format, independent of the viewer's perspective. Even though this approach has been developed primarily for vision, it should be capable of accepting inputs from haptic/kinesthetic sources as well, making it nonmodality-specific. Such an interpretation of the componential approach would be compatible with the extensive evidence that object shape is encoded in some detail by the congenitally blind (see, e.g., Landau 1991; Lederman & Klatzky 1987). Moreover, Jackendoff (1987a; 1987b) has shown that 3-D model representations can be translated in part into representations suitable for linguistic expression. Hence, the 3-D model level, insofar as it can be fleshed out, has the properties with which we have characterized spatial representation.

In most componential approaches, the primitives for 3-D object description are related to Binford's (1971) notion of the generalized cylinder. In Marr's model, for example, object representations are built up from (1) a set of principles for describing "generalized cones" in terms of an axis and a varying cross-section, and (2) a principle for elaborating a generalized cone by adding a subsidiary generalized cone whose axis is of a particular size and orientation relative to the main axis. Principle (2) applies recursively, so that objects in the 3-D model representation are composed of parts, each of which may have a further decomposition. As a result, highly detailed shapes can emerge, making the representations in principle rich enough to support the extensive vocabulary of object names seen in language.

As another example, Biederman (1987) proposes that the parts can be encoded in terms of a small specific set of generalized cones, 36 in number, which he calls "geons." These geons are meant to have "nonaccidental" properties (Lowe 1985) – properties in the image that are likely to represent true properties of the object rather than accidents of viewpoint. For example, the geons are meant to be contour-determined and invariant over size and viewpoint transformations.[5] Given a small set of attachment relationships among the cones – such as "end-to-end" or "top-to-side" – and only a few iterations, Biederman's system can generate at least as many object shapes as there are object names (as would be expected from any componential system of sufficient complexity).

The componential systems are all capable of generating a wide range of particular object shapes from primitives. For example, Marr describes schemata for the human figure and various animals, whereas Biederman describes such objects as airplanes and cameras (see Figure 2a and 2b). In principle, these systems should be able to generate enough descriptions to cover all named or unnamed objects. With a suitable similarity metric, they could account for differences among named categories such as *person* versus *gorilla*, whose shapes are quite similar.[6] Furthermore, using these componential systems, one can characterize transformations that preserve the geometric structure of objects capable of internal movement, capturing for example the possible limb movements of a human figure, as described by Marr and Vaina (1982) (see Figure 2c). Finally, the componential systems are capable of characterizing many kinds of object parts, an issue to which we turn next.

### *1.2. Names for object parts*

A significant part of the lexicon includes names for coherent object parts: handles, noses, legs, stems, and so on. For some parts, the componential approach provides straightforward ways of parsing the host object: Parts are represented by the individual generalized cones that combine to yield the whole object. Thus, some of the named parts of a camera, an airplane, or a person could be described by particular cones. Using Marr and Vaina's scheme, one can add that certain body parts – the head, legs, or arms of a person – follow from parsing at critical movement joints. The regions of the joints themselves correspond in turn to names for parts such as elbows, knees, and wrists.

For other parts, the componential approach seems less appropriate: Many named parts, especially parts of natural objects, are not well described by cylinders. For example, how does one tell where a forehead or a nose begins and where it ends? In such cases, two different approaches have been suggested. Hoffman and Richards (1984) suggest that parts may be defined by certain characteristics of the object's boundaries rather than by a set of prespecified primitive shapes. They propose that a part is perceptually defined as the segment between two consecutive contours of negative minima within the boundary's principal curvatures. The nose, for example, would begin and end at points of negative extrema along the external boundary of the face (see Figure 3a).

A second approach, offered by Leyton (1989), suggests that parts may be those units produced by the causal processes underlying shape formation. For example, processes such as "squashing," "protrusion," or "indentation" can operate on an initial shape to yield quite a different shape (which can nevertheless be seen as related to the original by the relevant causal processes). The nose, for example, could be the result of a growth process wherein a segment of a simple curved surface is pushed out from the inside along a principal axis to form a new segment (Figure 3b).

Each approach is promising in that it affords us a way of linking descriptions of further object parts with their labels. Of course, this linking is not one-to-one: We can often find parsable parts of objects that go unnamed.

Even with these enrichments, however, there are a number of cases in which evidence from language suggests the need for aspects of shape descriptions that are

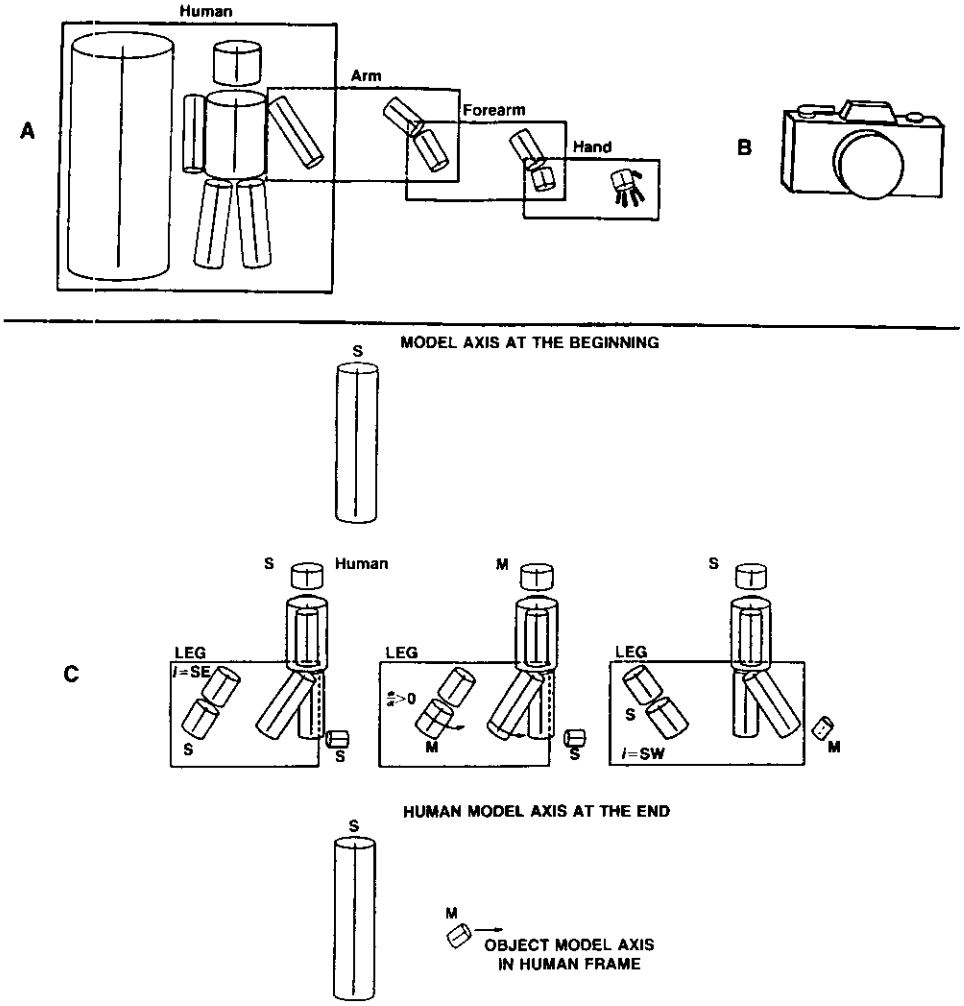



Figure 2. Componential theories of object shape can provide detailed descriptions of complex shapes based on the combination of a few primitives. A: Marr's (1982) schemata for the human figure. B: after Biederman's (1987) camera. C: Marr and Vaina's (1982) schemata for the human figure under movement transformations. (Figure 2A reprinted from Marr [1982] and 2C from Marr & Vaina [1982].)

not a prominent part of these proposals about object representation. The rest of this section makes some elementary observations that point to a number of such gaps. In each case we will suggest enrichments of spatial representation that seem compatible with the approaches described so far.

### 1.3. Names for spatial parts: Axes and axial parts

Many objects can be described as having a *top* and a *bottom*, a *front* and a *back*, and *sides* and/or *ends*. These terms do not describe parts in the same sense as, say, *handle* or *wing:* They are not subsidiary parts tacked onto the object. Rather, they denote regions of the object based on its inherent orientation (as opposed to contextually imposed uses; see sect. 2.7.2). How can these terms be derived from a spatial representation?

Leaving language aside for a moment, notice that the orientation of an object is necessary for describing certain relationships among parts: for example, the fact that one's nose, feet, and navel point in the same direction (hence are on the *front* of the body) or that one's arms are attached opposite one another and orthogonal to the front (hence are on the *sides*). Relative orientation of parts can be crucial in nonlinguistic tasks such as mental rotation, where judgments of object identity may require discriminating right-left reversals of an object (Parsons 1987; Shepard & Cooper 1982; see Tarr & Pinker 1989 for emphasis on orientation-dependence). A principal oriented object axis appears to be critical for our representations of objects in the "implicit" memory system (Cooper 1992).

Both the linguistic and the nonlinguistic facts therefore suggest the need for explicit representation of oriented and directed axes in object descriptions, that is, the axes required to distinguish top from bottom, back from front, and right from left. The componential approaches we have described provide for representing local axes (the axis of each part) and overall object axes (such as the principal axis of a canonically oriented object; see Marr & Nishihara 1978); but as far as we can tell, they make no provision for representing the orientation of these axes. In other domains, however, such as morphological development (Thompson 1961) or perceptual analysis of changes in shape (Leyton 1992), oriented axes are critical to understanding the structure and transformations of natural shapes.

One way to express regularities of orientation is to

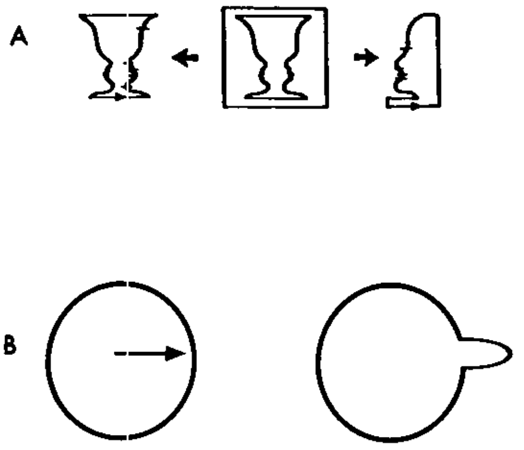



Figure 3. Theories of object parsing can provide descriptions of coherent object parts named by nouns such as nose and handle. A: Hoffman and Richards (1984) define a part as the segment between two consecutive contours of negative minima within the boundary's principal curvatures. B: Leyton's (1989) theory suggests that parts may be defined as those units produced by causal processes underlying shape formation. Here, a process of protrusion creates a "nose" within a uniformly curved boundary. (Figure 3A reprinted from Hoffman & Richards [1984].)

extend the theory of axes in two ways. First, we will call the axis that is expanded into a generalized cone the *generating axis;* this is the axis that is central to Marr's and Biederman's accounts of object shape. Let us impose on this cone up to two further axes, called *orienting axes,* which are orthogonal to the generating axis and to each other. These axes will serve to orient the principal cone radially (see Figure 4). In the case of the human body, the principal generating axis is vertical; it defines the generalized cone of the torso. The two orienting axes determine the front-to-back and side-to-side directions.

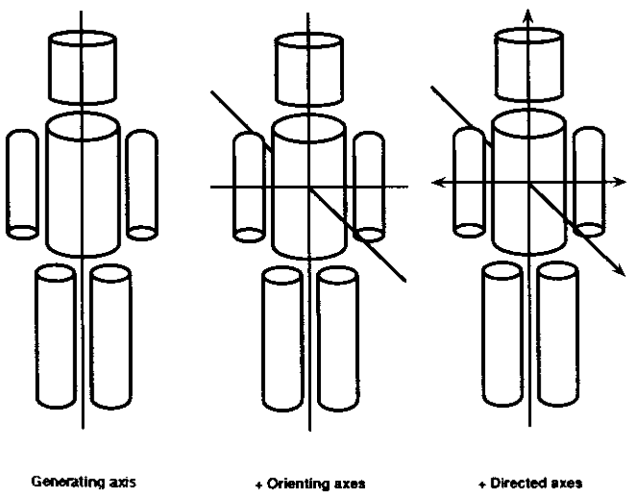



Figure 4. Three types of axes are required to account for linguistic terms describing aspects of an object's orientation. The *generating axis* is the object's principal axis as described by Marr (1982). In the case of a human, this axis is vertical. The *orienting axes* are secondary and orthogonal to the generating axis and to each other (e.g., corresponding to the front/back and side/side axes). The *directed axes* differentiate between the two ends of each axis, marking top vs. bottom or front vs. back.

Second, an axis can be marked optionally as intrinsically *directed* or *symmetric* (see Figure 4). A directed axis indicates inherent regularities that distinguish one end from the other: This can distinguish top from bottom or front from back. For example, the directed front-to-back axis of the human figure establishes the regularity of alignment for the nose, feet, and navel. A symmetric axis indicates equivalent elaborations of the object at both ends of the axis. For example, the side-to-side axis of the human figure establishes the symmetry of the limbs and face parts.

The generating axis of a cone as well as its orienting axes can be directed or symmetric. In the human figure, for example, the main generating axis is a directed one that distinguishes top from bottom. In Biederman's camera, if the long side-to-side dimension is the generating axis for the principal geon, then that axis will also be symmetric (since right/left need not be distinguished in a camera, aside from minor details).

Other combinations of generating and orienting axes are also possible. For example, an arrow has a directed generating axis but no significant orienting axes. The human hand has an oriented generating axis (wrist-to-fingers, following Marr), and two directed orienting axes (back-to-palm and pinky-to-thumb).

Returning to the linguistic description of objects, we can use the system of directed and orienting axes to define the terms brought up at the beginning of this section, which we can call "axially determined parts." The *top* and *bottom* of an object are the regions (or parts of the surface) of the object at the ends of whichever axis is vertical in the object's normal orientation. If the object is relatively long and narrow, that is, if it has a horizontal generating axis significantly longer than the other axes, it can be said to have *ends* – the regions at the termination of this axis. If the object has a horizontal directed axis, with one that normally faces the observer or determines the normal direction of motion, the region determined by that end of the axis is the object's *front;* the opposite end of this axis determines the *back.* Finally, the region determined by the termination of any other horizontal axis can be called a *side.*[7] Thus, linguistic and nonlinguistic facts about shape converge in motivating use of the axial system as an important part of object representation.

### *1.4. Names for objects best described as surfaces*

Consider the spatial representations of sheets of paper, phonograph records, crackers, table tops, blackboards, rugs, roads, and lakes. What these have in common is that they are principally extended in two dimensions, with a relatively negligible thickness (at least in the relevant context); the linear boundary of this surface can then be defined as its *edge.*

How are such objects to be encoded in spatial representation? It seems wrong to treat a phonograph record as a very fat cylinder with a very short main axis passing through the hole. However, this is the only way to generate it formally in the componential framework of volumetric primitives. Furthermore, a lake hardly lends itself at all to such a description. For example, if its generating axis is taken as going from the surface to the bottom, the description is entirely counterintuitive, especially if the lake is irregular in shape. Alternatively, its generating axis might be taken as parallel to the surface (say if the lake is relatively long and narrow). Our own intuition, however, is that it would be odd to have it fall in the interior of the geon (as in Biederman's [1987] repertoire): That would put the axis under water.

The problem in these cases is that models using volumetric primitives require one to generate a volume directly from a linear axis. A more intuitively satisfying analysis of these objects is that they are schematized as *surfaces*, possibly elaborated into a volume by adding a *thickness*. In this analysis, the record is schematized as a disk rather than a volume; the lake is schematized more as its surface, with depth as an elaboration. The *edge* of such objects is the linear boundary of the schematized surface.

In addition to object names for such "surface-like" objects and for the term *edge*, there are other words that benefit from having such analysis available in spatial representation. Two classes come to mind: (1) two-dimensional shape terms like *square*, *circle*, *oval*, *trapezoid*, and so on; (2) general terms for "thickened surfaces" such as *slab*, *sheet*, *layer*, *slice*, *lamina*, and *stratum*.

To further motivate the distinction between "surface-type" and "volume-type" objects, consider what happens when these different types of objects are modified by a dimensional term such as *big*. Like many adjectives, *big* selects different dimensions, depending on the nature of the object. If an object is inherently surface-like, it can

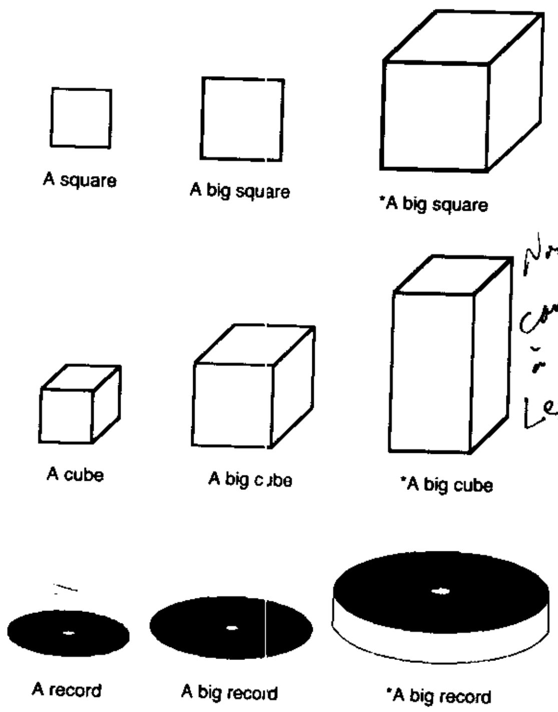



Figure 5. The distinction between "surface-type" and "volume-type" objects can be seen when they are modified by a dimensional term such as *big*, which has the function of enlarging the object in each of its critical dimensions. A *square* is inherently two-dimensional, so a "big square" can only be enlarged in those two dimensions, but not in the third dimension (in which case it becomes a cube). A *cube* is three-dimensional, so it must be enlarged equally in its three dimensions. In this analysis, a *record* behaves like the two-dimensional object: A "big record" must be large in its two primary dimensions, but not in the third.

only be enlarged in two dimensions: A big square is large in each of its two dimensions, a big cube is large in each of its three dimensions (see Figure 5). Surface-type objects, like *record* or *lake*, behave like *square* in this analysis: A big record is large in two dimensions, but if significantly enlarged in its third dimension, it ceases to be seen as a record (see Figure 5). Similarly, a big lake is one that is extended in length and width; its depth is irrelevant. Other dimensional adjectives have related application. For example, the adjectives *thick* and *thin* can be seen to place a metric on the elaboration of a surface into a volume: A thick record is one that is relatively large in the third dimension.

### 1.5. Names for "negative" object parts

Some entities may be best conceptualized as "negative parts" of objects, as alluded to by Hoffman and Richards (1984) and Herskovits (1986). Compare a *ridge* and a *groove*. A ridge is conceptualized as a protrusion from the surface of a host object. It has an extended linear generating axis parallel to the surface of the host object. In addition, it has a directed orienting axis that projects out of the surface of the host object, giving the ridge a top and a bottom, and a (roughly) symmetrical orienting axis that defines its sides. It can therefore be easily described as a part of the host object using the volumetric system as elaborated so far.

Now consider a groove. It is conceptualized as a depression in the surface of a host object. It has an extended linear generating axis parallel to the surface of the object plus a directed orienting axis that projects into the surface of the host object, giving the groove a top and a bottom, plus a (roughly) symmetrical axis that defines its sides. Although one could describe it within the volumetric system as an indented volume, we suggest an alternative representation.

A natural way to think of a groove is as a "negative part," a shaped volume scooped out of the object instead of added to it. That is, it is a shape defined by "lack of substance" rather than by presence of substance, as in the case of normal parts. Other than that, a negative part evidently has shape descriptors – and a linguistic description – essentially parallel to those of ordinary object parts. A groove not only has a top, bottom, and sides; it can be described as *long* or *short* (along its principal axis), *deep* or *shallow* (along its secondary axis), and *broad* or *narrow* (along its tertiary axis). Notice further that the terms *deep* and *shallow* play the same role for negative parts that *high* and *low* do for ordinary parts: We speak of a deep groove rather than a *high groove, for instance.

Other negative part names are *hole*, *pit* (a "negative" *bump*), *notch*, *slot*, *scratch*, *depression*, *cavity*, and possibly *dent*. Words that name "negative objects" are *valley*, *ditch*, *cave*, and *well*, as well as *door* and *window* (in the sense of "opening in a wall" rather than the object used to close off such an opening). Thus, a simple enrichment of spatial representation again affords revealing analyses for a wide variety of things we can name (though strictly speaking they are not *objects* this time).

### 1.6. Names for containers and related objects

As Miller and Johnson-Laird (1976) point out, English has an extensive set of labels for containers: objects like cups, bowls, boxes, jars, tanks, and so forth. What is their spatial representation? One possibility is that a cup, for instance, is a volume – a cylinder out of which a large coaxial negative cylinder has been scooped. An alternative with a certain intuitive appeal is that a cup should be represented as a thickened surface that encloses a cylindrical space – that is, the sides and bottom of the cup are not the residue of extensive scooping, they are surfaces.

Although these two alternative descriptions require an empirical test, there is some evidence that language distinguishes containers from solid objects; roughly, containers are objects that can hold things inside them (see sect. 2.4.1). To encode a class of containers in spatial representation, we tentatively adopt the second alternative and introduce the notion of "hollow" volumes, distinguished from the standard "solid" volumes (see Marr 1982, for a similar suggestion). Solid geons would be encoded as uniformly substantial; hollow geons would be shapes with a substance that is distributed only over their surfaces, leaving a shaped empty space inside.

In addition to the containers mentioned above, hollow volumes would permit the theory of spatial representation to neatly encode such objects as cars and other closed vehicles, houses and other buildings, stomachs, eggshells, balloons, bubbles, violins, and drums. Again, a simple parameter added to spatial representation affords an intuitively natural treatment of a significant new class of objects.

### 1.7. Summary

To name objects and object parts, spatial representation of objects by shape must be a rich combinatorial system. Its basic units include not only generalized cones, but also surfaces. Cones may be marked as "solid," "hollow," or "negative." Each unit has an axial structure: the generating axis around which the cone or surface is elaborated, plus up to two orthogonal orienting axes. In turn, each of the three axes may be directed or symmetric. These basic units are combined hierarchically to form complex object descriptions.[8]

In addition to our proposed additions to the descriptive power of the componential framework, there remains the task of showing that the whole system of decomposition into parts can be adapted to the tolerances necessary for object category discrimination. For example, descriptions must be potentially fine-grained enough that one can decide which objects are to be named *horse* and which *donkey*, or which *dog* and which *wolf*. On the other hand, they must be potentially indeterminate enough to allow considerable variation in shape within these named categories, for example, the differences between Dalmatians, German Shepherds, and Pekingese dogs, and to allow such variation as occurs in the number, placement, size, and shape of arms on a saguaro cactus. How these tolerances for discrimination are to be formalized is beyond the scope of this target article (see Note 6).

## 2. Talking about places

The componential framework and all the amplifications we have discussed concern object recognition and categorization, that is, *what* an object is. None of this addresses *where* the object is or, if in motion, its path of movement: That is an essential part of spatial cognition – and an essential part of spatial language.

There is a large and diverse literature on spatial representation in humans and other species. In addition to the vast body of research on space perception, the variety of perspectives includes that of psychologists interested in how perceptual-motor coordination is achieved (Hein & Jeannerod 1983); how spatial knowledge is structured (Gallistel 1990), how it develops (Stiles-David et al. 1988) and how it is represented neurally (O'Keefe & Nadel 1978). There is also a considerable tradition in city planning and environmental psychology that seeks to understand how we represent large spatial layouts such as the cities in which we live (Coucelis et al. 1987; Downs & Stea 1973; Hooper 1978; Kuipers 1978; Lynch 1960).

The variety of perspectives makes it difficult to identify a single unified theory of spatial representation that we might use as a model for thinking about how language encodes spatial relations. In taking these approaches as a whole, however, one theme is pervasive: Understanding our representations of space requires invoking mental elements corresponding to *places* and *paths*, where places are generally understood as *regions* often occupied by *landmarks* or *reference objects*. Objects (including oneself) are then located in these places. Paths are the routes along which one travels to get from place to place.

These elements are likely to be critical in any complete theory of spatial representation and will serve as a skeletal organization for our discussion of the language of space. We draw especially on the work of Bennett (1975), Clark (1973), Fillmore (1975), Hawkins (1984), Herskovits (1986), Jackendoff (1983; 1990), Miller and Johnson-Laird (1976), and Talmy (1978; 1983). As with our discussion of objects, we will show that linguistic evidence both converges with and enriches findings from nonlinguistic studies of spatial cognition.

### 2.1. Basic elements: Figure, reference object, region

The standard linguistic representation of an object's place requires three elements: the object to be located (or *figure*), the reference object (called *ground* by Talmy[9]), and their relationship. In the canonical English expression of an object's location, the figure and reference objects are encoded as noun phrases; the relationship is encoded as a spatial preposition that, with the reference object, defines a *region* in which the figure object is located. For example, in the sentence, "The cat is sitting on the mat," the figure (the cat) is located in the region described by the prepositional phrase *on the mat*. The region is in turn described by the reference object (the mat) and the spatial relation expressed by the preposition *on*, roughly, "contact with the surface of the reference object." In addition to prepositions, there are many verbs that incorporate spatial relations; these can (almost invariably) be paraphrased by a simpler verb plus a preposition. For example, *enter* can be paraphrased by *go into*, *approach* by *go toward*, and *cross* by *go across*. (See Jackendoff 1983; 1990 for formalization of these relations.)

Thus, the key element in the English expression of place is the preposition: We can develop a fairly comprehensive idea of the spatial relations expressed in language by focusing on spatial prepositions. We present a fairly

Table 1. *Prepositions of English*

| | | |
|---|---|---|
| about | between | outside |
| above | betwixt | over |
| across | beyond | past |
| after | by | through |
| against | down | throughout |
| along | from | to |
| alongside | in | toward |
| amid(st) | inside | under |
| among(st) | into | underneath |
| around | near | up |
| at | nearby | upon |
| atop | off | via |
| behind | on | with |
| below | onto | within |
| beneath | opposite | without |
| beside | out | |
| *Compounds* | | |
| far from | on top of | |
| in back of | to the left of | |
| in between | to the right of | |
| in front of | to the side of | |
| in line with | | |
| *Intransitive prepositions* | | |
| afterward(s) | forward | right |
| apart | here | sideways |
| away | inward | south |
| back | left | there |
| backward | N-ward (e.g., homeward) | together |
| downstairs | | upstairs |
| downward | north | upward |
| east | outward | west |
| *Nonspatial prepositions* | | |
| ago | for | |
| as | like | |
| because of | of | |
| before | since | |
| despite | until | |
| during | | |

complete list of the prepositional repertoire of English in Table 1.[10] Additional prepositions exist, but these are typically either archaic or reserved for technical usage (e.g., *betwixt, athwart, abaft,* etc.) and, in any case, these prepositions do not violate the principles set forth below.

A salient fact about prepositions is that there seems to be surprisingly few of them in comparison to the number of names for different kinds of objects. (In fact, there are few enough prepositions that they are usually considered part of the "closed-class" vocabulary, along with auxiliaries, determiners, and inflections.) We can get an idea of the order of magnitude of different spatial relations expressed in English by counting the prepositions (see Table 1). There is something on the order of 80 to 100, depending on how one counts. Many of these are polysemous (e.g., the different senses of *over*), and quite a few are nonspatial (*during*, for instance, is purely temporal), so this estimate gives us only a ballpark figure. But compare it to the number of count nouns in English – tens of thousands. Again, many of these are polysemous, and many are not object names, so this estimate is only rough. But even supposing the estimate is drastically biased, there is a difference of approximately two orders of magnitude: For every spatial relation expressible in English, there are perhaps a hundred object names. This qualitative difference is reproduced in every language we know of. (If there were a language with even a thousand prepositions, someone would certainly have raised a big hue and cry about it.)

Given the small number of prepositions, the word class most clearly devoted to expressing spatial relations, one is led to ask what constrains the range of possibilities so severely. Our hypothesis is that there are so few prepositions because the class of spatial relations available to be expressed in language – the notions prepositions can mean – is extremely limited. This section presents what we believe is a rather comprehensive enumeration of the factors involved in defining the spatial relations expressed in English: the totality of meanings of spatial prepositions.

These factors divide into four classes. The first, the asymmetry between figure and reference objects, sets the basic parameters for spatial relations (sect. 2.2). The remaining three concern geometric possibilities for the three key elements of the spatial relation: the reference object (sect. 2.4), the figure object (sect. 2.5), and the region based on the reference object (sect. 2.7).

### 2.2. Asymmetry between figure and reference object

It is logically possible that spatial relations could be mentally encoded as binary relations between objects, that is, as propositional functions of the form $R(a,b)$, where $a$ and $b$ are the objects to be related. In human languages, however, a predominant way to express spatial relations is asymmetrical. In the canonical form, the figure is encoded as grammatical subject, and the reference object is encoded as the object of the spatial preposition or of the verb itself.

To illustrate, we have annotated the figure and reference objects in the sentences in (1).

(1) a. The book (figure) is lying on the table (reference object).
b. The train (figure) reached the station (reference object).
c. The star (figure) is inside the circle (reference object).
d. The circle (figure) lies around (surrounds) the star (reference object).

Note that (1c) and (1d) can describe the very same physical stimulus (see Figure 6A). They organize it differently, however, exchanging figure and reference object. These different organizations appear to reflect differences in the encoding of the stimulus in spatial representation, with primary attention switching from one object to the other.

There appears to be a canonical way of expressing linguistically the assessment of object to the roles of figure and reference object. Whereas the exchanges listed above are not unusual, they are not always possible. As noted by Miller and Johnson-Laird (1976) and Talmy (1983), if the objects are unequal in size or mobility, the larger and more stable is invariably encoded as the reference object. For example, in (2), an exchange comparable to (1c,d) produces the odd-sounding result (2b).

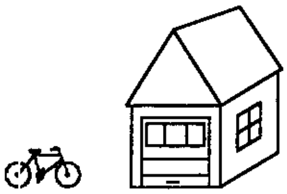
(2) a. The book is on the table.
b. ?The table is under the book.

A

(a) The star is inside the circle.

(b) The circle lies around (surrounds) the star.

B



(a) The bike is near the garage.

(b) ? The garage is near the bike.

Figure 6. A: The same scene can be described in two different ways, by reversing the figure and ground objects. B: Even an apparently symmetrical spatial relation – "nearness" – is subject to asymmetry of figure and ground, as is shown by the relative naturalness of the two sentences.

Even what would seem to be a symmetrical spatial relation – adjacency – is subject to the asymmetry of figure-reference object dyads, as shown in (3) and Figure 6b.

(3) a. The bicycle is next to the house.
b. ?The house is next to the bicycle.

This is not to say that spatial expressions *must* be asymmetrical (consider, e.g., "the bicycle and the house are adjacent to each other"; see Landau & Gleitman 1985, for discussion), but only that this is the norm.

What causes these asymmetries? It does not seem to follow from any fact specifically pertaining to language that, in these contexts, the table and the house are more plausible reference objects and the book and bicycle are more plausible figures. (In particular, if "the house" in (3b) happens to refer to a toy house, which is smaller and more mobile, the sentence becomes much more acceptable.) Rather, we believe that this linguistic asymmetry follows from principles of spatial organization, which require that an object be anchored (or located) relative to some other object. Reference objects should have properties that facilitate search: In many contexts, they should be large, stable, and distinctive (and in environmental contexts they are often landmarks; Lynch 1960). That is, in this case, the organization of language parallels the organization of spatial cognition.[11]

A variety of experimental evidence supports the notion that these asymmetries are fundamental to our spatial representations – in particular, that people expect figure and reference objects to differ in size and stability. For example, in perception, some cases of "induced movement" show the importance of differentially marking figure and reference frame: If a stationary dot is placed within a moving rectangular frame, observers see the dot as moving rather than the frame (Duncker 1929). In such cases, the object that surrounds the other or tends to dominate it by greater size or intensity will tend to become the reference object (Oppenheimer 1934).

A similar asymmetry appears to guide our assignment of objects to the role of figure versus ground when making explicit judgments of the relative distance between objects in spatial layouts. Sadalla et al. (1980) asked people to judge distances between different campus landmarks that were thought to be either good or poor reference objects. Good reference objects had been shown by independent ratings to be relatively large, stable, familiar, or culturally significant. People tended to judge the distance of a poor reference object (e.g., the architecture building) from a good reference object (e.g., the student union) as shorter than the reverse, suggesting that reference objects are critical anchors in structuring cognitive maps (see also Coucelis et al. 1987). Thus, converging with the linguistic evidence presented above, a logically symmetrical relationship – distance – is treated as a psychologically asymmetrical relationship when people make certain kinds of distance judgments.

Evidence from language tasks also confirms the linguistic reflection of figure-reference object asymmetry, even in young children. A classic study by Huttenlocher and Strauss (1968) showed that children and adults respond more quickly to sentences in which the mobile object is named as grammatical subject of the sentence and the stable object is named as grammatical object of a preposition than they do to sentences in which these roles are reversed. For example, subjects were quicker and more accurate when asked to "Make it so the (mobile) block is on top of the (fixed) block" than when asked to "Make it so the (fixed) block is on top of the (mobile) block" (i.e., by placing the mobile block under the fixed block).

This effect even appears for the symmetrical predicate *near.* Landau et al. (forthcoming) showed children and adults pairs of objects and asked them to "Make it so the (mobile object) is near the (fixed object)" or the reverse. Reaction times to the canonical form were much quicker, and some adults even claimed they could not carry out the request in its noncanonical form (e.g., "Make it so the house is near the bicycle"), because the fixed object could not be moved.

Finally, this sensitivity to asymmetry appears in explicit judgment tasks. Landau et al. (forthcoming) showed children (2-, 4-, and 6-year-olds) and adults drawings of pairs of objects, and asked them to make judgments of asymmetry. In each pair of objects was one that could naturally serve as a reference object for the other because it was larger or more stable – for example, a picture of a house (reference object) adjacent to a bicycle (figure). Subjects were asked, "Which one is near which one?" or, in some cases, "Is the house near the bicycle, or is the bicycle near the house?" Both of these requests pushed subjects to make an asymmetry judgment for a predicate (*near*) that is logically symmetrical. Even the youngest children tended to place the smaller and more mobile object in subject position and the larger and more stable object in prepositional object position, saying, for example, "The bicycle is near the house" rather than "The house is near the bicycle." A similar finding was obtained by Sadalla et al. (1980): When people were asked to assign two well-known campus buildings to the sentence "_____ is close to _____," they tended to assign the clear reference objects to the prepositional object position, again consistent with linguistic observations on asymmetry.

In sum, asymmetry between figure and reference object emerges in a variety of tasks, ranging from strictly perceptual judgments to cognitive judgments of distance and linguistic judgments of semantic naturalness.[12] This suggests that the asymmetry arises as a consequence of our nonlinguistic representations of space and that language draws on them as a means of expressing a critical aspect of these representations.

### 2.3. The sparse constraints on the geometry of figure and reference object

Next we explore in more detail the components of spatial relations, beginning with how the system of spatial relations expressed in language makes use of the shapes of the objects being related. Having just discussed the intricate shape descriptions required for object naming, it is worth asking to what extent the same descriptors are used when describing an object's place. What proves surprising is how sparsely both the figure and reference objects appear to be represented.

Taking a simple case, there seem to be no prepositions with a figure or reference object that must be analyzed in terms of a particular geon. A hypothetical example would be the preposition *sprough*, "reaching from end to end of a cigar-shaped object," appearing in sentences like (4a) but not (4b) (see Figure 7).

(4) a. The rug extended sprough the airplane.
The weevil bored sprough the cigar.
The major axis of an ellipse goes right sprough it.
b. *The rug extended sprough my dining room.
*The weevil bored sprough the chair.
*The major axis of a cup handle goes right sprough it.

Similarly, there are no prepositions that insist on analysis of the figure or reference object into its constituent parts. An example might be the hypothetical preposition *betwaft* in (5), which requires the reference object to have a protruding part.

(5) a. The bug crawled betwaft my face.
"The bug crawled down the junction between my nose and the main body of my face."
b. The water ran betwaft the airplane.
"The water ran down the junction between the wing and the fuselage."
c. A stripe extended betwaft the cup.
"A stripe extended along/down the junction between the body of the cup and the handle."

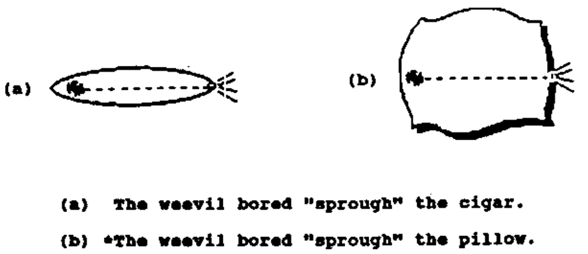



Figure 7. There is no preposition in English whose reference object must be analyzed as a particular geon, such as one that would require a cigar-shaped reference object, making only (a) grammatical.

For a third case, it is easy to think of many dozens of container terms of the sort discussed in section 1.6, for example, *cup, bowl, vase, box, carton, crate, coffin, bag, pouch, sack, sheath, case, tank, barrel,* and so on. Yet there are essentially only two spatial relations that pertain to containers: Something can be either *in* a container or *out of it* (but see discussion of verbs in sect. 2.6). For example, there is no preposition *plin* that describes contact with the inner surface of a container, so that one can say:

Bill crawled *plin* the water tank, in the sense
"Bill crawled along the inner surface of the tank"; or
Bill spread paint *plin* the carton, in the sense
"Bill spread paint on the inner surface of the carton."

The inner surface must be explicitly mentioned. Like *sprough* and *betwaft*, this hypothetical term represents a perfectly plausible spatial relation, but a perfectly horrible preposition.

About the most complicated cases we have found in English – in which some elements of object shape are relevant to the preposition's meaning – are the terms *along* and *across*. *Along* requires its reference object to have a principal axis of significant elongation, so one can travel along a road or along a beach, but not along a chair or along a round table. One can travel along the edge of a round table, but then the linear edge, not the table as a whole, is serving as reference object. In addition, this principal linear axis must be (more or less) horizontal: A bug can be said to crawl along a flagpole only if the flagpole is lying down. (We treat *across* and some further wrinkles in *along* shortly.)

And that is more or less it with regard to specific shape requirements. The detailed descriptions of shape relevant to the naming of objects appear to be irrelevant to the descriptions of the same objects in their role as figures or reference objects. Only very sparse schematization of the objects is relevant.

Landau and Stecker (1990) produced experimental evidence for this bifurcation between the description of objects as objects and the description of objects in their role in constructing spatial relations. They showed 3-year-olds, 5-year-olds, and adults a single scene in which a novel object was being placed on the top front corner of a box. Half of the subjects were told, "This is *a corp*," as if the novel word named the object itself. The other half were told, "This is *acorp* my box," with the same novel phonetic sequence now playing the role of a preposition, as if it described the object's location. All subjects were then shown a series of novel objects being placed one at a time in novel locations on a second identical box. Each time, they were asked either, "Is this a corp?" or, "Is this acorp my box?" with the choice of question matching the syntactic context of the introductory sentence they had heard.

Subjects who had heard the sentence with a novel count noun generalized it to objects of the same *shape* as the original, completely ignoring the object's location on the box. In contrast, subjects who had heard the sentence with a novel preposition generalized it to objects of *any* shape, as long as they were in roughly the same *location* as the original. This suggests that both children and adults attended to details of an object's shape when it was named as an object, but completely ignored the same object's shape when what was at issue was its role as figure in a locational expression.

In a further experiment, Landau and Stecker asked whether subjects would ever preserve any components of object shape in the preposition condition. This time, subjects were shown a straight rod lying across the top of a box. The test objects included a straight rod, a squiggly rod of the same extent, and a cube, each placed in different positions on and off the box. As in the first experiment, children and adults preserved precise shape in the count noun condition (accepting only the straight rod) and location in the preposition condition. But in the preposition condition, they also accepted both rods of equal extent, even though they were quite different in precise shape. Thus, a novel preposition led subjects to preserve both location and a very general component of the figure object's shape – linearity – while still ignoring the detailed shape.

These studies tell us that even young children pay attention to only very sparse properties of objects when what is at issue is their location: Either they ignore shape entirely or they preserve only very crudely schematized components of shape. We assume that children come prepared or learn very rapidly to draw on qualitatively different kinds of representations when learning the names of objects (count nouns) versus the names of spatial locations. We return to this issue in section 3.2.3.

We will now enumerate the linguistic evidence for overall constraints on the range of geometric properties relevant to the description of figure and reference objects. Reference objects are the more complex of the two and we describe their constraints first. (A summary of constraints appears in Table 2 at the end of sect. 2.)

### *2.4. Reference objects*

The few restrictions on reference objects concern their treatment as certain geometric types (volumes, surfaces, points, and lines), their axial structure, and their quantity. We take these up in turn.

**2.4.1. Volumes, surfaces, points and lines.** The terms *in*, *on*, *near*, and *at* require very little in the way of detailed geometry (see Herskovits 1986, for details on normal use types). There is not even any requirement for particular axes. For something to be *in X*, *X* must have an interior, but nothing more is necessary. In other words, the reference object for *in* needs a form descriptor less specific than any particular generalized cone, something like a "lump" or "blob" that would indicate its capacity to surround or contain. Similarly, *near* and *at* require only that the reference object be bounded in extent; they place no requirements at all on its shape. *On* is slightly more complex: It requires that its reference object possess a surface, whether it be a line (on the border), a surface (on the square), or an object with a boundary that is a line or a surface (a house on the lake or on the hill, respectively).

A minor distinction appears in the contrast between *inside* and *in*. *Inside* is the more specific, and seems to require that its reference object be or contain a bounded enclosure (a negative part or the interior of a hollow volume). Thus, as pointed out by Talmy (1983), one can be either in or inside a cave or a bottle; but one can be only in, not inside, a swimming pool or lake, because these are not conceptualized as enclosures or containers (see sect. 2.9 for other usage distinctions based on convention and idiom).

**2.4.2. Axial structure.** A sizable number of prepositions, such as *on top of*, *under*, *in front of*, *in back of*, and *beside*, make reference to an object's axial structure and its axially determined parts, such as *top*, *bottom*, and *sides* (sect. 1.3). In the case of an object that lacks inherent axes, such as a sphere, axes are contextually imposed (see sect. 2.7.2). But these prepositions can also make use of the reference object's inherent axes. *On top of* and *under* project a region from whichever directed axis is vertical in the object's normal orientation. *In front of*, *in back of*, and *behind* make use of the directed horizontal front-to-back axis; *beside* and *alongside* make use of a horizontal axis perpendicular to the front-to-back axis. For the purposes of these prepositions, it does not matter whether the axes in question are generating axes or orienting axes (see sect. 1.3).

As mentioned above, *along* requires its reference object to be basically linear and horizontal. Its partner *across* appears to require its reference object to be or to have a surface with sides, so that one can go across, "from one side to the other." The best case appears to involve opposite sides, although adults accept almost all cases where one object intersects two segments of another object (e.g., two arcs of a circle or two adjacent sides of a rectangle, see Williams 1991). Just in case the reference object has a significant linear elongation, the *sides* are distinguished from the *ends:* A square table has four sides, but a long rectangular table has two sides and two ends. In such a case, *across* appears to pertain specifically to the sides and not to the ends, so that *across the rectangular table* describes a region that traverses the table's shorter dimension.[13] Nonrectilinear objects in this framework tend to be schematized as though they were rectilinear, so that, with respect to *across*, a round table behaves like a square table and an oval table like a rectangular one.

A further restriction on *across* is that, like *along*, it describes a horizontally oriented region. For example, one draws a line across a blackboard in the horizontal direction, not the vertical. This follows from the stipulation that *across* pertains to the sides of the object, which are normally the boundaries of a horizontal axis.

**2.4.3. Quantity.** A different sort of restriction on the reference objects appears in the distinction among the prepositions *between*, *among*, and *amidst*. For *between*, the reference object is not a single object, but rather a pair. In the case of *among* and *amidst*, the reference object is an aggregate (or collection of objects), as in *among the people* or *amidst the waves*.

To sum up, the restrictions placed on the form of the reference object by expressions for spatial relations are not at all severe, compared to the potential complexity of objects themselves. At most, these restrictions appeal to the very gross geometry of the coarsest level of representation of the object – whether it is a container or a surface. In addition, the object's axial structure plays a crucial role, that is, whether the object is relatively elongated, whether the elongation is horizontal, and whether it has sides.[14]

### 2.5. *Figure objects*

There are even fewer constraints on figure objects than on reference objects. For the most part, the figure object needs no geometric specification at all; Talmy (1983) suggests that the figure is usually conceived as "point-like." We have found only two specifications of figural geometry for the English prepositions: axial structure and quantity.

**2.5.1. Axial structure.** Four prepositions, *along*, *in line with*, *across*, and *around*, express spatial relations between the reference object and the linear axis of the figure object. For example, consider "The road is along the river." This specifies that the main axis of the figure object (the road) is parallel to (as well as horizontally proximate to) the main axis of the river. If the figure object is an aggregate, as in "The trees are along the river," this aggregate is preferentially understood as forming a virtual object with an axis that is parallel to the main axis of the river (see Figure 8). The adjectival form *parallel to* places constraints similar to those of *along*. (Note, however, that not all uses of *along* impose this linearity condition, in particular, if the figure has no main horizontal axis, as in "The tree is along the river." Similarly, if the figure object is in motion, as in "the dog loped along the river," it is the trajectory of the figure rather than the figure itself that must be conceptualized as linear and parallel to the main axis of the river.)

The compound preposition *in line with*, like *along*, requires a linear reference object. Unlike *along*, however, the reference object need not be horizontal; moreover, it can be a pair of objects or an aggregate that forms a linear virtual object, so that one can say "The house is in line with the trees." If the figure object is not linear (e.g., the house in the previous sentence), it is simply understood as lying on the axis of the reference object or the extension of this axis at any distance. If the figure object has a linear axis, however, as in "The road is in line with the trees," this axis must be aligned with that of the reference object.

*Across*, as mentioned above, involves a linear region that goes from one side to the other of the reference object. Various senses of *across* locate the figure object differently with respect to this region. The two most relevant ones place the figure object within the region, where it is either (1) linear and coaxial with this region "the stick lay across the road"), or (2) distributed along the axis of the region ("The trees extend across the field"). As with *along*, other senses do not require a linear figure. For example, one sense (3) places the figure on the other side of the region in relation to the observer or a secondary reference object ("Bill is across the road (from here)"). Another (4) expresses a figure moving along the axis of the region ("Bill ran across the road").[15]

*Around* also has a number of variants. Other than the one that means roughly *near* ("There are lots of trees around here"), they all specify a hollow region with an interior that contains the reference object. The constraint on the figure is that it occupies this region, either as a linear object surrounding a two-dimensional reference object ("The road goes around the city") or as a shell or thickened surface surrounding a three-dimensional reference object ("There is chocolate around the core of the candy"). A distributed figure object is again acceptable ("There are trees around the house"). As with *along* and *across*, a moving figure need not be a line or a surface; it may either circumnavigate the reference object (go all the way around) or detour around it.

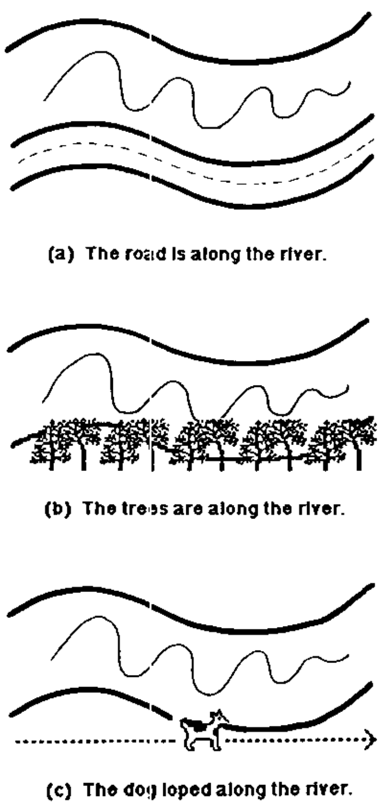



Figure 8. The preposition *along* requires a linear figure and reference object. In (a), both figure and ground are ribbonlike; in (b), the aggregate of trees forms the linear figure; and in (c), the dog's path is the figure that is linear.

**2.5.2. Quantity.** A different class of prepositions requires the figure object to be distributed in space, either as a substance or as an aggregate. Consider: "There was water all over the floor." *All over* specifies a figure object (water) distributed over and in contact with the entire extent of the surface of the reference object. In "There were raisins throughout the pudding," *throughout* specifies an aggregate figure object (raisins) more or less evenly distributed in the volume of the reference object. Thus, *all over* and *throughout* are "distributive" forms of the spatial relations normally expressed as *on* and *in*, respectively. *All along*, *all around*, and *all across* are similar distributive forms corresponding to the prepositions *along*, *around*, and *across*.

### 2.6. *Summary*

In comparison to the rich description of objects as category members (named by count nouns), most of the geometrical distinctions among objects are disregarded when specifying their role as either figure or reference object in a spatial relation. A reference object can be schematized as a point, a container, or a surface, as a unit with axial structure, or as a single versus aggregate entity. No more detail is necessary. Similarly, a figure object is schematized at most as either a simple lump or blob (with no geometric structure whatsoever), a unit with axial structure along at most one of its dimensions, or a single versus distributed entity. Further internal complexity of the object, describable in terms of componential configuration or object parts, is simply irrelevant.

Before turning to a discussion of regions, we should note some examples from other languages that might appear to conflict with the description set forth above. These concern spatial verbs with an application that is restricted to figure or reference objects belonging to a certain object category or to those possessing specific configurational properties. For example, Talmy (1983) describes Atsugewi, a Native American language of California, in which there are roughly a dozen verb suffixes that mark distinctions finer than the English preposition *into*. The distinctions concern the geometry of the reference object, with separate suffixes for "into an aggregate," "into a gravitic container" (e.g., basket, pocket), "into an areal enclosure" (e.g., a field or corral), "over the rim into a volume" (e.g., gopher hole or mouth), and so on. English prepositions distinguish a few such cases; for example, *inside* can be used only for enclosures.

There are also spatial verbs that differentiate according to figure object type. Talmy (1985) describes a set of spatial verbs in Atsugewi that have restrictions on the figure object, for example, "for a small shiny spherical object to move/be-located," "for a slimy lumpish object to move/be-located," "for runny icky material to move/be-located," and so on. Analogous English verbs include "to spit" and "to rain." Similar distinctions are reported by Bowerman (1991) in her description of Tzeltal. There, spatial relationships are apparently expressed with "closed-class positional" verbs that classify on the basis of the figure. Thus, there are separate terms for the following types "to be located": (for) a bowl-shaped figure (to be located), a narrow-mouthed container in upright position, an inverted object with flat surface down, a small sphere, a large sphere, and things in a bulging bag. Analogous distinctions in English might be verbs like *smear* or *pin*, which describe kinds of attachment restricted to different classes of figural objects: The kinds of things you can smear on the wall are quite different from those you can pin on the wall.

These cases appear to encode more detailed geometry in the figure and reference object than do English prepositions, although they still do not encode as much detail as is found in object names. From our perspective, these cases are somewhat different because they involve verbs that are open class and clearly encode more than just location. It is well known that verbs impose all kinds of restrictions: *Drink* applies only to ingestion of liquids, *disperse* applies only to aggregates, *diagonalize* applies only to matrices, and the German verb *fressen* can be predicated (nonironically) only of animals. However, the cross-linguistic evidence does offer interesting challenges, to which we return after presenting the remaining evidence and our hypotheses.

### 2.7. Constraints on regions

Recall that the third element of a linguistically expressed spatial relation is the *region* (of the reference object), in which the figure is located. A spatial preposition can be thought of as a function that determines the relevant region of the reference object. For example, the cat can be said to be *on* the mat just in case it is located in contact with the uppermost surface of the mat; that is, *on* maps the reference object (mat) into the region consisting of its (upper) surface. The bicycle is *near* the house just in case it is located within a region surrounding the house and extending up to some critical distance; *near* thus maps the reference object into such a region.

The notion of region appears necessary to describe the meanings of spatial prepositions (see e.g., Miller & Johnson-Laird 1976). It is also necessary to describe people's nonlinguistic representation of space. Kevin Lynch (1960) was one of the first to show the importance of components such as "nodes" and "districts," which are not evidently circumscribed in physical space but serve as foundational elements in people's conceptualization of the cities in which they live.

Several studies by cognitive psychologists have pointed to the importance of regions in organizing people's knowledge of space. Stephens and Coup (1978) showed that people's distance and direction judgments were not isomorphic to physical space but rather were biased by the organization of individual cities into larger regions defined by the states in which they were located. Hirtle and Jonides (1985) showed that people tend to judge distances within a particular region as smaller than they do when the same distances cross regions.

Given that regions are part of our spatial representations, we can ask how many ways spatial prepositions can encode regions. What kinds of distinctions among regions are encoded by the set of spatial prepositions? In principle, they could represent regions in acute detail, for example, in terms of precise distance and direction from the reference object, using coordinate systems and the specialized domain of measurement terms. This degree of precision is surely represented in the spatial cognitive system and is used in some cases, although not in all cases as evidence from nonlinguistic studies such as Hirtle and Jonides suggests (see sect. 3 for further discussion).

Within language, the very restricted number of spatial prepositions suggests that precise location is not encoded in any individual term. And, as it turns out, the spatial relations expressed in English factor into just a few independent features that combine to produce some of the complexity of the system. The principal features include several degrees of distance and several kinds of direction.

**2.7.1. Relative distance.** This factor concerns the distance between figure and reference object. The most salient fact is that distance is digitized into several discrete categories. There appear to be four levels described by English prepositions: (1) location in the region interior to the reference object (*in*, *inside*); (2) location in the region exterior to the reference object but in contact with it (*on*, *against*); (3) location in the region proximate to the reference object (*near*); and (4) location distant from the reference object (*far* and perhaps *beyond*).

Some languages provide additional values for the distance feature, for example, "not near but within reach" or "not near but visible." One such language is Korean (Soo-Won Kim, personal communication), in which the expressions *yup* and *kiyut* both translate as "near," but *yup* is confined to more immediately proximate cases. (Other examples are cited by Anderson and Keenan [1985] in connection with systems of spatial deixis corresponding to English *here* and *there*.) On the other hand, some languages have prepositions that collapse two adjacent values of distance in English. For example, English *in* and *on* can both be translated by Spanish *en*. Thus, English represents an intermediate degree of complexity in the distance parameter. To our knowledge, no language encodes more than five or six levels of distance, although we are able to represent distance at much finer levels for other cognitive purposes.

At least two of the degrees of relative distance found in English have a corresponding "negative," which actually means "farther away from the reference object than." These are "farther away than the interior" (*out of, outside*), and "farther away than in contact" (*off of*). The pair *near* and *far from* might also form such a contrast.

Several prepositions involve the distance feature in a less obvious way. As mentioned above, the reference object for *among* is an aggregation of objects, which together define a group or virtual object that contains them all. The figure object is then specified as interior to this virtual object. The case of *between* is similar, except that the virtual object is the minimal space bounded by the pair of reference objects.

**2.7.2. Direction.** Direction of the figure object from the reference object provides the second key parameter in specifying spatial relations. Crucially, the entire set of directions derives from the axial structure of the reference object (or its axially determined parts): The three principal axes can be viewed as extending from the center of the reference object to provide six possible directions. Centered around each half-axis is a region that defines the acceptable space for different prepositions.

Regions determined by the vertical axis are given (in the canonical case) by gravitation, defining *over, above, under, below,* and *beneath*. Orthogonal to gravitation is the horizontal plane, which helps define *beside, by, alongside,* and *next to*. To see that the horizontal is crucial to these terms, notice that if a bird is beside, by, alongside, or next to a house, it must not be on the roof or flying overhead: It must be in proximity to the house and no higher than the house. Thus, these prepositions designate the relation "proximate to the reference object in the horizontal direction." If, in addition, the reference object has inherent axes that distinguish front and back from sides, the terms *beside, by, alongside,* and *next to* tend to mean "horizontally proximate to the *sides* of the reference object." For example, Bill is not *beside* me if I am facing him.

Similarly, if the reference object has an axis that distinguishes an inherent front and back, *in front of* can mean "horizontally proximate to the inherent front of the reference object," and *in back of* can mean "horizontally proximate to the inherent back of the reference object." However, an alternative interpretation of these prepositions results from *contextually* assigning a front-to-back axis to the reference object: The front is the surface facing the speaker (or addressee) and the back is the surface opposite. In this case *in front of* and *in back of* mean "horizontally proximate to the contextual front/back of the reference object." A parallel ambiguity occurs with *on top of*: If a flagpole is lying on its side, one can paint the ball on top of it (referring to the inherent top), or one can sit on top of it (referring to the contextually determined top, in this case a long horizontal surface). Levelt (1983) and Olson and Bialystok (1983) describe the many complexities that arise from assigning different frames of reference.

Not all spatial expressions involving axes leave this choice of reference system open. *On the top of*, by contrast with *on top of*, refers only to the inherent top of the reference object (presumably because it contains the full noun phrase *the top*). *Beyond*, by contrast with *behind*, refers only to the region projected to the contextually determined rear of the reference object. Whichever type of reference system is used, however, its structure still depends on analyzing the reference object into its three principal axes.

**2.7.3. Combinations of distance and direction.** Further distinctions can be derived among prepositions by various combinations of distance and direction. For example, compare *over, above,* and *on top of*. *Over* is indifferent to contact versus noncontact: A cloth may be put over a table (contact), and clouds may fly over a city (noncontact). *Above*, however, specifies noncontact: Though clouds may fly above a city, one can only put a cloth above a table by putting it on a higher shelf. Finally, *on top of* strongly favors a contact reading.

*In back of* and *behind*, which share directionality, also differ in distance. A tree may be right behind (proximal), way behind (distal), or right in back of a house, but "The tree is way in back of the house" sounds odd or colloquial (to us, anyway). The standard use of *in back of* seems to be restricted to proximal distance (and possibly contact), whereas *behind* and the colloquial *in back of* are unrestricted. In any case, distance can be combined with direction to yield finer distinctions.

For a somewhat different case, to move *up* or *down* a mountain, tree, or wall is to move in an upward/downward direction while maintaining contact with (or, marginally, proximity to) the surface of the reference object.

Sergey Avrutin (personal communication) has pointed out that the conceptual features of distance and direction can be used to predict the case-marking pattern of Russian place prepositions. According to his analysis, the two closest grades of distance – interior and contact – assign the prepositional case to their objects; prepositions that involve an axis-based directional feature (*over, under, behind,* etc.) assign the instrumental case; the remaining prepositions (*at, near, close to,* etc.) assign the genitive case.[16] Thus, these features are grammatically as well as conceptually significant.

**2.7.4. Visibility and occlusion.** A further distinction appears to be subsidiary to (but to some degree independent of) the distance and direction distinctions. A case in English where this distinction is evident is in speaking of paint on a wall being *on top of* or *underneath* the wallpaper. Here *on top of* evidently means "in contact with visible surface," whatever its orientation, and *underneath* means "in contact with the surface opposite the visible surface." (Notice, by the way, that one cannot speak of the bottom of the wallpaper in this context; not all the words of vertical orientation generalize to this use.) Vandeloise (1986) argues that occlusion of the reference object is the main relation expressed by French *devant* ("in front of"). We would not go quite so far, but we believe this criterion does play a secondary role, possibly forming a preference rule system (Jackendoff 1983) with the directional criteria.

Visibility and occlusion may well play a role in children's learning of spatial prepositions. Johnston (1984) found that 3-year-olds understood *behind* to mean "occluded from sight." For example, they agreed that X was behind Y only when Y was an object large enough to occlude X. In contrast, adults agree that X is behind Y whatever their relative sizes, as long as the figure object is aligned properly with the reference object's front-back axis. Combined with the evidence from French, this suggests that visibility versus occlusion may indeed be an additional distinction in spatial relationships.

### 2.8. Spatial relations defining paths

In addition to regions, language expresses another spatial category of *paths* or *trajectories* to specify a figure's motion ("The bird flew to the house") or orientation ("The sign points to New York"). There are only a few simple ways of constructing trajectories, none of which draw any further on the geometry of the figure or reference objects than we already have. Once again, the main geometric property involved is axial structure.

One class of paths specifies the figure's motion in terms of its own inherent horizontal axes: *forward, backward,* and *sideways*. Another specifies change of the figure object's orientation, again in terms of its own axes: turn *around*, turn *over*, turn *left*, and turn *right*. Another class draws on the axial structure of the earth: These terms are the environmentally oriented directions *up, down, north, south, east*, and *west*.

The largest class of paths, however, is constructed from the class of regions by attaching one of a set of five operators: *via, to, toward, from,* and *away from* (Jackendoff 1983). The operator "via" creates a path that passes through a region that is in turn defined by the parameters described in section 2.7. For example, to *go through a room* involves a path that at some point involves being *in* the room; to *run by the house* is to traverse a path that at some point involves being *near* the house; to *walk under a bridge* is to follow a route that at some point involves being *under* the bridge. If the region in question is linear for example, as in *along, across,* and *around,* the "via" path is coaxial with the region, so that *going along X* involves moving parallel to the axis of X.

The operator "to" creates a path that terminates at the region in question. For example, *to X* expresses a path that terminates *at X*. *Into X* and *onto X* express paths that terminate *in X* and *on X*, respectively. Similarly, the operator "toward" constructs a path that would terminate at the region if extended, but does not in fact reach the region. To *go toward X* is therefore to undergo a motion that if extended would terminate *at X*.

The operator "from" is just the reverse of "to," perhaps forming a pair similar to *in/out* and *on/off* (Gruber 1976): It constructs a path that begins at the region in question. Examples are: "The bird emerged from under the table" and "The train came from inside the terminal." The operator "away from" is the reverse of "toward": "Bill ran away from the explosion" describes him as traversing a path that if extended backward would begin at the explosion."

### 2.9. Other factors

The discussion above has presented more than just a sample of the spatial relations expressed by English prepositions. Rather, we believe we have extracted from the complicated grammatical and pragmatic facts about prepositions all the purely spatial information they are capable of encoding. This section will list briefly some of the complications that remain, none of which involves geometric properties per se.

First, there are uses of prepositions that involve special functional situations. Herskovits (1986) points out that to be *at a desk* or *at a sink* usually implies more than being located close to it; one is probably performing characteristic actions, such as writing at the desk or washing at the sink. For another case involving *at*, to *throw a ball at X* involves more than *throwing it toward X*, namely, an intention to hit X. This difference accounts for the contrasts in (6):

(6) a. Bill threw the ball toward/?at Harry without meaning to hit him.
  b. Bill shot at/?toward Harry.

Other special situations involve conventionalized conceptualizations of the reference object. For example, when traveling, one is *in a bus* or *on a bus* but only *in*, not *on, a car*. It seems that in English, large vehicles (buses, yachts, trains, large airplanes) are conceptualized either as containers that one is *in* or sorts of platforms that one is *on*, but small vehicles (cars, rowboats, small airplanes) are only conceptualized as containers. The choices here show that alternative conceptualizations are available; presumably, the particular choice has less to do with principles of spatial representation than with historical and pragmatic issues. For a somewhat different case, a container can be conceptualized either in terms of the volume it surrounds or in terms of the body of its substance, so we can speak of either *the water in the cup* or *the crack in the cup*.

Some uses of prepositions appear to involve forces exerted between the figure and the reference object. For example, the preposition *on* is frequently said to involve support by the reference object (Cienki 1988; Herskovits 1986). This is not always the case, because we can speak of *the fly on the ceiling;* but it may be a default interpretation. According to Bowerman (1989), the Dutch preposition *aan* also involves support or attachment, specifically a figure object hanging or projecting from a reference object that is something other than a horizontal surface (for example leaves *aan* a twig, a coathook *aan* a wall, clothes *aan* a line). The English preposition *against*, as in "Bill leaned against the wall" or "The tree fell against the house," describes contact with exertion of force, usually in a horizontal or oblique direction. Among expressions of path, there is a reading of *into* found in "The car ran into the pole," which means not traversal to the interior of the reference object, but rather coming into contact with the reference object with considerable force.

Bowerman (1989) describes other uses of force-dynamic properties encoded in Korean spatial verbs. It appears that the verb *kki-ta*, roughly "put in, put together," applies to situations in which the figure object fits fairly tightly into or around the reference object, such as a ring on a finger, a hand in a glove, a lid on a jar, and a button in a buttonhole. (The English verb *insert* appears to cover part but not all of the same semantic territory; the verb *fit* also appears close in meaning.) The verb *ppay-ta* describes the removal of the figure from a reference object with which it has been configured in this fashion. This application of force-dynamic properties also occurs in English verbs such as *clasp, snap,* and *impress* (see Pinker 1989 for discussion). The Korean verbs again make little reference to the detailed geometry of the figure and the reference object, however, other than the fact that there is a match between a positive part of one and a negative part of the other.

Beyond this sort of complication, most of the complexity of English prepositions appears to involve (1) how spatial configurations that are nonstereotypical or ambiguous are forced into the expressions available in the language, (2) how particular prepositions are extended from core place meanings to different sorts of related paths and places (for example, the variants of *across* mentioned above), (3) how preposition meanings are extended to nonspatial domains such as time and possession, and (4) how prepositions are used as purely grammatical markers ("Bill believes **in** capitalism," "The letter was received **by** Bill," and "a picture **of** Bill"). Extended discussions appear in Brugman (1981), Cienki (1988), Herskovits (1986), Jackendoff (1983; 1990), Lakoff (1987, Chapter II.2), Miller and Johnson-Laird (1976), and Vandeloise (1986).[18] Our goal here has been to use linguistic research to see through all this complication in order to find the specific characteristics of spatial representation necessary for the range of preposition meanings. These characteristics have proven to be quite limited.

### 2.10. Summary

The description of figures, reference objects, and regions in English (both places and trajectories) recruit just a few geometric properties and distinctions. The geometry of figure objects specifies at most a single axis, whereas that of reference objects specifies at most three principal axes of the object. The regions relevant to describing places make further use of the axes, adding qualitative distinctions pertaining to the distance and direction of the figure from the reference object. The regions relevant to describing trajectories or object motions then draw on these place descriptions, adding operators that specify the location of the path relative to a given place, and where that path begins and ends. Table 2 provides a summary of the factors discussed here.

Table 2. *Features of spatial relations (with sample prepositions)*

| | |
|---|---|
| ***Reference object geometry*** | |
| Volumes, surfaces, and lines: | in, on, near, at, inside |
| One or two axes | |
| Vertical: | on top of |
| Horizontal: | in front of, in back of, beside, along, across |
| Quantity: | between, among, amidst |
| ***Figure object geometry*** | |
| One axis: | along, across, around, in line with |
| Distributed figure (substance or aggregate): | all over, throughout, all along, all around, all across |
| ***Relation of region to reference object*** | |
| Relative distance | |
| Interior: | in, inside, throughout |
| Contact: | on, all over |
| Proximal: | near, all around |
| Distal: | far |
| "Negatives" | |
| Beyond interior: | out of |
| Beyond contact: | off of |
| Beyond proximal: | far |
| Direction | |
| Vertical: | over, above, under, below, beneath |
| Horizontal | |
| Side-to-side: | beside, by, alongside, next to |
| Front-to-back: | in front of, ahead of, in back of, behind, beyond |
| Choice of axis system: | |
| Inherent: | on the top of, in front of, ahead of, behind |
| Contextual: | on top of, in front of, behind, beyond |
| Visibility and occlusion: | on top of, underneath |
| ***Paths (trajectories)*** | |
| Earth-oriented: | up, down, east, west, north, south |
| Figure object axis-oriented: | forward, ahead, backward, sideways, left, right |
| Operators on regions | |
| Via: | through (= via inside), along (= via along) |
| To: | to, into (= to in), onto (= to on) |
| Toward: | toward |
| From: | from, from under, from inside |
| Away from: | away, away from |

## 3. Why do spatial prepositions make so little use of object shape?

The picture that emerges from this brief overview (if it is anywhere near complete, as we believe it is) is that the geometric descriptions relevant to words describing *what* an object is are very different from those describing *where* an object is. This is initially suggested by the fact that spatial prepositions in English are quite few in number relative to the class of object names (even taking polysemy into account, just compare the number of nouns mentioned in sect. 1 with essentially the entire repertoire of English spatial prepositions as described in sect. 2).

It is more significant that our investigation shows severe constraints on the ways in which the meanings of spatial prepositions can invoke object geometries. It is as if the spatial relations expressed by prepositions filter object descriptions, removing much of the detail of object shape and preserving only certain key properties, primarily the boundedness, surface, or volumetric nature of an object and its axial structure. Because of this extreme limitation, more complex meanings such as those ascribed to the hypothetical words *sprough, betwaft,* and *plin* are simply not available as possible meanings in the language's basic stock of prepositions.

This is not to say that one cannot express such meanings; we clearly can, using combinatorial expressions. Our claim is that such meanings will arise either through derivation (e.g., *aboard*), as technical terms (e.g., *abaft*), or by composition with open class words (e.g., through a *cigar*, between the *nose and mouth*, *3 meters* from the door, etc.).

One could stop here and accept this result as an interesting fact about language. But to us it cries out for explanation. We see two possible approaches. The first, which we will call the Design of Language Hypothesis, claims that the limitation on spatial relations expressible in language is indeed just a fact about language, and that spatial cognition is just generally much richer in the spatial relations it can encode. According to the other approach, which we will call the Design of Spatial Representation Hypothesis, this limitation in language reflects a deeper constraint on how spatial cognition encodes the relations among objects. We will argue that both factors contribute to the relative scarcity of preposition meanings.

### 3.1. Design of Language Hypothesis

According to the language design hypothesis, spatial representation can itself encode a rich range of spatial relations, making use of detailed properties of object shapes. Most of these are "invisible" to the language faculty, however, and therefore neutralized or leveled out in the translation into linguistic format. This hypothesis implies that filtering should apply equally to the two domains of object and place and that it might be a design feature required by any system that must collapse numerous complex distinctions into a finite set of elements.[19]

There is abundant evidence that language does indeed filter representations of spatial properties and relationships. For instance, language does not typically represent the dimensions of objects in analog fashion, but rather digitizes them. Thus, dimensional adjectives such as *big/small*, *thick/thin*, and *tall/short* refer to continuous dimensions of size, but the linguistic terms bifurcate these dimensions into pairs of relative contrasting terms. Such binary (or ternary) relative contrasts are characteristic of most adjective domains; they occur across most languages and may be more natural than absolute contrasts in first language learning (Landau & Gleitman 1985).

Especially compelling evidence for this binary/relative constraint is provided by Newport (1988) and Supalla (1990) from the study of American Sign Language. As Newport points out, the manual mode affords a straightforward means of expressing continuous quantities in an analog fashion. For example, the sign for *slow* is represented by moving the finger slowly along the opposite forearm. It would be feasible to "mimic" different rates in a much more precise fashion, perhaps even matching the object's actual rate of movement in one's sign. ASL, however, like spoken languages, expresses different speeds using only two or three distinctions, qualitatively categorizing the physically continuous dimension despite the fact that the modality affords a more detailed representation of speed.

As another argument for filtering, notice that the spatial representations recruited by the motor system must be much richer than those appearing in language. Consider the task of inserting one's hand through a narrow slot. One might accomplish this by predicting from the beginning of the reach the hand's exact angle of orientation relative to the slot. Or one might initially make a rough ballistic reach toward the slot and then modify it using visual feedback about error. In either case, precise spatial representations of distance, angle, and joint position must be available to be converted into equally precise muscular forces. A similar situation exists for the task of throwing a ball to someone, catching a ball, and, in general, any act of navigation based on either visual or haptic-kinesthetic perception (see Gallistel 1990, for a discussion of the geometries underlying navigation).

Such metric representations of space emerge spontaneously early in life; they are not the product of lengthy formal tutoring and are therefore probably part of our biological endowment. For example, evidence from von Hofsten (1980) suggests that even 4-month-old infants can "catch" both stable and moving objects successfully under some circumstances. It is significant that the infant's initial angle of trajectory is mapped more closely to the object's (predicted) *final* position than to its actual position at the initiation of the reach. Other evidence shows that by around 2 or 3 years of age, children can use knowledge of metric properties of spatial layouts to search for objects, guide navigation, and use maps (DeLoache 1987; Landau 1986; Landau et al. 1984; Rieser & Heiman 1982).

Yet, despite its importance to motor control and navigation and its naturalness in human perceptual and cognitive development, precise metric information is simply not encoded in the language's stock of spatial terms, a point emphasized by Talmy (1983). It is possible to be precise in expressing distances and orientations, but to do so, one must invoke a culturally stipulated system of measurement that operates by counting units such as meters or degrees (*go 30 meters, turn 30 degrees*).

Could this filtering account for the severe limitation in the ways spatial prepositions can take object shape into account? It cannot be the whole story, for filtering also takes place in translating object shape descriptions into language. For one thing, not everything called *dog* is precisely the same shape; that is why a similarity metric is needed for categorizing objects. And on the nonlinguistic side there is abundant evidence that humans can encode detailed aspects of shape that do not appear in language.

For instance, one can recognize with great accuracy complicated contours and surface patterns (and this improves with perceptual learning (Gibson 1969). To succeed in tasks such as mental rotation and composition, we apparently represent in detail the metric composition and relationships among object parts (Cooper 1989), yet these detailed descriptions are very hard to describe to someone else in words. Imagine, for example, trying to describe the Shepard-Metzler objects or the Attneave figures used in visual image rotation experiments (Shepard & Cooper 1982), the pattern of stripes on a particular zebra, the shape of a violin, or one's mother's chin. What actually happens when we try to describe complex figures is that we describe them in terms of their parts, or with allusions to familiar objects, such as "spider" for a thing with "legs" projecting from a round center (Fussell & Krauss 1989).

These difficulties in describing precise shape are exacerbated by the absence of linguistic terms for describing exact sizes of objects (again apart from a culturally stipulated system of measurement). That is, the filtering out of metric information that occurs in the expression of spatial relations also occurs in the expression of objects.

In short, language does not convey all the representational richness we have for encoding either locations or objects. (That is why a picture is worth a thousand words.) However, this filtering out of metric information does not by itself explain why the encodings of object shape are limited in the particular way they are when spatial relations are at issue. If the problem were merely a design need to filter, prepositional meanings could filter objects coarsely on many different properties, by representing just two values of brightness or color, size, texture, animacy, and so on. But prepositions filter objects in particular ways, preserving just global and axis-based structure. We are therefore still left with the question of why objects that are being *named* are differentiated in relatively complex geometric terms, whereas objects that are being *located* and the regions in which they are located are treated in terms of relatively simple schematic geometric descriptions. What accounts for this difference?

### 3.2. Design of Spatial Representation Hypothesis: The "what" and "where" systems

One possibility is that the disparity may be inherent in the spatial representations underlying language. According to this hypothesis, spatial representation is relatively rich in its possibilities for describing object shape, but relatively limited in the way it can use object shape to encode spatial relations. If this is the case, the disparity observed in language is a consequence of the disparity in the spatial representations that language encodes.

We conjecture that this is indeed the case, and that the disparity in spatial representation is partly reflected in some basic organizational facts about the human brain, in particular, that it arises in part from a functional bifurcation of the system of spatial representation (perhaps into "submodules" in the sense of Fodor 1983 [see also *BBS* multiple book review of Fodor's *The modularity of mind*, *BBS* 8(1) 1985], as refined by Jackendoff 1987b, Chapter 12). One part of the system is devoted primarily to objects and their identification (mostly by shape), the other to locating objects in space relative to each other and to the observer. The expressive power of the system of nouns that identify objects is linked to the shape identification submodule; the expressive power of the spatial preposition system is linked to the submodule governing the location of objects relative to each other.

**3.2.1. Nonlinguistic evidence.** Our conjecture finds interesting correlative support in neurological, psychological, and computational evidence.

Neurological evidence reviewed by Ungerleider and Mishkin (1982), building on previous work by Schneider; (1969; see also Ingle et al. 1967), suggests that the brains of monkeys contain separate areas specialized for object identification (the "what" system) and object location (the "where" system). These specializations have been inferred from selective deficits following damage to different areas of the cortex. For example, in one kind of task, an animal might learn to find food when it is hidden in one of a pair of distinctively patterned objects, regardless of its location. Such a task clearly requires the ability to select on the basis of pattern (or, in some cases, shape); it is selectively impaired by damage to the inferior temporal cortex.

On another kind of task, an animal might be required to find food when it is hidden in one of a pair of identical objects that is nearer a particular reference object. This requires the ability to select on the basis of position and it is selectively impaired by damage to the posterior parietal cortex. More generally, damage to the posterior parietal cortex impairs route following, reaching for objects, and using landmarks to locate objects.

Ungerleider and Mishkin also review evidence suggesting that the neurons in the two cortical regions in question have distinctly different receptive field properties. Most neurons in the inferior temporal cortex (the "what" system) are driven by complex sets of features and their receptive field is large: More than half have a bilateral receptive field and all include the fovea. This would mean that information about the position of any particular object shape being processed by these cells would be lost – that is, the position of the pattern would simply not be represented. In contrast, neurons in the parietal lobe (the "where" system) are not sensitive to stimulus features and most do not include the fovea in their receptive field.

Although Ungerleider and Mishkin's evidence is derived primarily from lesion studies on animals, there is converging evidence from human psychophysical studies that there may be two distinct streams of visual processing: the "parvo cellular" system, specialized for detecting color and detailed object shape, and the "magno cellular" system, which is color-blind but specialized for detecting motion, depth, and location. These streams are segregated at relatively low levels of the visual system and the segregation appears to become more pronounced at higher levels, providing converging evidence for the what and where distinction (Livingstone & Hubel 1989; but see Van Essen et al. 1992 for a suggestion that the systems overlap considerably at early stages of processing).

Human clinical evidence appears to support the what/where distinction as well. Farah et al. (1988) document a case in which bilateral damage to inferior temporal areas with sparing of parietal regions produced a deficit in a wide range of tasks involving shape recognition but preserved normal performance in tasks involving object localization and spatial relations.[20] Evidence from Levine et al. (1985) suggests that the imagery system might also carry these distinctions. One of their patients could imagine object shapes but not spatial relationships or object layouts; another could imagine spatial layouts, but not the shapes of individual objects.

Finally, evidence from formal modeling of simple learning systems supports the distinction. Rueckl et al. (1988) found that in a PDP model of a very simple visual system a certain degree of extra efficiency accrues to a system that strongly separates computation of the "what" and "where" functions as long as both subsystems have sufficient computational resources. In their study, Rueckl et al. used a stimulus space containing only nine different shapes, each of which could occur in nine partially overlapping locations. Within this tiny system it was found that the optimal allocation of resources between the "what" and "where" systems used over three times as lated system of measurement). That is, the filtering out of metric information that occurs in the expression of spatial relations also occurs in the expression of objects.

In short, language does not convey all the representational richness we have for encoding either locations or objects. (That is why a picture is worth a thousand words.) However, this filtering out of metric information does not by itself explain why the encodings of object shape are limited in the particular way they are when spatial relations are at issue. If the problem were merely a design need to filter, prepositional meanings could filter objects coarsely on many different properties, by representing just two values of brightness or color, size, texture, animacy, and so on. But prepositions filter objects in particular ways, preserving just global and axis-based structure. We are therefore still left with the question of why objects that are being *named* are differentiated in relatively complex geometric terms, whereas objects that are being *located* and the regions in which they are located are treated in terms of relatively simple schematic geometric descriptions. What accounts for this difference?

### 3.2. Design of Spatial Representation Hypothesis: The "what" and "where" systems

One possibility is that the disparity may be inherent in the spatial representations underlying language. According to this hypothesis, spatial representation is relatively rich in its possibilities for describing object shape, but relatively limited in the way it can use object shape to encode spatial relations. If this is the case, the disparity observed in language is a consequence of the disparity in the spatial representations that language encodes.

We conjecture that this is indeed the case, and that the disparity in spatial representation is partly reflected in some basic organizational facts about the human brain, in particular, that it arises in part from a functional bifurcation of the system of spatial representation (perhaps into "submodules" in the sense of Fodor 1983 [see also *BBS* multiple book review of Fodor's *The modularity of mind*, *BBS* 8(1) 1985], as refined by Jackendoff 1987b, Chapter 12). One part of the system is devoted primarily to objects and their identification (mostly by shape), the other to locating objects in space relative to each other and to the observer. The expressive power of the system of nouns that identify objects is linked to the shape identification submodule; the expressive power of the spatial preposition system is linked to the submodule governing the location of objects relative to each other.

**3.2.1. Nonlinguistic evidence.** Our conjecture finds interesting correlative support in neurological, psychological, and computational evidence.

Neurological evidence reviewed by Ungerleider and Mishkin (1982), building on previous work by Schneider; (1969; see also Ingle et al. 1967), suggests that the brains of monkeys contain separate areas specialized for object identification (the "what" system) and object location (the "where" system). These specializations have been inferred from selective deficits following damage to different areas of the cortex. For example, in one kind of task, an animal might learn to find food when it is hidden in one of a pair of distinctively patterned objects, regardless of its location. Such a task clearly requires the ability to select on the basis of pattern (or, in some cases, shape); it is selectively impaired by damage to the inferior temporal cortex.

On another kind of task, an animal might be required to find food when it is hidden in one of a pair of identical objects that is nearer a particular reference object. This requires the ability to select on the basis of position and it is selectively impaired by damage to the posterior parietal cortex. More generally, damage to the posterior parietal cortex impairs route following, reaching for objects, and using landmarks to locate objects.

Ungerleider and Mishkin also review evidence suggesting that the neurons in the two cortical regions in question have distinctly different receptive field properties. Most neurons in the inferior temporal cortex (the "what" system) are driven by complex sets of features and their receptive field is large: More than half have a bilateral receptive field and all include the fovea. This would mean that information about the position of any particular object shape being processed by these cells would be lost – that is, the position of the pattern would simply not be represented. In contrast, neurons in the parietal lobe (the "where" system) are not sensitive to stimulus features and most do not include the fovea in their receptive field.

Although Ungerleider and Mishkin's evidence is derived primarily from lesion studies on animals, there is converging evidence from human psychophysical studies that there may be two distinct streams of visual processing: the "parvo cellular" system, specialized for detecting color and detailed object shape, and the "magno cellular" system, which is color-blind but specialized for detecting motion, depth, and location. These streams are segregated at relatively low levels of the visual system and the segregation appears to become more pronounced at higher levels, providing converging evidence for the what and where distinction (Livingstone & Hubel 1989; but see Van Essen et al. 1992 for a suggestion that the systems overlap considerably at early stages of processing).

Human clinical evidence appears to support the what/where distinction as well. Farah et al. (1988) document a case in which bilateral damage to inferior temporal areas with sparing of parietal regions produced a deficit in a wide range of tasks involving shape recognition but preserved normal performance in tasks involving object localization and spatial relations.[20] Evidence from Levine et al. (1985) suggests that the imagery system might also carry these distinctions. One of their patients could imagine object shapes but not spatial relationships or object layouts; another could imagine spatial layouts, but not the shapes of individual objects.

Finally, evidence from formal modeling of simple learning systems supports the distinction. Rueckl et al. (1988) found that in a PDP model of a very simple visual system a certain degree of extra efficiency accrues to a system that strongly separates computation of the "what" and "where" functions as long as both subsystems have sufficient computational resources. In their study, Rueckl et al. used a stimulus space containing only nine different shapes, each of which could occur in nine partially overlapping locations. Within this tiny system it was found that the optimal allocation of resources between the "what" and "where" systems used over three times as many "what" units as "where" units. It is hard to know how this case generalizes to a more realistic system (or to other different learning systems), but the disparity is reminiscent of the one we have found between noun and preposition systems. Rueckl et al. in fact argue that the disparity can only increase as one moves to a more realistic system.

Thus, our Design of Spatial Representation Hypothesis is suggestively supported on a variety of empirical grounds. We say "suggestively," because the evidence described above will no doubt be subject to revision as investigation proceeds. For example, recent evidence challenges the notion that the "what" and "where" systems are as separate and independent as originally thought.[21] Another problem concerns how object and place information is combined. In many cases, object recognition will require the assembly of object parts on the basis of their spatial relations. It is unclear how this would be accomplished if the what and where systems were totally separate. We assume that the solution to these and other problems will be found in increasingly complex models of how these systems work and how they interdigitate. Recent work suggests a considerable degree of "cross-talk" and the possibility that the systems underlying the identifying and locating objects are considerably overlapping at early levels of the visual system (Van Essen et al. 1992). For the time being, we draw attention to the fact that there is evidence for a what and where bifurcation, pointing out what we view as an intriguing parallel to the representation of object and place in language.

**3.2.2. "What" and "where" in spatial language.** How does the "what" and "where" distinction bear on our observations about spatial language – and especially on their relation to the Design of Spatial Representation Hypothesis? Let us consider for a moment the logic of a representation that separates "what" from "where." What information does the "where" system have to encode? At the very least, it must have a space of possible locations and a way to mark which ones are occupied. But this is obviously not enough. It would not do just for the "what" system to know that one is seeing a cat and a dog, and for the "where" system to know that positions A and B of the visual field are occupied: Is the cat at A and the dog at B, or vice versa? To keep track of which objects are where, there must be a liaison between the two representations (see Van Essen et al. 1992 for neurological perspective).

A simple way to accomplish this liaison formally is by coindexing or linking the object representations in the two systems (a similar proposal is suggested by Kosslyn 1990). The "where" system could then encode very rudimentary representations of the objects being located, perhaps as simple as "thing here." Such extremely schematized objects would place only minimal demands on information-bearing capacity in the "where" system. These schematized objects, however, would also be linked to or associated with representations in the "what" system that encode the objects' detailed shape. In other words, the "where" system can get by with including just a little object information, as long as it can link its object tokens to those in the "what" system.[22]

An ordinary-life analogy to the situation in the "where" system is the conventionalized representation in maps. What is at issue in a map is how to navigate through a region (or perform the many tasks requiring object location). A map cannot just specify "wheres": It has to have something to stand in for the objects being located. Typically, these stand-ins are points and lines, with some conventionalized symbols to distinguish different sorts of objects from each other (big cities vs. small cities, main roads vs. subsidiary roads, churches vs. hospitals, etc.). If a map had to distinguish all the objects by their shapes, it would be much more complex and quite possibly unusable.

We are not suggesting that the "where" system necessarily encodes something like an internalized map. The point is only that many of the same design criteria are applicable, in particular, the need to represent objects as tokens in the representation but to compress their encoding by eliminating most information about their form.

Based on the evidence and arguments presented here, we conjecture that the relatively simple shape specifications observed in the prepositional system reveal the extent of detail possible in object descriptions in the "where" system. These details go somewhat beyond "thing here," but not much. In particular, as we have seen, the way the system works is not by locating objects absolutely but by locating one object in terms of another. The little detail there is in the system's shape descriptions is concentrated on the reference object, which defines the space in which the figure is located – and even that is highly restricted. The geometry of the figure object goes beyond "thing here" only in the small class of cases in which the issue is its orientation (*along, across, around*) or its distribution through a region (*all over, throughout*).

**3.2.3. "What" and "where" as universals.** If we are right, then the structure of object and place systems should be correlated with universals in languages and this universal spatial-semantic system should serve as an important constraint on first language learning. Presumably, the universal system would exhibit properties at the grain needed to distinguish among named objects and named places across languages. Each of these conclusions awaits considerable further research, but in the meantime we would like to suggest several hypotheses and lines of investigation that could provide testing ground for our claims.

The Design of Spatial Representation Hypothesis gives rise to three different sorts of predictions. One concerns the universal structure of language. If the properties of object and place language are caused in part by the structure of nonlinguistic representations of object and place then we should find broad similarities in the expression of object and place across languages. Specifically, we should find that object names universally draw on detailed descriptions of object shape, but that terms for location draw on quite sparse descriptions of object shape. Although our discussion has focused on English, a variety of cross-linguistic studies suggest that the broad outlines of our proposal are correct (Bowerman 1989; Cienki 1988; Talmy 1983; Vandeloise 1986).

We have also noted several examples that might provide a challenge to our view (sect. 2.6). These are primarily languages in which a range of spatial verbs restricts their application to figure or reference objects whose geometries are more detailed than those found among English prepositions. We believe there are at least two possibilities consistent with our hypothesis. One is that spatial verbs, which are open class and can clearly express more than just location, are different from English prepositions, which are closed class elements dedicated to expressing location only. For example, there are many spatial verbs in English that concern the motion of something *into* another object, where the reference object is defined quite specifically for each verb: These are denominal verbs such as *to bottle*, *can*, *house*, and so on. The fact that these verbs (or the spatial verbs in other languages) show quite specific restrictions does not invalidate our hypothesis, which is explicitly concerned with those closed class forms that are dedicated to expressing location.

A second possibility is that different languages do in fact have sets of elements dedicated to expressing location only and that these elements do incorporate more detail about figure and reference objects than is found in English. Bowerman's examples from Tzeltal do so (see sect. 2.6), although the degree of detail is still radically less than that captured by object count nouns. The question is not whether any languages exist that include *some* shape information in place terms; as we have seen, even English prepositions include some shape specification of figure and reference object. Rather, the question is what kind of detail and how much exists across languages.

These questions clearly require additional cross-linguistic research.[23] One possibility is that there are typologies among languages, with certain languages (like English) drawing on extremely sparse geometric descriptions and others drawing on some additional degree of detail. We would predict, however, that no natural language exists whose pure locational elements consistently draw on object geometries even close to the level of detail seen in the object naming system.

A second set of predictions concerns language learning. Given the proposed structural constraints on object versus place language, children should come to language learning prepared to attend only to certain properties when learning names for objects versus names for places. In particular, children should attend to object shape for object count nouns but only sparse shape elements (or none at all) for place words. Empirical evidence is consistent with these predictions. Young children learning object names attend to shape rather than texture or size (see sect. 1 and Landau 1993). In contrast, there are no reports of children overrestricting figure or reference object shape when using prepositions in spontaneous speech or when learning novel prepositions (see Landau 1993 for review).

Note that our hypothesis provides a general framework for testing other acquisition patterns. Children should be predisposed to learn spatial terms representing quite sparse figure and reference object geometries and nonmetric regions – roughly the properties listed in Table 2 (subject to revision from cross-linguistic evidence). For example, they should resist learning a new preposition meaning "plin," "sprough," or "2 inches away from" the reference object. On the other hand, there should be a reasonable degree of flexibility in exactly how these object geometries and regions combine to yield spatial meanings in any particular language, thereby allowing an important role for learning. For example, children learning any language should be sensitive to the reference object's principal vertical axis. In addition, however, children learning English should be able to learn two separate terms for regions that (1) are or (2) are not in contact with the object (*on* and *above*, respectively), whereas children learning Warlpiri should be able to collapse this distinction into a single term (Bavin 1990).

A final set of questions arises from using our evidence on language to make predictions about neuropsychology. We are extremely tentative here, as we recognize the complexity and extent of the unsettled issues concerning the "what" and "where" systems. If we are right, however, then the "where" system should incorporate some shape information, specifically, that having to do with objects as volumes, their axes and orientations, and their distribution in space (sparse/dense/linear). In addition, it should be capable of encoding the kind of qualitative (nonmetric) spatial relations we have discussed, such as distance and (axis-based) direction. We would be delighted to find neurons capable of responding to such properties. At higher levels of analysis we would be interested to discover, for example, whether damage to the "where" system would have repercussions for the use and understanding of spatial prepositions but not, say, nonspatial prepositions or other prepositions used purely as grammatical markers.

## 4. Summary and conclusions

We began with the goal of using evidence from language to provide insight into the nature of spatial cognition, in particular to help understand how we represent objects and places. Our survey has shown that the domains of language describing objects and places draw on very different kinds of spatial representations. When objects are named as belonging to a category, their descriptions appear to draw on rather complex representations of shapes and surfaces. When the same objects play the role of figure or reference object in a locational expression, however, their descriptions appear to be highly schematized, preserving at most the axial structure of the object's principal volume. The spatial relationships that these objects engage are similarly sparse, including primarily qualitative distinctions of distance and direction. Thus, there are significant limitations on the kinds of descriptions represented by language, and the two domains of object and place exhibit quite different constraints.

Some of these limits appear to reflect partially a property of language design: filtering out of metric information. More important, however, these differences in how objects and places are represented may be correlated with a property of neurological design: a separation of spatial cognitive systems into "what" and "where." If our conjecture is correct, we have found a bifurcation in the expressive power of language that corresponds to a bifurcation in the functional and anatomical systems of the brain. To our knowledge, this is the first time a correlation has been made in cognitive science between a property of grammar – that is, the kinds of things that count nouns and prepositions standardly express – and descriptions of nonlinguistic systems represented in particular areas of the brain.

Without this correlation, language would still provide rich and systematic evidence on the character of spatial cognition, for as we argued in the beginning, anything we can talk about we must also be able to represent. What is exciting about finding this correlation is that whereas previous studies have documented the tasks performed by the two visual and neurological systems, linguistic evidence can now provide a new source of insight into the actual forms of information the systems encode and, more generally, into the nature of spatial representation.


## ACKNOWLEDGMENTS

A less detailed version of this paper appeared previously (Jackendoff & Landau 1991). We would like to dedicate this version to Lila Gleitman, who through many years of association has been a constant source of inspiration, wisdom, and good jokes.

This research was supported in part by NSF Grant IRI 88-08286 to Brandeis University, by Social and Behavioral Sciences Research Grant #12-214 from the March of Dimes Birth Defects Foundation to Barbara Landau, and by NIH Grant #1-RO1-HD28675-01 to Linda Smith, Susan Jones, and Barbara Landau. We are grateful to many colleagues for critical comments and discussion. These include Julian Hochberg, Michael Leyton, and Melissa Bowerman for detailed comments on previous drafts of this paper, and Don Hoffman, David LaBerge, David Murray, and Edgar Zurif for essential comments and references used in this study.


## NOTES

1. Clearly, the converse does not hold: We may be able to represent notions that are not so naturally encoded by language. In fact, the selectivity of language is the very wedge we use to gain insight into spatial representations. We take up the question of why language does not encode all aspects of spatial representation in section 3.

2. Our assumptions here may be oversimplified. For example, there is good reason to suspect that visual, haptic, and auditory information will connect with more than one motor system, for example, one that guides reaching and another that guides locomotion. In addition, there may be cases where visual information is directly translatable into motor commands, without being translated into amodal spatial terms. Nevertheless, translation into a common format is necessary to explain a large variety of tasks, including how language is readily learned and used by individuals whose modality of experience differs from the norm (e.g., see Landau & Gleitman 1985).

3. This is not to deny the complications involved in deciding what an object should be called, and how this relates to the object's category. We assume that what determines an object's name is its membership in some category and, to be sure, category membership (for adults, at least) is not always determined by object shape. Other considerations can be important, such as function, or, in the case of living things, descent (Carey 1985; Keil 1989; Murphy & Medin 1988). On the other hand, it is also clear that object shape is often critical to what an object is called. A toy bear may not really belong to the category *bear* (depending on one's theory of the nature of a category), but it does share its name – assigned by the head noun – with the animate versions in virtue of the similarity in appearance.

Whatever complex interactions are involved in assigning objects to categories, our interest for present purposes is in the mapping between spatial cognition and language. Limiting ourselves to these relationships, we note that shape-based representations are critical to object identification and that object identification is in turn critical to object naming. Hence our focus on object shape and its role in naming.

4. Recognition is apparently mediated by the lawful relationship between an object's shape and its silhouette: The sign of the curvature of points on the silhouette agrees with the sign of the Gaussian curvature of the corresponding points on the object. The visual system is apparently designed to register this correspondence, a fact that enters into our ability to parse objects (Hoffman & Richards 1984.) We thank Don Hoffman for pointing this out to us.

5. Whether geons are truly invariant over all viewpoint transformations is a question of current interest in theories of object recognition. This technical matter has little impact on our main point that componential theories of object recognition may play an important role in our understanding of how objects get named.

6. "Suitable" similarity metrics will no doubt be complex and the result will differ depending on the investigator's approach and goals. For example, one could imagine a bottom-up approach in which one could ask what kinds of discriminations can be made among different classes of shapes, or a top-down approach in which one could ask what role is played by world knowledge in shape classification (e.g., knowing that four-legged and two-legged animals fall into different classes). Different goals might also give rise to different similarity metrics. For example, an investigation of the morphological development of shape (Thompson 1961) and a study of what objects share the same "basic level" (Rosch et al. 1976) might give rise to different classes of shapes. Developing such similarity metrics is clearly vital to understanding the interaction of shape perception, classification, and naming.

7. Assignment of these terms can vary cross-linguistically, as in Hausa, where nonoriented objects are assumed to "face" in the same direction as the speaker, with the *front* assigned to the side farthest from the speaker. [See also Deregowski: "Real Space and Represented Space" *BBS* 12(1) 1989.] Note also that the criteria for all of these spatial part terms must be stated fairly carefully; they interact in curious ways depending on the shape and function of the object. For example, a house normally has a *front, back,* and *sides,* but no *ends,* because it is not long and narrow. A wide but shallow office building may be said to have a *front,* a *back,* and two *ends* (a left end and a right end); the axis defining front-back is orthogonal to that defining ends. There are therefore no axes left over to define sides. By contrast, a bus has a *front end,* a *back (rear) end,* and two *sides;* here the criteria for *front/back* and *end* project into the same horizontal axis and the orthogonal axis remains to define sides. Finally, a rectangular carton may be said to have two sides and two ends – but no front or back, because its axes are symmetric.

8. Some readers may protest that this theory of object shape appears too complicated and full of special-purpose devices – that the way humans understand objects cannot possibly require all this complexity. In reply, we can only challenge such critics to develop a simpler alternative that both answers the many concerns of object schematization and object constancy expressed by perceptual theorists and makes all the distinctions among objects that we have observed in language. We believe that any such theory will contain complexity comparable to ours.

9. Perceptual ground and reference object do not completely overlap in the linguistic representation of place. In some cases, they may be the same, for example, in the sentence "The cat is on the mat," the mat is the perceptual ground and the reference object. In other cases they may be different, as in "The cat is near the mat"; the mat is the reference object but not the perceptual ground. Also the encoded reference object is not always a landmark. Landmarks are typically reference objects, but the converse does not necessarily hold. (A piece of furniture might serve as a reference object but would not generally be considered a landmark.) The exact relationship among perceptual ground, reference object, and landmark is beyond the scope of this target article. We will focus on reference objects unless otherwise specified.

10. In Table 1, *compound* prepositions are combinations of words that function grammatically as a single preposition, more or less parallel to compound nouns, such as *garbage man* and *big top*. We do not consider compounds that incorporate open class materials (such as "on the top 3 inches of"). *Intransitive* prepositions, often classified as adverbs in traditional grammar, are locational and directional words that occur in all the usual grammatical positions for prepositional phrases but need not be followed by a noun phrase. Many of them, e.g., *outward* and *upstairs*, contain a preposition as a constituent. In addition, many standard prepositions such as *below, nearby,* and *through* can occur with or without an object. By analogy with the familiar distinction between transitive and intransitive verbs, then, these uses of prepositions with and without objects have been termed transitive and intransitive, respectively (see Jackendoff 1973 for more detail).

11. What makes a good reference object will vary enormously with context: In the majority of cases, the reference object must be "salient" or "distinctive." For example, in the context of a city, a good reference object will be large and stable and perhaps have significant cultural or emotional meaning (Lynch 1960). On the other hand, as Miller and Johnson-Laird (1976) point out, in a field of grey dots, something as simple as a red dot might make a good reference object just because it is perceptually salient. Extended to the domain of objects, certain features might function as landmarks if they are visible with peripheral vision and serve to distinguish the target from distractors (Hochberg & Gellman 1975).

12. These asymmetries are not confined to the spatial domain; rather, the existence of "cognitive reference points" (Rosch 1975) can be shown with color, number, and orientation, among others.

13. Going across a bridge is a special case, in that one goes from one *end* of the bridge to the other. Presumably, this is motivated by the fact that the bridge itself extends from one side to the other of something else, such as a road or a river.

14. There are a few exceptions to this overall generalization. Nautical terms like *port* and *starboard* require a boat as reference object. *Upstairs* and *downstairs* involve levels in a building (though not necessarily stairs, because one can go upstairs in a building using only elevators). And the compounds with *-ward* such as *homeward* and *shoreward* involve reference to the object named by the initial noun. The meaning of these terms may go outside the spatial relation system proper, involving an interaction with the object shape system.

15. The adjectival form *perpendicular to* places constraints similar to senses 1 and 2 of *across*. *Opposite* used as a preposition ("Bill is opposite Harry") means about the same as sense 3 of *across*, except that it leaves unexpressed the object that Bill and Harry are on opposite sides of.

16. In Avrutin's data, the prepositions that take the prepositional case include *na* ("on") and *v* ("in, inside"). Those that take the instrumental case include *za* ("behind"), *nad* ("above, over"), *pered* ("in front of"), *pod* ("under"), and *mezdu/sredi* ("between, among"). Those that take the genitive include *u* ("at, very close to") and *okolo/vozle* ("near, not far from, close to"). The one problematic example he has provided is *mezdu/sredi* ("between, among"), which we have analyzed as "interior to" and which therefore should take the prepositional case. It will take further research to decide whether this is truly exceptional or whether our feature analysis must be modified.

17. "Toward" and "away from" are more restricted than "to" and "from," in that the region they are constructed from is always "at X." There are no expressions **toward on X* or **away from under X*, parallel to *onto X* and *from under X*, for instance.

18. Pinxten et al. (1983) make an exhaustive exploration of Navajo spatial terms. Though they emphasize how different the Navajo spatial framework is from that of English, there are few surprises with respect to the parameters discussed here (insofar as we can follow their discussion without competence in Navajo). About the only case that involves a novel shape descriptor for an object, the issue with which we are most concerned, is *biniká*, a postposition meaning roughly "passing through a hole," as through the eye of a needle. It is interesting that the French verb *enfiler* implies a similar geometry of something slipping through a narrow opening (Jacques Mehler, personal communication). Thus, although we do not want to claim that English exhausts the spatial relations expressible in language, it does appear to provide a substantial and representative sampling.

19. Functional explanations of this filtering could take one of a number of forms. For example, one could propose that people can manage to communicate effectively while expressing only a small range of spatial relations, so language has evolved to have no more such expressions than necessary. This kind of explanation runs into problems in trying to explain why there is a proliferation of vocabulary in certain apparently "inessential" areas (e.g., color) but not in other apparently "essential" areas (e.g., spatial terms). What one needs is some reasonable design criteria for what *would* constitute a plausible function. Along these lines, Pinker and Bloom (1990) have suggested that by representing spatial relationships discretely, one can capture causal discontinuities in an efficient manner. For example, one is protected from the rain *under* a ledge regardless of how far under one is; and one is not so protected if one is not under the ledge, regardless of how far away from the ledge. Such qualitative causal facts could underlie the separation of continuous space into discrete regions such as are encoded in languages.

Our own thought about functional explanations is that they surely play some role in the evolution of vocabulary (as evidenced by the existence of technical or specialist vocabularies), but one would be hard-pressed to come up with a systematic account that explains vocabulary limitations across all domains. We are happy, however, to admire such attempts from afar.

20. Farah et al. (1988) call this a difference between "visual" and "spatial" capacities. One of us (R. J.) has a somewhat different interpretation: that we should think of both as subsystems of the multimodal spatial capacity. Because all the tasks tested by Farah et al. were exclusively visual – there were no haptic or motor tasks – the evidence so far does not distinguish the two possible interpretations.

21. For example, damage to the hippocampus results in impairment on tasks requiring memory for an object's location (Angeli et al. 1988), consistent with evidence reviewed by O'Keefe and Nadel (1978). Yet the hippocampus is one of the termini of the temporal pathway, which is the hypothesized locus of the object system.

22. Formal indexing does not answer the psychological or neurological question of how the link is effected. But this is altogether parallel to the familiar problem in language of how multiple representations are psychologically or neurologically connected – for example, what it means neurologically for the representation of the sound of a word to be linked to the encoding of its meaning.

23. An important methodological question is how one decides which elements in a language to consider. We have restricted our discussion to prepositions – closed-class elements – and have essentially left untouched spatial verbs, which are open class and seem to have much more latitude in what kinds of geometric and nongeometric elements they can represent. We agree with Talmy (1983) that cross-linguistic investigation should focus on closed-class elements (whether verb markers, prepositions, postpositions, etc.) that express spatial relationships.

# Open Peer Commentary

*Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.*

## There is more to location than prepositions

David C. Bennett
*Department of Phonetics and Linguistics, School of Oriental and African Studies, University of London, London WC1H 0XG, England*
**Electronic mail:** *d.ben@clus1.ulcc.ac.uk*

Landau & Jackendoff (L & J) present an illuminating account of the categorization of physical objects and of prepositional meanings. They are right, too, in pointing out that languages typically have tens of thousands of names for physical objects and less than a hundred spatial prepositions. They are wrong, however, when they imply that the rich system we have for describing things (the "what" system) contrasts strikingly with an impoverished system for describing locations (the "where" system). Descriptions of locations can be indefinitely detailed. A moderately complex example is given as the italicized phrase in:

(1) The jack must be positioned *beneath the jacking points at the outer ends of the cross-member.*

Admittedly, much of the information in this phrase is conveyed by noun phrases (*the jacking points, the outer ends, the cross-member*), but that does not affect the fact that the phrase as a whole answers a "where" question. Similarly, in the simpler sentence:

(2) There's a bike *in front of the house,*

the locational information is given by means of a preposition *and* a noun phrase. In general, locational expressions incorporate a "reference object." When L & J write (Introduction, para. 6) "place naming draws on quite sparse elements of object shape," they are referring only to the prepositional component of locational phrases. The noun phrase component can convey any amount of information about object shape. There is no great discrepancy therefore between the number of possible things and the number of possible places, or between the resources of our object naming and place naming systems. And if the assumed discrepancy does not exist, it is unnecessary to try to explain it by hypothesizing that "what" expressions and "where" expressions are processed by different parts of the brain.

I hasten to add that I do not disagree that the number of geometric features of reference objects, and of located objects, encoded by spatial prepositions is very limited. Indeed, I shall attempt to throw further light on this issue by mentioning one factor not noted by L & J. We approach the matter by comparing (2) with:

(3) There's a crack *at the thick end of the cue.*

It seems reasonable to suggest that the reference object in (3) is not just "the cue" but "the thick end of the cue" – that is where "a crack" is located (simple location being indicated by the preposition *at*). Assigning (2) a parallel semantic structure would entail treating "the front of the house" as the reference object: "a bike" is located "at-the-interior-of" (*in*) the space adjacent to "the front of the house." Such a semantic representation can certainly be defended. However, in addition to proposing a semantic representation for the whole of (2) and (3), it is relevant to indicate which bits of the semantic representations correspond to which bits of the syntactic representations. It is here that a difference emerges between *in front of* and *at the thick end of*. *In front of* is appropriately analyzed as a complex preposition: although originally syntactically complex, it is now well on the way to becoming "lexicalized," that is, to becoming a single lexical item. (In the case of *behind*, the lexicalization process has gone even further: unstressed *be-* is not generally equated with the preposition *by*, except by linguists; *hind* does not occur as a noun; and there is no following *of*.) By contrast, no one would suggest analyzing *at the thick end of* as a complex preposition. From a semantic point of view, lexicalization entails the combination of separate concepts into a single (albeit complex) concept. The location specified by (2) is best described therefore as involving the reference object "the house" combined with the relational concept "in front of." The location in (3), on the other hand, is indicated by the reference object "the thick end of the cue" combined with the relational concept "at." The question that now arises is why some strings of words become lexicalized whereas other, syntactically parallel, strings do not. The relevant factor would seem to be frequency of occurrence. It is presumably because objects are so frequently visualized as occupying the space adjacent to the front of some other object that the phrase "in front of" has acquired the status of a single concept. By contrast, it is only rarely that we find ourselves describing something as being located "at the thick end of" something else. In support of this line of reasoning, it may be noted that in an alternative world with very large numbers of cone-shaped and wedge-shaped objects, the situation would be different, because the meanings "at the thick end of" and "at the thin end of" would be far more readily applicable. In this case it might happen that the two strings of words in question would become lexicalized and perhaps give rise to the complex prepositions *thick-end* and *thin-end* (e.g., *A is thick-end B, C is thin-end D*).

This discussion of lexicalization indicates that frequency of occurrence is one factor influencing the formation of complex locational concepts and is therefore relevant to the question of the kinds of geometric information that can be incorporated into the meaning of prepositions. It is a mistake, however, to concentrate on prepositions to the exclusion of other parts of speech. Faced with verbs of a language such as Atsugewi, which place detailed geometric restrictions on the "figure object" (thing to be located), L & J write (sect. 3.2.3, para. 3): "Spatial verbs, which are open class and can clearly express more than just location, are different from English prepositions, which are closed class elements dedicated to expressing location only. . . . our hypothesis. . . is explicitly concerned with those closed class forms that are dedicated to expressing location." Yet, given that spatial information is expressed by prepositions (*in, inside*), adverbs (*nearby, inside*), verbs (*enter, contain*), adjectives (*long, thin*), and nouns (*top, inside*), and given the close relationship between such (nonsynonymous) expressions as *inside the jar* and *on the inside of the jar*, it seems unlikely that the semantic representations we derive, in the process of understanding utterances, are segregated into sections deriving from different parts of speech.

In view of the opinions expressed above, it is not surprising that I attach more importance to the problems for the Landau & Jackendoff hypothesis, which they acknowledge in the final paragraph of section 3.2.1, than to the claims they make in section 4, paragraph 2.

## Spatial and cognitive vision differentiate at low levels, but not in language

Bruce Bridgeman
*Program in Experimental Psychology, University of California, Santa Cruz, Santa Cruz, CA 95064*
**Electronic mail:** *bruceb@cats.ucsc.edu*

Vision seems to be a single, unified sense – we see things, describe them, and interact with them in a seamless continuity. On closer analysis, however, the unity of vision splinters into specialized subsystems. Characterizing these subsystems has become one of the central concerns of vision research; a central source of confusion has been that different methods yield different bifurcations of visual function. Landau & Jackendoff (L & J) contrast the rich and varied vocabulary of object names with the sparse geometric description of visual space, concluding that the difference in language may be related to physiological differences between spatial and object systems that have been identified in monkeys (Schneider 1969; Ungerleider & Mishkin 1982) and in lesioned humans (Farah et al. 1988). I suggest that the situation is a bit more complicated; L & J identify two modes of linguistic function in vision, whereas some other contrasts identify one system that is accessible to language and another that is not. The two linguistic modes may be part of a broader cognitive system.

One requirement for differentiating two modes of vision is the study of normal humans rather than lesioned humans or other primates. There is now extensive evidence that two visual modes are represented in the normal human brain and that they follow different rules (Bridgeman 1981; 1986; Paillard 1987). Following the nomenclature of Paillard (1987), these will be called cognitive and sensorimotor, respectively. The two systems have been applied in many contexts and given many different names. It is not yet clear whether all of the names refer to the same neurological machinery, for some of them address different aspects of behavior. All, however, share a common distinction between a uniquely spatial, generally unconscious, motor-oriented system, and a more symbolic system whose contents are at least partially conscious, forming the basis for perception.

In our first experiment on this problem (Bridgeman et al. 1979), two conflicting observations (saccadic suppression of perceived target position on one hand and accurate reaching to a target following a saccade on the other) were combined by asking subjects to point to the position of a target that had been displaced during a saccadic eye movement. Subjects were also asked whether the target had been displaced. Extinguishing the target and preventing the subjects from viewing their hands (open-loop pointing) guaranteed that only internally stored spatial information could be used for pointing. On some trials the displacement was detected, whereas on others it went undetected, but pointing was accurate whether or not the displacement was detected. Pelisson et al. (1986) found a similar result. This result implies that visuomotor localization is unaffected by the perceptual detectability of target position. A further test of this possibility was a two-alternative forced-choice measure of saccadic suppression of displacement. Even this criterion-free measure showed no information about displacement to be available to the cognitive system under conditions in which pointing was affected (Bridgeman & Stark 1979).

A more rigorous way to separate cognitive and motor systems is to stimulate only the motor system in one condition and only the cognitive system in another. This was done with induced motion, the apparent motion of a target when its background moves. We know that induced motion affects the cognitive system because we experience the effect and subjects can make verbal judgments of it. The above experiments implied, however, that information used for pointing might come from sources unavailable to perception. We inserted a signal selectivity into the cognitive system with stroboscopic induced motion (Bridgeman et al. 1981). A surrounding frame was displaced, creating the illusion that a target had jumped, although it remained fixed relative to the subject. After all stimuli were extinguished, the subject pointed open-loop to the last position of the target. Trials in which the target had seemed to be on the left were compared with trials in which it had seemed to be on the right. Pointing was not significantly different in the two kinds of trials, showing that induced motion did not affect pointing.

In a second condition of the same experiment, information was inserted selectively into the motor system. Each subject adjusted a real motion of the target, which was jumped in phase with the frame, until the target seemed stationary, so that the cognitive system specified a stable target. Subjects nevertheless pointed in significantly different directions when the target was extinguished in the left or the right positions, showing that the difference in real target positions was still available to the motor system. This is a double dissociation: In one condition the target displacement affected only the cognitive system and in the other it affected only motor behavior.

Dissociation of cognitive and motor function has also been demonstrated by giving the cognitive and motor systems opposite signals at the same time. Again, the experiment involved stroboscopic induced motion; a target jumped in the same direction as a frame but not far enough to cancel the induced motion. The target appeared to jump in the direction opposite the frame when it actually jumped in the same direction. Saccadic eye movements followed the veridical direction even though subjects perceived stroboscopic motion in the opposite direction (Wong & Mack 1981). If a delay in responding was required, however, eye movements followed the perceptual illusion, implying that the motor system has no memory and must import information from the cognitive system after a delay.

All of these experiments involve motion or displacement, leaving open the possibility that the dissociations are related to a confounding of motion and position rather than to a representation of visual space per se. A new method tests dissociation of cognitive and motor function without motion of the eye or the stimuli at any time. The dissociation is based on the Roelofs effect, a tendency to misperceive the position of a target presented against an off-center background. Subjects were always biased by a Roelofs effect in judging which of five possible positions of a target was presented; an off-center surrounding frame caused judgments to deviate to the opposite side. For half of the subjects, however, the frame had no effect on pointing (a measure of the sensorimotor system). Again, if the sensorimotor system has no memory, some subjects may have switched to the cognitive system to point to the target position after it was extinguished; indeed, with a long enough delay we could force all subjects to show a Roelofs effect even in pointing (Bridgeman 1991).

Taken together, these psychophysical experiments show that cognitive and spatial visual systems can be distinguished on a lower level than that of Landau & Jackendoff, a level that differentiates linguistic from nonlinguistic coding. Their fascinating contrasts between object and spatial language may simply reflect the physical fact that there are a lot of objects, but only a few Euclidean geometric relations.

## The role of cerebral lateralization in expression of spatial cognition

Halle D. Brown
*Department of Psychology, Harvard University, Cambridge, MA 02138*
**Electronic mail:** *hb@huelings.harvard.edu*

Although the analysis of linguistic descriptions given by Landau & Jackendoff (L & J) is an innovative approach to the study of spatial cognition, I would suggest that only half of the issue has been considered. In particular, their analysis has informed us about the geometric properties of objects and spatial relations that we know to be the specialty of the left cerebral hemisphere. Recent studies, however, suggest that some aspects of spatial cognition are performed by the right hemisphere. These are as important as those performed by the left hemisphere but are not readily evident from an analysis of linguistic expressions. The contributions of both hemispheres to location and object representation will be considered in this commentary.

We can certainly express spatial representations verbally, but we also express them in nonverbal behavior. For example, to reach for an object, we need to have computed its location in space. Both behavioral studies with humans (Kosslyn et al. 1989) and computer simulation models (Kosslyn et al. 1992) have provided evidence for two distinct types of spatial-relation representations. Categorical spatial-relation representations, used to judge whether a dot is above or below a line, can be distinguished from coordinate spatial-relation representations, needed when judging whether a dot is within 3 millimeters of a line. Kosslyn and his colleagues have found that categorical spatial relations are computed more efficiently in the left hemisphere, whereas coordinate spatial relations are computed more efficiently in the right hemisphere. Only the categorical type of spatial relation underlies the prepositional descriptions listed in L & J's Table 2.

Coordinate spatial relations are equally important for the study of spatial cognition, yet they are not expressed in as straightforward a manner with prepositions as categorical spatial relations are. Kosslyn and his colleagues have argued that coordinate spatial-relation representations are most useful to guide action (see Kosslyn & Koenig 1992). We need to know precisely where the cup is located to guide the trajectory of our hand toward the handle; a verbal description of the movement would be superfluous. The locations of objects in space and of parts relative to each other may be encoded procedurally rather than descriptively, as a set of coordinates to guide motor programming.

In their analysis of geometric properties conveyed by object nouns, L & J have again given us only half the story. Object representations that have associated count noun names focus mainly on invariant features of objects. These geometric properties allow us to access the same object name even when the object is presented in varying forms (e.g., words written in different type fonts are read the same way). Yet we also have the ability to distinguish particular exemplars of a type of object. Recent studies of cerebral lateralization have indicated that two fundamentally different object representations can be computed, at least in the visual system. An "abstract-visual-form" representation is activated by different instances of the same type of object; as long as the inputs have the same invariant relations among their parts, the same abstract object output will be computed. Marsolek (1992) has found that abstract-visual-form representations are processed more efficiently in the left hemisphere. On the other hand, form-specific representations are activated by different instances of the same abstract type of object. Marsolek et al. (1992) found that the right cerebral hemisphere distinguishes more effectively between different form-specific representations of the same abstract type of object than the left hemisphere.

Why is there no distinction in language for types versus tokens that maps onto this hemispheric difference in the brain, as there is with object nouns ("what") versus prepositions ("where"), which map onto the difference between ventral and dorsal system processing in the brain? Although object types can be distinguished by different spatial relations within and between parts, object tokens can be distinguished more precisely. For example, in the study by Marsolek et al. (1992), implicit memory for words, measured as stem-completion priming, was reduced when words were presented in different cases at study and test, but only in the right hemisphere. That is, although "w-o-r-d" is read as the same object type whether it is printed in upper or lower case letters, the right hemisphere apparently distinguishes tokens (e.g., "WORD" vs. "word") in memory. The physical features or visual details of objects are needed to distinguish tokens. Why do we not have nouns to express the geometric properties of objects that distinguish tokens? Perhaps because this analysis of featural variations is carried out more effectively in the right hemisphere, which does not have language output capabilities. The output from the token system must cross the corpus callosum into the left hemisphere, where a discrimination between tokens and types would no longer be distinguished in the system generating output for "what" descriptions.

The alternative hypotheses suggested by L & J regarding differential linguistic expression of object properties and spatial locations can now be reevaluated in light of this consideration of cerebral lateralization. Their "Design of Language Hypothesis" may be correct precisely because language-processing systems in the left hemisphere have access only to generalized object and location representations. Note, however, that language does not "filter" otherwise precise spatial relations; rather, spatial relations computed more efficiently in the left hemisphere serve a different purpose from the precise, coordinate relations computed more efficiently in the right hemisphere. Their "Design of Spatial Representation Hypothesis" can be better characterized in terms of processing in the "what" and "where" pathways in the left hemisphere, with computations of object "types" underlying count nouns and computations of categorical spatial relations underlying prepositional descriptions.

The study of spatial cognition must clearly include an analysis of cerebral lateralization of function and should avoid research tools biased to examine the specializations of only one hemisphere. Rather, our tools should be general enough to examine and ultimately characterize the specializations of both hemispheres. Landau & Jackendoff have expanded our understanding of representations of spatial relations in two different systems ("what" vs. "where"), but the implications of their analysis are limited to one domain (the left hemisphere). Converging evidence from other methodologies is needed to complete our understanding of the range of spatial information humans can express.

ACKNOWLEDGMENTS
I would like to thank Stephen M. Kosslyn, Chad J. Marsolek, and Kevin N. Ochsner for their helpful comments on this manuscript.

## Frames of reference in the spatial representation system

David J. Bryant
*Department of Psychology, Northeastern University, Boston, MA 02115*
**Electronic mail:** *bryant@northeastern.edu*

Central to Landau & Jackendoff's (L & J's) argument is the assumption that a single spatial representation system underlies perception and language. I agree with L & J that human cognition uses separate "what" and "where" systems, my purpose is not to criticize their analysis. Instead, I will briefly discuss additional evidence for a common perceptual/linguistic spatial representation system, examine some properties of the spatial system, and suggest one way in which the analysis of spatial language might be extended.

***Mental models of space.*** A large body of research has determined that people represent texts in situational mental models rather than as a propositional record. It is relevant to the present discussion that mental models preserve explicit and inferred geometric relations between objects in described situations. In particular, research has demonstrated that mental models preserve spatial features such as distance (e.g., Glenberg et al. 1987; Morrow et al. 1987) and relative direction (Bryant et al. 1992; Franklin & Tversky 1990) Moreover, these features guide the retrieval of information, determining the accessibility of objects in mental models. When analyzing a verbal description, readers normally extract information about objects and their locations and represent it in a form that is spatial and perceptual (Johnson-Laird 1983, pp. 156–62).

The close correspondence between mental spatial models and perception can be seen in a class of mental models called spatial frameworks (Franklin & Tversky 1990). A spatial framework organizes objects and their locations within the frame of reference created by an individual's three body axes (head/feet, front/back, left/right), in the same way that body axes are used in perception to locate objects (Gallistel 1990, pp. 106–18; Hintzman et al. 1981). Spatial frameworks also render certain spatial locations more accessible at retrieval, depending on the perceptual and physical asymmetries of body axes and the perspective of the observer (Bryant et al. 1992). Recent studies have demonstrated that people use the same kinds of spatial frameworks during visual perception of scenes (Bryant & Tversky 1991; Logan 1991), indicating that representations of perceived and described environments have equivalent structure.

As further evidence of a common perceptual/linguistic spatial system, people's spatial representations of descriptions interact with perceptual representations. Easton and Bentzen (1987) found that performance on a spatial finger-maze task was impaired when subjects simultaneously verified verbal spatial statements but not when they verified nonspatial statements. Similarly, performing a visuospatial tracking task interferes with a person's ability to form a coherent spatial model from a verbal description (Oakhill & Johnson-Laird 1984). In both cases, interpreting verbal spatial directions and performing a spatial task seem to compete for the resources of the same spatial system.

***The spatial representation system.*** All this reinforces L & J's assumption that a common spatial system underlies spatial cognition in perception and language. To explore further the geometric properties of spatial language and cognition, however, it is worth wondering what this spatial representation system is like. The purpose of the spatial system is to represent the layout of objects in the environment. To do this, it must use some *frame of reference* that establishes three spatial axes. The spatial system can then determine each object's position along each dimension in the resulting coordinate space. Humans commonly use the egocentric frame of reference, defined by the three body axes (head/feet, front/back, left/right), and the allocentric frame of reference, defined by orthogonal axes set outside the observer and anchored on prominent landmarks or aligned with environmental features. People may also use object-centered reference frames (see Shepard & Hurwitz 1984).

The allocentric and egocentric reference frames are crucial to spatial perception. We need allocentric cognitive maps to navigate in the world and the egocentric frame of reference to guide our actions in immediate space. Also, because we can perceive the world only from our own position, we can create allocentric representations only through transformations of egocentric representations (Gallistel 1990, pp. 106–9). People constantly update allocentric maps from egocentric perception and likewise direct egocentric interaction with the environment on the basis of allocentric maps. It is not surprising, then, that both frames of reference are available for the representation of linguistically described space. Readers create allocentric and egocentric spatial models to represent described environments and update both types of representation regardless of the perspective of the description (Taylor & Tversky 1992).

***Frames of reference and spatial language.*** Although frame of reference is clearly necessary to represent space, spatial prepositions generally do not specify a frame of reference (Retz-Schmidt 1988). At most, prepositions are constrained by the alignment of a major axis of the figure or reference object (L & J). Thus, I can say "the computer is behind the filing cabinet," and mean that the computer occupies a position to the filing cabinet's intrinsic back. On the other hand, I could say "the computer is behind the filing cabinet," and mean that both objects are to *my* front, but the computer is farther away. The word *behind* does not specify whether I am using an egocentric or an allocentric frame of reference. Likewise, most spatial prepositions can be used in both of these frames of reference (exceptions seem to be words like "through" and "inside," which require an object-centered interpretation).

Although spatial prepositions do not specify a frame of reference, language at the discourse level does. Levelt (1984) among others has distinguished between the deictic system of spatial reference, in which spatial prepositions are interpreted relative to one's own egocentric origin and body axes, and the intrinsic system, in which spatial terms are interpreted with respect to external axes of a referent object or the environment itself. These two systems map onto egocentric and allocentric coordinate frames respectively, but they coexist at the level of spatial prepositions. To say that prepositions are *interpreted* with respect to egocentric or allocentric referents is crucial because these spatial terms have meaning in both systems. Thus, other cues are needed to specify a frame of reference in language, cues that can only be included at a level beyond the single word (see Retz-Schmidt 1988). The deictic system, for example, must refer to the speaker's body sides and establish this context in discourse for the interpretation of spatial terms. Likewise, the intrinsic system must establish a set of environmental axes to distinguish between the possible meanings of spatial prepositions. These discourse-level linguistic systems, like prepositions, are influenced by the way the spatial system represents the perceptual world. For example, Levelt (1984) has noted that perceptual and spatial features such as the orientation of objects, the world's gravitational axis, and the posture of the observer limit what spatial representation can be derived from a verbal description.

***Conclusion.*** The purpose of my commentary has been to flesh out the idea of a common spatial representation system as it has guided the evolution of spatial language. In addition to L & J's argument for separate "what" and "where" systems, I would suggest that other vital spatial concepts are embodied in the spatial system and constrain language at some level. In particular, to understand how we represent space, we need to consider how frames of reference are used by the spatial system in perception and language.

## Generative versus nongenerative thought

Michael C. Corballis
*Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand*
**Electronic mail:** *mikec@ccu1.aukuni.ac.nz*

I should like to argue that many of the distinctive characteristics of spatial language and spatial cognition described by Landau & Jackendoff (L & J) are uniquely human and primarily a function of the left cerebral hemisphere.

Like language, human manufacture is characterized by an open-endedness that seems to be unique to our species; there seems no limit to the number and variety of different objects humans can create. Animals do use and make tools (Beck 1980), but in a "one-off" rather than a generative fashion. Moreover, the procedures by which humans make objects are similar to those by which they generate propositional language, and involve some of the same principles (Corballis 1991; Greenfield 1991). Just as phonemes are combined to form words, words to form phrases, phrases to form sentences, and so on, so standard parts are combined in hierarchical and recursive fashion to create complex objects. Most of those parts have names – wheels, axles, handles, knobs, shafts, blades, nuts, bolts, screws, nails, bulbs, sockets, sprockets, switches, transistors, and so on (and on).

This combinatorial, recursive mode may characterize not only the language and objects we produce but also many of our cognitive representations. Biederman (1987) drew attention to the parallel between his recognition-by-components theory of object representation and the combinatorial nature of language. Although his "geons" are not directly or obviously related to manufacture, and indeed are readily applied to nonmanufactured objects (such as a penguin, to take one of his own examples), I think it reasonable to suppose that both may be linked to the same underlying, distinctively human cognitive system. Elsewhere, I have referred to this system as a generative assembling device or GAD (Corballis 1991).

GAD may be distinguished from an (evolutionarily) much older system that represents things in a nonsymbolic way. Because this older system is a direct product of gradual selective adaptation, it is much more finely tuned to the subtle nuances of the natural environment. I do not think that a geon-based theory could ever account for our ability to recognize the faces of the individuals we know; geons or generalized cones might capture the prototypical shape of the human body (Marr 1982), but could not easily represent the idiosyncrasies of shape or gait that single us out as individuals. Similarly, as L & J point out, language permits only crude statements about the locations of things in space, yet our behavior in reaching or navigation displays a fine spatial tuning that belies our clumsy words. As a rough analogy, I like to think of the GADly system as comparable to a Lego construction set, whereas the unGADly one might be compared to Play-Doh (Corballis 1992), or for those who like to play with adult toys, to a digital computer and a connectionist network, respectively.

I have speculated elsewhere on the possible time course of the evolution of GAD (Corballis 1991; 1992). Here, however, it is pertinent to ask why it was adaptive to superimpose a new representational system on the old. Despite the open-endedness afforded by GAD, I do not think the answer hinges on memory capacity – we can, after all, recognize hundreds if not thousands of faces, and many other natural objects that do not readily lend themselves to geonetric (or geonological?) principles. A network with as many connections as a human brain would have enormous storage capacity without having to resort to symbols or recursive principles.

The advantage of GAD may lie partly in communication. A channel transmitting categorical distinctions is intrinsically less noisy than one transmitting metric distinctions. L & J draw attention to the fact that even in sign language, where the potential for analog mapping is high, a continuous dimension like speed is reduced to two or three categories – a reflection no doubt of the limit on our ability to make absolute judgments (Miller 1956). A more important consideration, though, may have been the need for improvisation. The unGADly system is the product of slow evolutionary adaptation, and as such was poorly equipped to deal with the rapid changes in environment that we hominids inflicted on ourselves, partly through persistent emigration to different terrains, but more conspicuously through the manufacture of objects, such as dwellings, tools, clothes, and so forth. Arguably, only a recursive geon-like system could keep up with cultural and manufacturing change, much of which occurs within an individual's life span.

It may please the more theologically inclined to know that GAD seems to reside primarily in the left cerebral hemisphere. We all know this to be true of language, but Kosslyn et al. (1989) have also shown that in making judgments about the relative locations of objects, the left hemisphere seems to rely on a categorical code, the right on a metric one. There is also evidence that the representation of partwise or manufactured objects may be primarily left-hemispheric, whereas the representation of more naturalistic ones is more bilateral or right-hemispheric (Farah 1991; Warrington & McCarthy 1987; Warrington & Shallice 1984; but see also Farah & McClelland 1991). However, I do not think that the distinction between the GADly and the unGADly coincides precisely with that between the manufactured and the naturalistic. GADly principles have no doubt invaded our representations of animals or plants, culminating in the sciences of zoology and botany, and there are few more GADly sights than the models of molecules that lie around chemistry laboratories. Conversely, the craft of the sculptor, potter, or painter may be based on holistic rather than combinatorial principles.

I suggest that the contrast between the language of objects and the language of places is neither exclusively a property of language nor of spatial cognition, but was bestowed by GAD.

## Are spatial representations flattish?

J. B. Deregowski
*Department of Psychology, University of Aberdeen, King's College, Old Aberdeen AB9 2UB, Scotland*
**Electronic mail:** *psy022@aberdeen.ac.uk*

Landau and Jackendoff (L & J) find a great disparity between the way the English language describes objects and the way it describes spatial relationships amongst objects. The latter descriptions, they observe, are markedly more rudimentary.

Although naming an object acknowledges its spatial presence, it does not necessarily imply that the object's three-dimensionality is a dominant characteristic of its mental representation. It may be that this characteristic, although important in the context of *immediate* perceptual activity, is less important for encoding and retention. Perception obviously has to take into account the spatial attributes of objects and their mutual spatial relationships. This is necessary for survival; but it can be questioned, perhaps, whether the encoding of objects in memory by reference to their three-dimensional characteristics is of equal importance in this context. Recognition of objects may not rely primarily on their three-dimensionality, which may constitute a secondary factor, present but not salient, akin to, say, surface colour; and, like colour, it may in certain circumstances gain greater importance than it normally has; apart from such circumstances, however, its importance is small. The hypothesis that three-dimensionality of objects is not of prime importance for their encoding will be examined using a source of evidence largely ignored by L & J: pictures, that is, images on flattish surfaces.

The axes that pass through Marr's (1982) cylinders, threading them together to form a figure (Figure 2A in L & J's target article), themselves constitute an easily recognisable pin-figure. This mannequin and its perceptual kith and kin have been much used by artists for millenia. A band of pin-man hunters is shown chasing deer in a Mesolithic picture in Cueva de los Caballos and the chase continues in numerous Bushman shelters of South Africa (see, e.g., Pager 1972). This type of portrayal was readily accepted by illiterate Mekan (Me'en) living in a pictureless culture (Deregowski et al. 1972) and is universally popular with children. It is unlikely, therefore, that the use of pin-figures calls for great pictorial sophistication. This is confirmed by data obtained by Fortes (1940; 1981; Deregowski 1978), which show that men who have never drawn before use pin-figures quite spontaneously when requested to draw. [See also Deregowski: "Real Space and Represented Space: Cross-Cultural Perspectives" *BBS* 12(1) 1989.]

The notable feature of pin-figures is that they readily convey their meaning without any attempt to show directly that the depicted objects are three-dimensional. Silhouettes and outline figures (which can be thought of as unfilled silhouettes) also represent objects without any direct hint of depth. Unlike pin-figures, these figures are constructed by portraying selected contours of a solid's surfaces. These "typical contours" have been found to be of great importance in children's drawings as well as in drawings of certain artistic schools (Deregowski 1990a; Dziurawiec & Deregowski 1992). The use of silhouettes in road signs is likewise evidence of the efficacy of this type of pictorial communication (Deregowski 1990c).

Pictures that have no illusory element of depth, in which the depicted objects are seen as flat (although the portrayals are readily recognisable as showing three-dimensional objects) have been termed epitomic (Conley 1985; Deregowski 1980; 1990b; Parker & Deregowski 1990). Pin-figures, silhouettes, and outline drawings clearly belong to this category. This has implications for L & J's analysis because it shows that the recognition of depicted objects proceeds quite well in the absence of perceptual cues as to their three-dimensionality; it also raises a question: If solids can be adequately depicted by treating indications of their three-dimensionality as distinctly secondary, is it necessary to postulate that mental representations must treat such indications otherwise?

Chwistek (1960; see Parker & Deregowski 1990) observed that there is a type of figure that represents objects as they are and as the artist knows them to be. A uniformly red vase, say, is painted in such a picture as uniformly red without any reference to shadows or reflection caused by the ambience in which it is placed and the direction of falling light, or to the artist's experience of the object, or indeed to the artist's mental condition. Such a picture shows the vase so that its typical contour appears in the picture's plane.

If the line of argument followed above is valid, then spatial qualities are not central to the encoding of objects. Many objects can be readily encoded, given names, or otherwise classified purely on the basis of their encoding attributes: axial arrays or typical contours. The encoding attributes, however, are not equally readily derivable from all objects in the environment although they are easily derivable from a large majority of objects present when languages were formed. Most animals, for example, have readily detectable, typical, idiosyncratic contours that generally run along the animals' spine so that the side view of an animal is the most typical one (but see Dziurawiec & Deregowski 1992). In contrast, man-made objects often lack such distinctiveness of shape (there is little difference between the essential shape of a television set and a packet of cigarettes), although for reasons of commerce attempts are often made to introduce shape distinctions (e.g., in the shape of containers for liquid, where none is needed functionally).

The argument is therefore that spatial qualities are not as important in the encoding of objects, as would appear to be the case *prima facie*. The disparity between descriptions of objects and the description of spatial relationships is therefore not as great as L & J suggest, simply because the naming of objects does not imply that their spatial qualities are of prime importance for mental representation.

It may also be worth observing that many terms used to describe spatial qualities of objects are those that are *sensu stricto* applicable to flat objects. Boxes, for example, are said to be square and no distinction is made between the roundness of an orange and the roundness of a coin. This usage is not confined to the hoi polloi but is found in psychological writings as well. Greenfield et al. (1967) are agreed, for example, that an orange and an alarm clock are round. This observation, by suggesting that the third dimension is not of prime importance in linguistic descriptions, strengthens the notion of the relative unimportance of the third dimension in the mental representation of objects.

## Causal models of spatial categories

Jacob Feldman

*Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08903*

**Electronic mail:** *jacob@ruccs.rutgers.edu*

Landau & Jackendoff's (L & J's) discussion of the verbal correlates of spatial categories is a welcome addition to research on categorization phenomena, filling out the sometimes vague territory between perceptual models and verbal conceptual categories. It dovetails nicely, for example, with my own attempt to classify spatial categories on purely perception-theoretic grounds, with the added benefit of cross-fertilization from linguistic (or at least lexical) concerns. L & J's enumeration of the possible meanings of prepositions, in particular, has an impressive aura of completeness. As they argue, these simple spatial categories are apparently just those that are so perceptually salient and have such general utility in our descriptions of the relative positions of things that we are actually willing to assign them each fixed words – closed-class words, no less.

At an underlying formal level, though, categories of structure and categories of position are more intimately related than L & J are willing to let on, notwithstanding the differences they discover in the way they are eventually expressed – in fact making these differences all the more interesting. My own account of these types of categories, though differing in motivation, methodology, and scope, results in a list of "natural" categories that L & J would find quite familiar. Hence, I think my account tends to reinforce their position, laying a somewhat firmer mathematical foundation under the sometimes vague conceptual arguments from which they get their enumerations. In particular, my account suggests a justification, on formal epistemological grounds, of this amazingly constant list of ubiquitous spatial categories. I will get back to that account below.

***"What" is more complex than "where"?*** L & J express surprise that encoding *shape* turns out to be so much more complex than encoding (relative) *position*. Their surprise seems puzzling, however, when one considers that structure is in fact *intrinsically* more complex than location. The position something occupies, after all, is at most a three-dimensional entity (because space is 3-D); what complexity there is comes from the complicated ways in which coordinate systems for encoding a 3-D position can be affixed to something, namely, to a gravity or viewpoint vector, or to vectors defined by aspects of some object's shape (above a surface, near a vertex, along an axis, and so forth).

The *shape* of an object, on the other hand, is a very high-dimensional construct; just *how* high-dimensional depends on one's model of shape. (This is in fact why Rueckl et al. [1988] feel certain that the discrepancy they find between the computational difficulty of "what" and "where" would only increase with more realistic object models. More realistic locational models never get more complex than three dimensions – only object models do.) Intrinsically, that is, with a maximally dumb model, structural models could be infinite-dimensional, if every nuance of shape were taken to be a plausible category distinction. What is impressive about, say, Marr's (1982) generalized cone theory is that shape could be meaningfully captured with so few dimensions, that shape could be reduced to such a simple description without losing much descriptive power vis-à-vis the categories of things actually extant in the world. The power of a shape theory is in constraining the useful dimensionality of shape without giving up too many desirable distinctions among classes of things. But you can never get it down to as low as three dimensions, like position – even as simple a class of geometric objects as 4-geons has four intrinsic shape dimensions. (See Feldman [1991] or Kendall [1989] for an indication of how complex this problem can get when the purely mathematical aspects are investigated more deeply.) Hence, we would always expect the linguistic system for encoding shape to be larger and richer than that encoding location; again, the surprise is that a finite but useful representation of shape is possible at all.

***Good categories as cues to causal structure.*** The more difficult question, though, is: What elevates the particular set of relative position categories that people evidently find natural (and that they express in prepositions) above the host of unnatural but mathematically similar alternative categories? L & J repeatedly suggest that relative-location categories are simply *coarser* than pure metric information – that the "bins" of relative position picked out by prepositions are distinguished simply by being big, and borders between them, by being blurry. It is easy to see that it is not quite so simple. For example, the regions "above" and "below' the ground are separated by a surface that is very fine compared to the size of the two regions. The region "near" or "at" a point is infinitely smaller than the complementary region (the rest of the universe). Perhaps most interesting (for reasons elaborated below), the region "in line with" or "along" an axis or an edge picks out a subspace of lower dimension than the 3-D space in which it is embedded. Conversely, a coarse metric grid does not automatically produce natural spatial concepts: "Less than 100 meters away" and "less than 200 meters away" are both lexicalized equally often, namely, never. Rather, the distinction between good and bad spatial concepts has a subtle structure. The outlines of this structure are probably best looked for, as Marr (1982) famously prescribed, in a consideration of the essential constraints governing a useful representation of (in this case) the locations of things.

In particular, consider the following general constraint on perceptually natural concepts about relative location. Given the regularities extant in the particular world being described, each such concept should contain a set of relative positions, all of which are qualitatively interchangeable with respect to the causal relationship between object and referent. For example, the objects "in" the refrigerator tend to be perishable food items that were all intentionally placed there for the same reason; they differ systematically from objects "outside of" the refrigerator. Similarly, to take a more violent example, the objects "along" the line of a rifle sight have in common with each other a particular potential causal relationship with the rifle. That is, the prepositions that capture these spatial relationships are doing something more than simply encoding them, in the sense of reducing image data to a more compact representation; rather, they are particularly tuned to encapsulating plausible hypotheses about causal stories.

It turns out that the above aim can be satisfied, more or less, in a formal scheme such as the one I described briefly in Feldman (1992). Leaving the details aside, we can treat a spatial position as the result of a sum of translation operations in some coordinate frame, in much the same way that Leyton (1992) treats objects as the result of some sequence of group-theoretic operations, and for much the same reason: The position of an object, with respect to the origin of the coordinate frame, is then seen as having a causal history behind it. (That is the sense in which place is a special case of structure: Structure can be modeled as the result of a sequence of arbitrary generative operations, but with position it is always translation.) The coordinate frame is in turn defined in terms of critical values (right angles, equal lengths, etc.) of commonplace structural parameters (orientation, length, etc.) that are liable to have causal significance because of the regularities of the physical, mechanical, and biological laws governing our environment. These critical values tend, over and over again, to mark critical distinctions between regions of configuration spaces of objects in the environment – that is, to mark boundaries between causally interchangeable objects.

The critical formal trick is to represent location as a sum of transverse translations in some coordinate frame; then discrete subcategories correspond to discrete, algebraically distinct subsets of the full inventory of translations. For example, one of the models would be the set of positions that can be expressed mathematically as just a translation along an edge (say), with translations away from the edge measuring zero and hence algebraically dropping out. Enumerating all such models we get a lattice (a kind of hierarchy; see Jepson & Richards 1992) of subcategories of location, each of which putatively has a privileged status as a cue to a distinct causal model. Because each causal model ends up corresponding to a distinct formal object, this scheme attaches a very literal definition to the idea of qualitative equivalence with respect to causality. "Meaning" thus accrues to the relative location of object and referent, in a way that makes the lexicalization Landau & Jackendoff have discovered seem particularly intriguing.

## On places, prepositions and other relations

Angela D. Friederici
*Cognitive Science Lab, Berlin Institute for Psychology, Free University Berlin, Berlin, Germany*
**Electronic mail:** *friederici@zedat.fu-berlin.dbp.de*

Landau & Jackendoff (L & J) propose a theory in which the linguistic and conceptual systems are viewed as isomorphic and sharing a common neural basis – at least for spatial aspects. In the following, I will focus on the isomorphism assumption and leave the neuropsychological claim undiscussed.

Concerning the isomorphism assumption, L & J claim in particular that the linguistic and the conceptual representations of objects and places are identical with respect to the (geometrical) aspects they encode. Representations of objects linguistically encoded in count nouns are rich in detail concerning the object's shape. Places are linguistically encoded in prepositions, and these represent information about the object only sparsely. The evidence L & J provide in support of this claim is sparse itself, however.

Linguistically seen, prepositions encode spatial relations (a definition used in sect. 2.1) between objects, between objects and places/regions (e.g., the tall building at Washington Square), as well as between places/regions (e.g., Washington Square is close to Kennedy Square) but they do not encode places/regions as such. Because there is only a small number of possible spatial relations, the number of linguistic elements that encode these relations is limited. L & J use the fact that the number of nouns is much larger than the number of prepositions as one of the arguments that objects and places are not only represented differently linguistically, but that this difference reflects a contrast between objects and places at a more general cognitive level. The fact that a language contains more nouns than prepositions is not taken as a reflection of the few existing spatial relations that must be encoded, because the number of prepositions could be larger in principle, they claim, if prepositions were to encode properties of the objects whose relation they describe. But why should a linguistic system do this? If prepositions are considered to encode relations similar to formal logical relations (e.g., $<$, $>$, $=$), the linguistic sign to encode such a relation can be used most efficiently if it abstracts from the parts whose relation it encodes. A very general principle for the structure of relational expressions could be: Specify the parts in any necessary detail, but do not encode these details in the element expressing the relation itself.

L & J use a lot of space in their target article to list linguistic observations (and experimental evidence) in support of their claim. Although their final statement is that a linguistic analysis would be sufficient to model the nature of spatial representations, they do not use this approach themselves. Rather than staying within the linguistic domain, they switch from the linguistic level to the cognitive-psychological level and back, just as they need it. The argumentation, therefore, becomes circular. Given the isomorphism assumption, this might be tempting, but given that they try to provide evidence for the isomorphism claim in the first instance, this is not a valid procedure.

The particular experimental data they cite in support of their claim are not convincing. Take, for example, the study by Landau and Stecker (1990) in which children and adults had to learn new words, in this case nouns and prepositions. The very elegantly designed experiments show that children at the age of 3 years are able to use syntactic information (prepositional phrase vs. noun phrase) to interpret a given phonological shape as encoding either an object or a location. This is an exciting result. Children, like adults, are able to generalize over different object shapes when the new word appears in a prepositional phrase, but they are sensitive to shape when they have to learn a new noun. L & J, however, take this finding as evidence that children and adults do not represent details about object shape when they learn prepositions. The fact that children and adults are able to generalize over different shapes when it comes to learning the use of a preposition does not mean that they represent the object in a sparse way during these learning instances. A necessary test for this claim would be to compare recognition memory for object shape under both the noun learning and the location learning situation. It may well be that even in preposition learning details of the object's shape are represented but these are just not relevant to the use of a preposition.

It seems that there is no solid empirical evidence to support L & J's otherwise interesting claim that objects are sparsely represented when a spatial relation is encoded. What is shown is that object shape information is more relevant in noun use than in preposition use but not that the underlying representation is of equal sparseness or richness, respectively. It may well be that there is a rich underlying representation for given spatial scenes and that only certain features are extracted when a spatial relation is linguistically encoded.

It appears that the advantage and the drawback of language in general is that a given linguistic form generalizes over different instances, objects, and even actions (people have different habits of drinking), but that at the same time this means a linguistic form is underspecified with respect to a particular instance, object, or action.

ACKNOWLEDGMENT
This work was supported by the Alfried Krupp von Bohlen und Halbach Stiftung.

## Is spatial information imprecise or just coarsely coded?

P. Bryan Heidorn & Stephen C. Hirtle
*Department of Information Science, University of Pittsburgh, Pittsburgh, PA 15260*
**Electronic mail:** *pbh@lis.pitt.edu; sch@lis.pitt.edu*

Landau & Jackendoff's (L & J's) target article makes a strong case for the assertion that there are different cognitive systems for representing "what" and "where" in spatial language and spatial cognition. We generally agree with this assertion and the arguments made for the usefulness of comparing language and cognition data. We also concur with the premise that any aspect of space that can be expressed in language must also be present in nonlinguistic spatial representations. Our main difference is with the conclusion that the representation of geometry of objects is "rich" relative to the representation of place, as argued from the language data, and, as a result, the representation of place is imprecise.

To explain the differences in spatial properties of count nouns and spatial prepositions we argue for a "functional design of language" hypothesis, consistent with the conversational maxims of Grice (1975), which extend the hypotheses of "design of language" and "design of spatial representation" presented in the target article. By this hypothesis, we stress the two separate communicative goals served by count nouns and spatial prepositions. Count nouns communicate object identity, providing access to the "what" system, whereas spatial prepositions communicate location information, providing access to the "where" system. Furthermore, in normal conversation, spatial communication does not exist outside of either a defining context or the perceptual-cognitive loop, be it explicit or implicit.

Why are there so many count nouns and so few spatial prepositions? Count nouns provide a token to refer to representations in a very powerful pattern recognition apparatus. As they are normally used, however, count nouns are not intended to tell the listener anything new about the spatial properties of the objects. For example, consider the phrase "gorilla on a mountain." Our spatial system has stored the complex geometry of a gorilla. We can use this knowledge to recognize a gorilla or to imagine one when the word "gorilla" is used. However, there is no new information being communicated about the shape of the gorilla in the linguistic act. The rich spatial knowledge of the pattern recognition system is isolated from the language system. The preposition "on" is the only word in this example that provides new spatial information. If we want to communicate new spatial information about the geometry of a gorilla, we need to use spatial prepositions (or verbs) to convey the new information.

Our language apparatus does not give us a very powerful mechanism to talk about shape. When we say "gorilla," there is no implicit shape built into the syntax of the word. The word provides access to a powerful pattern recognition system. The system does not seem to lend itself to the general build-up of a linguistic description of gorillas from its parts. There are many count nouns because we cannot rely on any system other than rote matching of complex shape to "word." Each count noun is a token for some fuzzy set of shapes around some prototype.

Spatial prepositions do something very different. They tell you where things are and they do so in a very structured manner. When I say one object is "in front" of another, I am telling you something new. Furthermore, the context of the situation and interaction with the perceptual apparatus often resolve any ambiguity. If I say, "The pencil is near the telephone," it is most likely that the pencil is within a relatively small area, say between 2 cm and 20 cm away from the phone. If I say that the "mailbox is near the restaurant" or the "airport is near the city," the figure is certainly more than 20 cm away from the reference object. In addition, the spatial information conveyed by the term "near" in the example of the pencil is enough to locate the object. In contrast, the purpose of the linguistic act in the airport example is not to specify an exact location but convey class information about the metropolis (e.g., "We could hold the conference there, as there is an airport . . ."). Thus, spatial prepositions transmit enough information to locate an object, which is the purpose of the linguistic act.

The interesting thing here is the different underlying system that is being used: It is not the pattern recognition system, but the perceptual system. If one draws the listener's attention to the correct region, the listener's perceptual system can resolve any minor ambiguity. Likewise, the geometry of the subject and object of a spatial preposition is unimportant for the communicative task of putting things in "close enough so you can find them" locations in space. The "where" system is powerful enough to resolve the ambiguity.

Furthermore, if we consider the type of coarse coding often used in neural networks (sect. 3.2.1; Hinton et al. 1986), we find that sufficient, and often quite accurate, spatial information can be encoded using relatively few "where" units. It appears that L & J have equated the coarseness of the representational code with the coarseness of the spatial information conveyed by the code.

The argument that the different word types use separate underlying cognitive apparatus and serve different functions also supports the idea that there are separate and different "what" and "where" systems. According to this "functional design of language hypothesis" language develops to support a set of specific functions. The mechanisms that develop in the language are matched to the system being used. There are few spatial prepositions because that is all that is needed to support the spatial-perceptual system. There are many count nouns because we need to communicate the many object distinctions (but not shape distinctions) that need to be communicated. One can no more say that the "shape" system is rich and the "spatial" system is weak than that human pattern recognition is rich and human perception is weak.

## No perception without representation

Donald D. Hoffman

*Department of Cognitive Science, University of California, Irvine, CA 92717*
**Electronic mail:** *dhoffman@orion.oac.uci.edu*

Whatever we can talk about we can also represent. This is a key principle behind the arguments of Landau & Jackendoff (L & J), and, stated at this level of generality, it is certainly hard to take exception. No claim is made about the form of the representation – it could be abstract symbols, single neuron activity, or population firing patterns. The claim is simply that language is not magic. Each linguistic ability requires some underlying representational ability.

What is true here for language is, most likely, true more generally. There is no perception without representation. Indeed, there is no perceptual or motor ability without representation. Denoting such an ability $A$, and a representation $R$, we can write this dictum as $A \rightarrow R$.

L & J actually use the converse of this principle: $\neg A \rightarrow \neg R$. They show, with very thorough and illuminating examples, that our linguistic abilities for describing "where" are very limited. Thus, $\neg A$. From this they conclude $\neg R$, namely, our "where" representations are very limited.

Airtight? Only if L & J have really established $\neg A$, and they have not. To establish $\neg A$ it is not enough to show a limited ability with "where" in our language system; one must show this limitation in *all* of our perceptual and motor systems. If there is even one perceptual or motor system in which we have a richer ability with "where," then, by $A \rightarrow R$, we must posit the corresponding richer representation.

I think there are such systems. Consider Michael Jordan, for example. During the NBA playoffs he sank six 3-point shots and amassed 35 points in a single half – from every position on the court and every possible orientation (or so it seemed) of his body in space. Put Jordan somewhere on the court and ask him to describe where the basket is. Ten to one, you could not make even a layup based on his description. Now hand him a basketball and ask him where the basket is. Ten to one, you will soon believe he knows, and without a word spoken. After asking him repeatedly from different positions and orientations, you will soon believe he has a very rich ability with "where" – and thus, by $A \rightarrow R$, a very rich representation of "where."

Objection: This is a highly trained ability in a talented man. True. But most of us can at least hit the rim most of the time, and swishes are not needed to make this point.

Objection: Baskets are simple objects, hardly a case in which complex "where" interrelations are needed. True. But remember that those 35 points came with 9 other men on the court (5 desperately trying to stop him by any means the refs could not see), and with thousands of fans not exactly in quiet meditation. "Where" relations were changing rapidly even after he began a shot. (Most of us, once again, will not score in such conditions, but we will at least hit the rim.) And in the NBA it is often the spatial relationships of objects around and above the rim that are changing all the time.

Objection: Though other things may change position, the basket does not, so this is still a very special case of "where" ability, and therefore does not imply a more generally rich "where" representational system. Perhaps. But Jordan also made a number of assists, and the "where" targets for his passes, as well as those trying to defend against them, were, well, all over the place and not standing still. It is hard to imagine a much richer environment in which to exercise and display your "wheres." To get to Pippin you must go through Drexler and around so and so and over so and so and . . . well, no wonder his tongue hangs out. Most spatial relationships for which there are prepositions and, I claim, many more for which there are none, are being exercised.

Perhaps there are objections that are fatal to the Jordan example. It seems certain, though, that many other plausible examples can replace it – examples from other sports, everyday physical activities, experiments in stereo probe placements or structure-from-motion probe placements (Braunstein et al. 1992), or other psychophysical experiments. And that is the problem with the task that Landau & Jackendoff have set themselves. They are trying to set *upper* bounds on the complexity of our "where" representations, but the data they collect can really only set *lower* bounds. And in this regard they have done a great service. I come away from their target article with a new respect for and understanding of the representational capacity of our "where" system. It is almost surely no less sophisticated than they have described. And many NBA fans count on it being much more.

## Evolution and physiology of "what" versus "where"

David Ingle

*Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139*

The target article by Landau & Jackendoff (L & J) documents and explores important differences between spatial and object-based representations as expressed in language. Because their article necessarily skimps on the evolutionary background and physiological underpinnings of these two realms of perception, my comments will highlight those issues.

It might be noted that the first explicit discussion of "what-versus-where" systems was a four-man symposium published in 1967 in *Psychologische Forschung*. Schneider (1967) dealt mainly with an attempt to distinguish tectal and cortical visual functions as orienting versus discriminating, and one could argue that his stripe-discriminations do not constitute "object vision." The other three participants (Trevarthen [1968], Held [1968], and Ingle [1967]) actually focused more on perceptual distinctions between spatial and object vision, using examples from humans, monkeys, and fish.

It would be of great evolutionary interest to devise tests comparing the number of spatial distinctions versus the number of object discriminations animals make. Most animal psychologists would probably support my prediction that there are smaller differences between rodents and primates (including man) in spatial route-finding or object-retrieval than in object discrimination skills. Thus, an evolutionary explosion of object-classification abilities in the higher mammals probably precedes language capacities, and may be a precondition of language.

L & J seem to fall into an error common among physiologists in discussing the relationship of low-level cortical processing (color, orientation, motion coding) to the shape/space dichotomy. It is quite clear that lines and edges can define the three-dimensional layout of surfaces as well as the shapes of objects. More dramatically, it is clear from the "biological motion" demonstrations of Johansson (1973) that rather complex arrays of moving spots (with no contours visible) can yield highly specific identification (e.g., a man hammering, a couple dancing) as well as yielding percepts of three-dimensionality via parallex on expansion effects. There are many examples to show that motion and contour processes can feed into either spatial or object shape recognition systems. Milner and Goodale (1992) have recently made a similar argument in discussing parietal (spatial) mechanisms involved in grasping oriented objects.

L & J do recognize that spatial vision involves unconscious computations that do not require language, but the extent of these hidden computations exceeds the examples they have listed. In my own recent work, it is clear that humans (even young children and mental retardates) accurately localize remembered targets on the floor of a large room after walking and turning along a disjointed route without vision. What is consciously perceived is the target's stable location in space, not the distances walked or angles turned; those computations remain hidden.

Other examples include the prediction of trajectories of missiles (baseballs in our era, spears in Paleolithic times) that must be caught or avoided. No outfielder can describe to a trainee how he judges whether a fly ball will carry beyond him. Somewhat more complex patterns (not yet analyzed in the laboratory) include estimating the timing and constraints of a human arm swinging a weapon so as to efficiently duck and counterattack. In general, the fighting skills of carnivores, birds, and even lizards (where jaw fencing is a skill comparable to that of the Three Musketeers) require some impressive computations at an unconscious level. As we begin to analyze spatial skills involving complex motion of both subject and object, we will perhaps realize that a large part of the mammalian brain is devoted to largely unconscious (at least routinized) computations for acting within a spatial framework.

Object and spatial vision should be compared not in their implicit complexity, but in terms of which representations normally become part of conscious deliberation. This is a dichotomy *within* the recognition mode as well (e.g., recognizing up to 10,000 different faces is accomplished without verbal analysis, whereas identifying the kind of emotion or social implication in a generic facial expression does map onto subtle verbal labeling).

Finally, I suggest that Landau & Jackendoff's ideas be further validated by counting the number of object categories that can be remembered by very young children before and after language becomes differentiated. Is the richness of object memory more related to *chronological* age or to *verbal* performance in cases where the two can be dissociated?

## Distinguishing the linguistic from the sublinguistic and the objective from the configurational

Scott D. Mainwaring
*Psychology Department, Stanford University, Stanford, CA 94305-2130*
**Electronic mail:** *sdm@psych.stanford.edu*

Landau & Jackendoff's (L & J's) analysis centers on two key distinctions, *what* versus *where* and (what I will label) *linguistic* versus *sublinguistic*. Conjoining these, a four-way classification of geometric representational systems underlying spatial thought and language results. Not only are there separate *what* and *where* systems, but also distinguishable *linguistic* and *sublinguistic* partitions within each. This represents a significant elaboration of the "3-D model structure" in Jackendoff's (1987a) theory of the interface between the spatial and linguistic faculties. Is this multiplication of theoretical entities justified? I will consider each distinction separately.

***The linguistic/sublinguistic distinction.*** In partially accepting the Design of Language Hypothesis, L & J explain that the language faculty has access only to part of the general spatial representational system. This part is a "filtered" version of the richer metric descriptions at the sublinguistic levels accessible to perceptual and motor systems. Filtering, they say, is likely to be "a design feature required by any system that must collapse numerous complex distinctions onto a finite set of elements." These elements are still geometric categories (such as geons, directed axes, and angles), but are more qualitative and schematic than the corresponding sublinguistic encodings.

This filtered/unfiltered distinction brings to mind such venerable dichotomies as conceptual/perceptual, discrete/analog, and descriptive/depictive. Indeed, the filtered representations are more abstract, less visualizable. However, *both* the linguistic (filtered) and sublinguistic (unfiltered) systems seem to be conceptual, discrete, and descriptive – modifications of, not major departures from, data structures such as Marr's (1982) 3-D model. Both are based on a "finite set of elements" and make "numerous complex distinctions." The filtered and unfiltered systems, it appears (though L & J do not go into detail), differ only in the amount of detail they can encode, not in the basic mechanism of representation (as, for example, the discrete/analog distinction standardly implies).

It is therefore unclear why language is restricted to using only the set of "filtered" representations. For example, what barrier prevents a lexical entry for a count noun from referring directly to a geometric model constructed in the fine-grained, sublinguistic medium, given that it can refer to one in the coarse-grained medium? Such a barrier is necessary to make plausible the major claim of L & J's target article – that the linguistic *where* system is a much coarser medium than the linguistic *what* system – because the sublinguistic *what* and *where* systems are both admittedly quite fine-grained and capable of capturing metric details of objects and configurations. But independent justification for, and explanation of, this linguistic/sublinguistic barrier is needed.

Such analytic support could come out of a recent movement for theoretical restructuring within cognitive science: "situativity theory" (Greeno 1992; Greeno & Moore, in press). As a way of reconciling traditional information-processing psychology with ecological psychology, these theorists argue that linguistic and other symbolic knowledge should be analyzed in a fundamentally different way from sublinguistic knowledge. For example, much basic knowledge, such as a mammal's understanding of Euclidean space, is better characterized as based on "attunements" (Greeno et al., in press) or "internalized constraints" (Shepard 1984) rather than "representations." Instead of being used ubiquitously as in traditional cognitive science, "representation" is reserved for a particular kind of attunement, mediated by symbol systems explicitly used and assigned meaning by the agent. Verbally expressible knowledge (such as that measured via think-aloud protocols) falls into this category. That this distinction between attunement-knowledge and representation-knowledge could usefully elaborate a theory such as L & J's is an exciting possibility.

***The what/where distinction.*** Whereas the sublinguistic/linguistic distinction categorizes geometric knowledge based on something like degree of abstraction, the what/where distinction divides it according to subject matter. According to the Design of Spatial Representation Hypothesis, the *what* system underlies knowledge of objects, including complexes composed of multiple parts; the *where* system underlies knowledge of the spatial relationships between objects comprising a configuration. The principal claim is that, at the linguistic level, the *what* system encodes part shapes and interpart connections in detail, whereas the *where* system encodes object shapes and interobject relationships more schematically. (How the *what* and *where* systems differ at the metric, sublinguistic level is less clear.) Loosely, the object system is to the configuration system as a set of aerial photographs of cities is to a low-resolution road map: One directly encodes detailed shape information of delimited areas, the other compresses and schematizes shape and location information to convey regional spatial relationships succinctly. By linking the two via "indexing," integrated environmental knowledge emerges (see sect. 3.2.2).

As L & J point out, *what* and *where* knowledge is complexly interdigitated. Both object and configuration knowledge consist of elements associated together spatially. The heart of the proposal, as I read it, is that the way this association is accomplished in the object/*what* system differs from that in the configuration/*where* system. This suggests that the same physical complex could be encoded as an object or as a configuration, resulting in two very different ways of understanding it. It would be interesting to work out an analysis of environmental learning, typically glossed as progressing from landmark to route to survey knowledge (e.g., Moore 1976), as involving initial development of direct but loosely bound *configuration* knowledge followed by reconceptualization of the environment as a single complex *object* that, like a map, can compactly though indirectly orient one within the actual environment.

Recent work by Neisser (1989; 1992) presents an additional perspective on the kind of information used by the *what* versus *where* systems. He argues that the *what* system operates by recognition, comparing gradually accumulating evidence with stored representations (criteria for belief). The *where* system operates by "direct perception," picking up invariants from the rich flux of sensory information (Gibson 1966). Considering such epistemological issues in object recognition versus configuration perception may be a fruitful direction for further elaboration of Landau & Jackendoff's thought-provoking model.

## Spatial development

David R. Olson
*Ontario Institute for Studies in Education, 252 Bloor St. W., Toronto, Ontario, Canada M5S 1V6*
**Electronic mail:** *internetd_olson@oise1.oise.on.ca*

Landau & Jackendoff's (L & J's) impressive synthesis leaves unresolved two of the issues that so puzzled us (Olson & Bialystok 1983) in our theory of children's spatial development.

First, what is the relation between representing an object such as a face in terms of spatial properties including roundness, and representing a face as round. L & J use the terms "description" and "representation" interchangeably, as does much of the cognitive science literature, to mean both a property that figures in some way in the representation of an object (sense 1), and a concept, category, or word that represents that property as a member of a class (sense 2). Preverbal children recognize balls on the basis of their roundness, but they represent them as round, and thereby include them in the class of round things, only when they are around 4 years of age. We argued that spatial cognitive development consisted largely of translating the spatial features implicit in the representation (sense 1) of objects and events into conceptual spatial representations (sense 2). To fail to mark this distinction is to lose, I suggest, the possibility of explaining cognitive development.

Second, L & J's willingness to adopt the "what" versus "where" distinction from neuropsychology leads them to overlook an important aspect of the relation between objects and locations. Objects are merely collections of parts with fixed locations relative to each other. Conversely, locations specify variable relations between parts, often of macro "objects." Eyes above nose is what makes a drawing of a face a face; but mole above the eyes is irrelevant to its being a face, so the location of the mole relative to fixed features has to be worked out anew for each face.

Again, a face and a farmyard are both "objects" differing only in scale: The nose at the front/the house at the front; the hair at the back/the barn at the back; the nose in the middle (of the front/the well in the middle; the mole at the side of the nose/the garage at the side of the barn; and so on. What makes the location of an object relative to some reference frame complex, we suggested, was that it is often variable, whereas the location of spatial features of objects relative to each other is fixed. Variable locations may be difficult, not because they involve new spatial predicates, but because they have to be constructed anew.

These two issues may also be related. Perhaps what makes the representation of variable location possible is the possession of a set of concepts, concepts that would permit an object or feature of an object to be represented (sense 2) as being round or being at X. Indeed, it is reasonable to expect that representing (sense 2) the location of a mole as being to the left of the nose is, in principle, no different from representing the eyes as being above the nose. Neither is required for recognizing a face or recognizing Aunt Maggie.

## Is spatial language a special case?

Dan I. Slobin
*Department of Psychology, University of California at Berkeley, Berkeley, CA 94720*
**Electronic mail:** *slobin@cogsci.berkeley.edu*

Landau & Jackendoff (L & J) describe two different kinds of phenomena and try to put them together into a single story: Information about objects and locations is separately represented in the mind/brain, and information about objects and locations is encoded by different means in language. But the connective tissue is missing from their argument. Why should objects be encoded by nouns and spatial relations (in some languages) by prepositions? And why should nouns point to richly detailed objects, whereas prepositions point to schematic relations? I submit that the linguistic division of labor reflects the way in which the cognitive distinctions are mobilized online for communicative purposes.

L & J pose the question in the following terms: Why is it that object names represent the geometric richness of shape whereas spatial prepositions represent a limited and coarse encoding of "place"? They say that the naming of objects requires "detailed descriptions of shape;" however, object names do not *describe* the shapes of objects; rather, the name *references* the object. When I say "frog," I do not describe the shape of a frog, but evoke whatever representation of frog you have in your mind. But when I saw "the frog is in the jar" I invite you to represent a spatial relationship between a frog and a jar. The question is not why some object categories are definable in terms of shape, but rather why there are so few grammaticized terms for spatial relationships, and why these terms tend to ignore details of the figure and of the reference object.

First consider the meanings conveyed by spatial terms. Note that (in languages like English) spatial prepositions do not occur alone but are accompanied by nouns that reference figure and ground. The words "frog" and "jar" already tell you a good deal about the possible relations between these two objects. This is why many languages do without the precision of English spatial prepositions when describing canonical or highly predictable relations between two objects. Spanish, for example, makes do with a single preposition, *en*, to describe both frog-*in*-jar and frog-*on*-table. Turkish does not make use of available locative postpositions at all to describe these scenes; an all-purpose locative inflection is sufficient: "frog jar-LOC" and "frog table-LOC." Both language types have recourse to more specific relational expressions for noncanonical locations – the equivalents of "the frog is on top of the jar, under the table," and so on. These facts suggest that a good deal of *inferencing* is carried out, making use of representations of the objects associated with locative terms to more fully interpret the grammaticized spatial representation.

Other languages, as noted by L & J in their discussions of Atsugewi and Tzeltal, provide verb-marked information about characteristics of figure and/or ground. It is interesting that such languages make sparse use of nouns in their discourse. In Atsugewi (Talmy 1975), for example, a verb meaning "don't-cause-dirtlike-substance-to-move-downwards-into-liquid" is sufficient, without the use of any nouns at all, to command you not to flick your cigarette ashes into my coffee, or not to dump the coffee grounds into the river, because the relevant object representations can be determined from visual evidence alone (i.e., contextually), given the schematic object representations encoded on the verb. Tzeltal (Levinson 1991) gets along with only one spatial preposition, an all-purpose term simply meaning "located with regard to." Thus, "sits-bowl-like LOCATIVE-PREPOSITION table gourd" conveys the same information as the English *the gourd is on the table*. Furthermore, given the information in the verb, one or both nouns may be omitted.

Across all languages it seems that information about spatial relations cannot be calculated without knowing what kinds of objects/substances are being related. Why, however, is the list of object and spatial-relational characteristics so limited? L & J suggest that this is due to a separation between the "what" and the "where" systems of the mind/brain. But this division does not dictate that the "where" system be so limited and schematized. One reason the "where" system can be linguistically limited is because of what can be inferred from the accompanying nouns (or the physical presence of their referents). This suggestion introduces a factor not considered by L & J: economy of information processing

In making such a proposal, it is appropriate to include spatial verbs and verb particles as well as adpositions (prepositions and postpositions). L & J exclude spatial verbs on the grounds that they are "open class elements" (i.e., elements of a large, unrestricted set that can be added to by speakers). Although it is true that verbs, in general, constitute one of the major open classes (along with nouns, and, in some languages, adjectives and adverbs), languages also have relatively small and relatively closed subclasses of verbs that function to schematize a semantic domain in the same way as prepositions (or personal pronouns, modal auxiliaries, or tense/aspect markers, etc.). For example, one such closed subclass of verbs deals with object destruction. In English, which has a fairly elaborate collection, we make distinctions of the nature of the object to be destroyed (e.g., *break, tear, smash*), force dynamics (e.g., *tear* vs. *rip*), the degree of destruction (e.g., *cut* vs. *shred*), the texture or constituency of the object (e.g., *crumple, crumble, shatter*), and so forth. The list of such verbs, however, is fairly short: I estimate that there are no more than 50 such verbs in English – fewer than the prepositions enumerated by L & J. There are many such "closed subclasses" of verbs, for example: verbs of manner of movement (*walk, run, jump, swim, fly*), manner of talking (*shout, scream, whisper, mumble, mutter*), posture (*sit, stand, lie, crouch*), and so forth. The Tzeltal spatial verbs seem no different: Though there may be about 300 of them (Levinson 1991), still they are a closed set, characterizable by a small collection of systematic distinctions. In each instance, we might ask, following the model of L & J: Why are there only 50 verbs of object destruction, or 25 verbs of manner of movement, or 6 verbs of posture? And, continuing their example, one might propose that this is precisely because the human mind/brain is constructed to conceive of object destruction or manner of movement or whatever just in terms of those features that can be used to describe these closed class lexical sets. Convenient, and maybe true, but, if so, not due to the difference between the "what" and the "where." It is, of course, a truism that whatever is expressed in language must also be present in nonlinguistic representations. This initial "premise" of L & J couldn't be otherwise. But why are there relatively closed and relatively open sets of lexical items?

L & J propose that one open class – "count nouns that label different kinds of objects" – is based on our capacity to represent shapes and they explain that the visual system is constructed to represent many different kinds of shapes. But the capacity to represent shape hardly accounts for the range of object names in language. L & J state that "descriptions must be potentially fine-grained enough that one can decide which objects are to be named *horse* and which *donkey*, or which *dog* and which *wolf*." Note, however, that these "descriptions" are not part of language; they are descriptions of how the visual system and brain might recognize objects. The reason we have words for horses and donkeys is that we can distinguish them – by whatever means (just as we have words such as *mistress, lover,* and *wife,* which cannot be distinguished by shape at all). L & J confuse the fact that we can recognize shapes and the fact that we can name objects, some of which can be characterized by shape. Languages have a large number of nouns because people categorize the world in a myriad of ways. Shape recognition is a fascinating scientific topic but it in no way explains why people need so many different ways to refer to objects of experience.

Why do we use nouns instead of descriptions? There are at least two answers: (1) Fully specified descriptions are impossible (e.g., there is no way to substitute the noun *dog* with a series of words that fully, or even adequately, describes a dog), and (2) partial descriptions are cumbersome. Our minds are so constructed as to quickly and easily access a particular mental representation in association with a particular word – a sound pattern that is totally arbitrary with regard to the associated concept. When necessary, we use word combinations to name objects – if they are new or complex or unfamiliar – but we reduce them to shorter expressions and ultimately to words when they become frequently used (*intercontinental ballistic missile* to *ICBM*, *analysis of variance* to *ANOVA*, etc.). One could imagine a language with a small number of words in which every referring expression was phrase-like or longer. Such a language would be far more transparent than our languages, but it would require a different kind of processor. The human processor has to make fast decisions: We can integrate verbal information only over short stretches of time. We economize by packing a great deal of information into brief bursts of sound, such as nouns.

It is not surprising, then, that we also economize in our relational expressions. Such economy is evidenced not only in the spatial prepositions on which L & J focus; it is characteristic of *all* grammatical choices. We might just as well ask why Russian has only six grammatical cases, or why Bantu languages have a dozen or so noun classes, or why the English aspect system has only perfect and progressive. Every utterance is only a schematic sketch of the content to be conveyed. The "open class" content words – nouns and/or verbs – point to the entities and events evoked by the utterance and the "closed class" grammatical elements suggest a framework of temporal, causal, spatial, and social relations in which those entities can be situated. It is characteristic of languages to have a collection of obligatory grammatical elements that must be used repeatedly, generally in every sentence. A grammatical/conceptual distinction that must be accessed so frequently operates under several constraints: (1) It must be maximally general, so as to apply to most instances. Thus, languages do not inflect nouns for the color of the referent object, perhaps because most objects do not have distinctive colors. (2) It must correspond to a way in which humans recognize and store events. Thus, for example, no known language marks its verbs to distinguish whether an event took place on a rainy or a dry day, perhaps because this distinction is not relevant to the ways in which we represent most events. (3) It must contrast maximally within a small class of distinctions. Here there seem to be two types of online constraints: (a) If a grammatical distinction is *obligatory*, that is, if it has to be expressed in every applicable sentence, the choice tends to be binary (e.g., perfective/imperfective, singular/plural) or ternary (e.g., singular/dual/plural, here/there/yonder). This is probably because online decisions cannot entertain many alternatives. Such categories must be general, salient, and minimally partitioned so as to allow for constant, rapid allocation of every noun or verb to the appropriate subcategory. Such obligatory markers lay out the basic framework of propositional content and relation: tense/aspect/modality, transitivity/causativity, person/number, speech act, and so on. (b) Spatial prepositions represent a different kind of online problem, in that every sentence does not require specification of spatial relations. If a sentence does have a prepositional phrase, it requires a particular preposition. (Similarly, if it does have a modal phrase, it requires the choice of a particular modal, and so forth.) Here, it seems, languages provide relatively small sets of choices: a handful of modals, several handfuls of prepositions, a small basket of spatial verbs.

The reason that sets of grammatical options are small is not because human beings are incapable of conceiving of more detail. We express more detail whenever we need to, with the extended possibilities of building phrases and clauses. The "where" system can be linguistically expressed in exhaustive detail, but not in the grammatical component. I propose that the constraint on what is grammaticized in human languages is not due to the structure of cognition alone, but is also due to what is useful, quick, and easy to access online: to enable us to program utterances while speaking and to make use of received utterances while listening in order to build up a mental representation of the communicator's intent. A processor of the human sort must make rapid decisions with regard to *all* grammaticized notions – not only notions of spatial relationship. Such decisions cannot require attention to a great deal of specific detail. To be sure, the schematization reflects the most natural predispositions of the human mind, but these are cognitive systems with a dynamic component. I suggest that the interface between grammar and cognition is influenced by our ability to schematize experience *for communicative purposes* just as much as it is influenced by our biologically determined predispositions to schematize experience as we do.

## From perception to cognition

**Michael J. Tarr**
*Department of Psychology, Yale University, New Haven, CT 06520-7447*
**Electronic mail:** *tarr@cs.yale.edu*

I am sympathetic to Landau & Jackendoff's (L & J's) proposal that linguistic structure may serve as a window on the nature of human spatial cognition, reflecting the "deeper" structural constraints of spatial representations. Indeed, mechanisms of human visual cognition may have evolved concurrently with those used for language (Tarr & Black 1991), and, in all likelihood, as the result of similar adaptive pressures (White 1989). Despite my generally favorable reaction to this approach, however, L & J's treatment of theories of spatial representation[1] raises concerns that may detract from their effort to draw plausible connections to linguistic structure. First, in adopting an approach to object representation of somewhat limited explanatory power – that of volumetric primitives – they are put in the position of having to amend the theory in an *ad hoc* manner. Second, because they confuse the different roles spatial relations play in visual recognition and representation, there is no clear mapping between spatial prepositions and spatial relations.

The volumetric primitive approach (Biederman 1987; Marr & Nishihara 1978) to object representation is attractive for many reasons, not the least of which is its computational elegance. Symptomatic of the problematic nature of this approach, however, is L & J's need to "enrich" it by incorporating constraints gleaned from linguistic evidence. It is not that three-dimensional part-based descriptions will not have some role in visual cognition but that these descriptions are fundamentally inadequate in and of themselves for explaining a great deal of human visual behavior. This is clearly evidenced by numerous experimental results on human object recognition that are not accounted for by volumetric models (Bulthoff & Edelman 1992; Jolicoeur 1985; Tarr 1989; Tarr & Pinker 1989). Indeed, proponents of the volumetric approach (Biederman 1987; Biederman & Cooper 1991) have been careful to state that such representations are restricted to *basic-level* access and that other forms of object representations and recognition mechanisms subserve "complex" recognition. The inherent need for these representations is indicated by L & J's introduction of surface representations to account for objects such as paper, phonograph records (compact discs?), and lakes. Unfortunately, as incorporated by L & J, this extension is *ad hoc* in that it is conceived solely to account for extant linguistic data. This same theoretical weakness occurs in a variety of guises: orienting axes, directed or symmetric axes, negative object parts, containment, and so on. Essentially, whenever a spatial property is linguistically represented, a corresponding property is posited in the spatial representation.

L & J attempt to sidestep the "special-purpose" characteristics of their theory by challenging "critics to develop a simpler alternative" (Note 8). Yet even if a better alternative could not be offered, it is crucial that elements of any theory of spatial representation play a *functional* role in spatial reasoning, recognition, and navigation. As formulated by L & J, the proposed additions serve a purely *explanatory* role. Moreover, there are plausible alternative theories of object representation that implicitly include many of the spatial properties that are to be found in linguistic structure.

Recently, both computer scientists (Koenderink 1987; Seibert & Waxman 1992; Ullman & Basri 1991) and psychologists (Bulthoff & Edelman 1992; Tarr 1989; Tarr & Pinker 1989) have proposed *multiple-views* theories of object representation. One important element of such theories is that shape representations are egocentric and thus encode the relative position of the object to the observer. That is, each *view* of an object depicts the appearance of the visible surfaces of that object from a small range of orientations. Multiple-views models include many of the properties that L & J have argued are necessary in theories of spatial representation. First, because properties such as left and right are inherently egocentric, they are implicitly encoded in each orientation-specific representation (Tarr & Pinker 1989). Second, qualitative spatial relationships between views (e.g., front/back and top/bottom, as well as explicit encoding of left/right) provide sufficient information to infer adjacency between views, whereas qualitative spatial relationships between object parts provide category information (the role Biederman [1987] has suggested for part-based models). Third, surfaces and containment are easily represented without modifications or additions to the theory. Both are simply encoded in the number and variety of views necessary to completely represent an object at a given scale (Koenderink 1987; Kriegman & Ponce 1990). In contrast to L & J's proposal, therefore, multiple-view theories of spatial representation have principled (e.g., functional) reasons for incorporating may of the properties identified as elements of spatial language.

My second concern is that in their mapping from spatial prepositions to spatial relations, L & J have equated the so-called "what/where" distinction with the recognition of objects and the recognition of places. In fact, the role of spatial relations in visual cognition is far more complex. There exist at least three levels at which spatial relations may be used to encode language: (1) between-part relationships within an object (i.e., "her head was *on top* of her body"); (2) between-object relationships within a scene (i.e., "the calculator was *on top* of the desk"); and, (3) between-place relationships within the environment (i.e., "the apartment *on top* of the Chinese restaurant"). Object location (e.g., "where" in sense 2) is computationally distinct from place location (e.g., "where" in sense 3). For the most part, the "where" system as discussed by L & J refers to object location – a problem equivalent to the problem of segmentation – for example, determining the position of an object independently of its identity. Both senses of "where," along with within-object relations, are used in spatial language, hence the parallel drawn between the bifurcation of language (count nouns and prepositions) and spatial cognition (what and where) is inadequate for capturing the complexity of the mapping.

An alternative bifurcation of spatial cognition may provide a clearer path for making comparisons. One possibility is that the visual system distinguishes between metric information and qualitative information. When placed in this framework, many levels of spatial representation can be shown to incorporate both kinds of information. As touched on earlier, there is experimental evidence that object representations may include metric information stored as views and qualitative information, as the relations between the views and object parts (Tarr & Pinker 1990). Likewise, neurophysiological studies (Paillard 1991) and implementations of exploration algorithms (Engelson & McDermott 1992) both suggest that representations of place also include both metric and qualitative information. No doubt the same holds true for the relations of objects in a scene. Given this sort of distinction, the connection between spatial prepositions and spatial relations may be revised, prepositions being used to capture qualitative relational information between parts, objects, or places, and count nouns being used to capture shape information about parts, objects, or places.

In summary, L & J raise many interesting and important points about the connection between spatial language and spatial cognition. Bridges between the two domains must be built upon a sound foundation, however, one in which there are principled reasons for the connection at both ends. Indeed, such cross-fertilization is essential if we are to understand how human information-processing progresses from perception to cognition.

NOTE

1. Marr's (Marr & Nishihara 1978) and Biederman's (1987) theories are both limited to *visual* representation and essentially only to object shape. In contrast, a complete theory of spatial representation would include models of the environment (Paillard 1991). For example, auditory perception in humans contributes only location information (e.g., for controlling eye movements, Jay & Sparks 1987) but tells us nothing specific about object shape (although identification may be made through auditory features).

## Prepositions aren't places

Barbara Tversky and Herbert H. Clark

*Department of Psychology, Stanford University, Stanford, CA 94305-2130*
**Electronic mail:** *bt@psych.stanford.edu & herb@psych.stanford.edu*

People have extensive knowledge of objects and places, what things are and where things are, and they are able to talk about both. Landau & Jackendoff (L & J) argue, however, that although language offers people a rich means for talking about objects, it offers only a meager means for talking about places. Moreover, they argue that these differences in language have parallels in the "nonlinguistic disparity between the representations of 'what' and 'where'" in the brain. In the course of their argument, L & J take us through an insightful review of both spatial cognition and the language of objects and location. In the end, however, their argument collapses. Let us see how.

L & J's first claim rests on the premise that objects are encoded by nouns and places by prepositions. "In English," they say, "objects are represented by count nouns and places are represented canonically by prepositions or prepositional phrases." Although they add the qualification "or prepositional phrases," they proceed in their argument as if they hadn't. They then note that nouns are much more highly developed in languages than prepositions. Languages have many more nouns than prepositions, nouns have more elaborate conventional meanings than prepositions, and so on. It follows, they maintain, that languages have a much richer means for talking about objects than places. But this only follows if the initial premise is correct, and it does not seem to be.

Prepositions are neither necessary nor sufficient to represent places or locations. Verbs like "support," "hold," and "contain," for example, take subjects that denote locations, without prepositions, as in "The table supported the statue," "The vase held the jewels," and "The boat contained illicit cargo." Likewise, there are no locative prepositions in "Chicago is windy," "His shoes are muddy," "His garden is full of flowers," or the prototypic expressions of place, "here" and "there." Prepositions are merely relational terms. In "The statue is on the table," "on" denotes a relation between an object (the statue) and a place (the table) just as the verb "support" denotes a relation in "The table supported the statue."

Places, therefore, are entities, not relations between entities, and are regularly represented by count nouns. "Table," "child," and "apple" can be used to express either objects or places, depending on your perspective. If you "move a table," "watch a child," or "pick up an apple," you are treating the table, child, and apple as objects. But if you put "a chair at the table," "a hat on the child," or "a knife through the apple," you are treating them as places. Yet places are entities we can relate other entities to, as expressed in *locative predicates* such as prepositional phrases (prepositions plus noun phrases). In English, the answers to "where" questions (like "Where did you put the chair?") are ordinarily locative predicates ("At the table"). They can never be prepositions alone ("At") unless the location can be inferred. In locative predicates, the place information carried in the noun phrases is essential.

Prepositions, L & J argue, make reference to only a limited set of properties of locations. "On" in "on the couch" calls on only a few of the many properties of couches. From that, L & J seem to conclude that people's representations of locations must themselves be limited. This argument has serious flaws. First, prepositions pick out only a fraction of the locative information that is actually exploited in locative predicates. Consider "There is/are X on the couch" and think of all the subtle inferences you make about location when X is "an adult," "two lovers," "a cushion," "a spider," "dirt," "a label," "four legs," "antimacassars," or "a bug (hidden listening device)." This locative information, we suggest, is in principle unlimited. Second, L & J seem to imply (in Whorfian fashion) that the poorer the prepositional system in a language, the poorer the representations of location. By this logic, speakers of Tseltal should have empty representations of entities viewed as places because Tseltal has only a single all-purpose preposition. On the contrary, speakers of Tseltal have an extraordinarily rich system for describing places. It just happens not to rely on prepositions (Levinson, in press). In short, the poverty of the semantics of prepositions cannot be used to argue for a poverty in the representation of place.

Intricate location information underlies our very conceptions of many objects. As L & J argue, languages have large vocabularies for classes of objects, and many of these terms are understood as referring to shapes. Yet critical to shape, as L & J note later, is location. An object gets its shape from its parts, but the parts must be arrayed in a particular configuration (Tversky & Hemenway 1984). The difference between a T and a +, or between a coherent object and a collect of parts, is precisely how the parts are located with respect to each other. Place, then, is critical to mental representations of objects.

Place must also be represented in the semantics of object names. Terms for parts, such as "back," "fender," and "leg" (of a person, chair, or table), must represent the location of the parts with respect to the wholes they are parts of. One large class of nouns must be represented in part by the potential locations of their denotata, for example, "blankets" (they go on beds), "roofs" (they go on buildings), and "plugs" (they go in holes). That makes it easy to turn them into denominal verbs that express subtle locational relations: "We blanketed the bed, roofed the garage, and plugged the holes" (Clark & Clark 1979). Another large class of nouns must be represented in part by the locative relations of other things with respect to their denotata, for example, "dock," "lodge," and "list." These are also useful as denominal verbs describing special locations: "We docked the boat, lodged the guests, and listed the groceries we needed." So place is inherent to – and richly represented in – the understandings of countless nouns.

A second important claim in L & J's target article is that "there are significant differences in the geometric richness with which objects and places are encoded," specifically, that "detailed geometric properties [of the object] – principally its shape. . . . – are represented," but for places, "only very coarse geometric properties are represented, primarily its [the object's] main axes." This could be a claim about encoding in language or a claim about encoding in the brain, or a claim about both. Clearly, language can be used to describe both objects and places vaguely or specifically. Objects can be referred to by nominals that convey minimal geometric properties, like "this," or "instrument," or "thing." Moreover, labels may be understood differently from their conventional meanings. In a given situation, we may know exactly what "thing" is being referred to, and encode quite specific features, yet in cases where terms and objects are unfamiliar, we may encode only vague features from specific terms. Also, there is no simple correspondence between what the perceptual system finds easy to represent and what speakers of a language find easy to talk about. Yes, people are good at identifying shapes and good at describing them for others to pick out. However, people are also good at identifying faces, and yet bad at describing them for others to pick out.

Similar to linguistic representations, mental representations of objects and locations may be rich or coarse. Given our ability to catch fly balls, sink holes in one, return tennis serves, play the piano or violin, avoid obstacles in familiar environments in the dark, and much more, it would be rash to claim that places are mentally represented any less richly than objects. Success at any of these activities seems to depend on detailed representations of geometric and other properties of objects with respect to locations. The neuropsychological literature on separate systems for "what" and "where" does not, as we understand it, include evidence for differences in coarseness of representation in the brain.

L & J's unifying thesis, then, remains unproven. Nouns and prepositions may very well differ in richness, as L & J have argued, but our linguistic and mental representations of objects and places do not differ in parallel ways. Representations of both objects and places may be rich or coarse, in language, in the brain, and in the mind.

## From observations on language to theories of visual perception

Johan Wagemans
*Laboratory of Experimental Psychology, University of Leuven, B-3000 Leuven, Belgium*
**Electronic mail:** *fpaas10@blekul11.bitnet*

Landau & Jackendoff (L & J) explore the language of objects and places to learn more about their underlying representations. They conclude that objects and places differ considerably in the geometric richness with which they are encoded. Object nouns imply more detailed representations that those that are currently suggested in theories of visual object recognition, whereas spatial prepositions imply much sparser spatial information than must be available for complex visual tasks such as eye-hand coordination and navigation. Although this linguistic perspective on mental representations provides some challenging hypotheses, there are many gaps to bridge when going from observations on language to theories of visual perception. I think L & J jumped to unwarranted conclusions in some cases. I will discuss two of them, formulating a question in relation to each of them in turn.

First of all, L & J propose some enrichments of spatial representations of objects that are needed to allow names for spatial parts (such as top, bottom, front, back, sides, and ends), for objects best described as surfaces (such as records and lakes), for "negative" object parts (such as ridges and grooves), and for containers. These enrichments do not seem too remote from current models of object representations, but I have my doubts whether all the fine distinctions needed can be derived from the available visual input. I am strongly convinced by Marr's (1982) argument that there are two constraints determining the content of a representation: further task requirements (such as being able to speak about it), heavily stressed by L & J, and the possibility of perceptual recovery from images or intermediate representations, unwarrantably neglected by L & J. For example, I can easily imagine that the visual system can compute object and part axes based on elongation or symmetry, but I do not see any obvious computational scheme to distinguish top from bottom or front from back. I believe that perceptual representations are what they are because of bottom-up as well as top-down constraints (available input information and task requirements, respectively). Perhaps later nonperceptual representations are more fully based on requirements from language and other cognitive faculties only. In sum, my question to L & J concerns whether or not they would consider their enriched formats of object representation to be perceptual. To answer this question L & J will need some dividing line, which might be on- versus off-line, bottom-up versus top-down, or some other criterion they like more.

A second concern is about the criteria for distinguishing between those task requirements that have implications for the underlying representations and those that do not. For example, L & J explicitly mention that metric information must be encoded to support performance in fine motor tasks such as reaching and grasping, whereas most of that information seems to be filtered out in the representations underlying spatial expressions. Only the very gross geometry of the coarsest level of representation of the object seems to be preserved. Even within the domain of linguistic expressions, some observations are used to derive statements about the representations of objects and places, whereas others are done away with as having to do with historical and pragmatic issues rather than principles of spatial representation (sect. 2.9, para. 3). For example, large vehicles (such as buses and yachts) are conceptualized either as containers that one is *in* or sorts of platforms that one is *on*, whereas small vehicles (such as cars and rowboats) are only conceptualized as containers. Although this seems to suggest that the representation of the reference object must include a

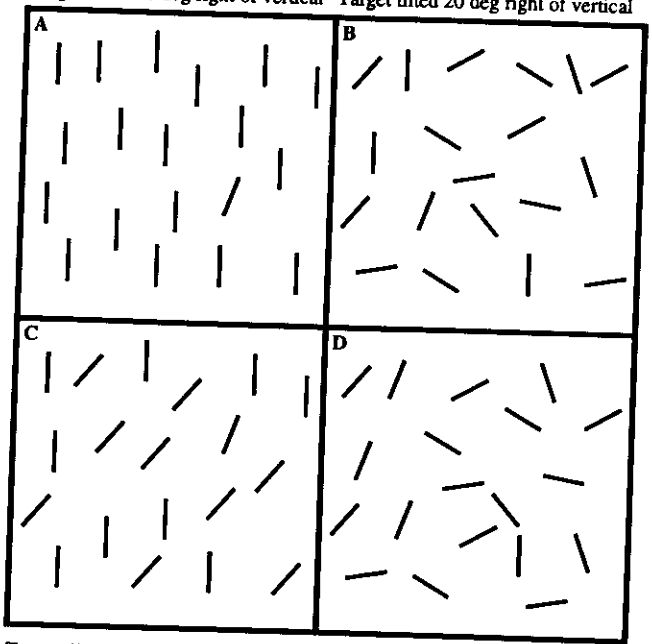



Figure 1 (Wolfe). How do we talk to ourselves about orientation? It is easy to find a target that is oriented 20 deg away from vertical in a homogeneous array of vertical items (1a). If the background items are heterogeneous it is quite hard (1b) even though most of the distractor lines differ from the target by more than 20 deg. It is hard even if there are only two distracting orientations (1c) and it is hard even if the target is vertical (1d). We cannot ask ourselves about "20 deg" or even about "vertical."

size parameter (which can be quite metric!), L & J refer to this as a special situation that involves conventionalized concepts. I would therefore like to have a principled account of when to use an observation on language (or any other perceptual or cognitive capacity) to derive predictions about the underlying representations. Or do L & J believe in an abundance of representations of objects and places, one for each task to the limit?

In general, I believe that L & J have formulated some challenging hypotheses to be tested by perceptual research. For example, the representation of a container (such as a cup) as a thickened surface that encloses a cylindrical space (sect. 1.6, para. 1) makes its functional use or affordance (Gibson 1979) much more explicit and its perceptual relevance might well be tested with more primitive "containers" such as bowls made from leaflets. Similarly, L & J's basic conjecture that the relatively simple shape specifications observed in the prepositional system reveal the extent of detail possible in object descriptions within the "where" system (sect. 3.2.2, para. 5), has much intuitive appeal and seems quite testable. L & J's case would be much stronger, however, if they clarified some of the conceptual fuzziness identified above.

ACKNOWLEDGMENT
The author is supported by a grant from the National Fund for Scientific Research (Belgium).

## Talking to yourself about *what* is *where*: What is the vocabulary of preattentive vision?

Jeremy M. Wolfe

*Center for Clinical Cataract Research, Harvard Medical School & Brigham and Women's Hospital, Boston, MA 02115*
**Electronic mail:** *wolfe@cccr.bwh.harvard.edu*

Landau and Jackendoff (L & J) argue that our language about *what* something is and about *where* something is reflects different representations of the *what* and *where* of objects. Although the language the authors discuss is the language between individuals, we can also consider the *language* we use to talk to ourselves. Of course, this is not a language in the normal sense but rather the internal vocabulary used by one part of us to tell another part what to do.

In visual perception, this internal speech is used routinely in certain visual search tasks. In a visual search task, subjects look for a target item among distractor items. In one standard version, we measure reaction time (RT), the time it takes to say "yes" there is a target or "no" there is not. In some searches, RT is roughly independent of the number of distractor items (set size). These searches make use of *parallel* mechanisms that can process all items in the display at once. Other searches appear to involve *serial* examination of objects at a rate of about 1 item every 50 msec. It is widely believed that the visual system can be broadly divided into a *preattentive* stage, in which a limited set of basic features (color, orientation, motion, etc.) can be processed in parallel across the visual field and a *subsequent* stage, in which more sophisticated processes can do elaborate tasks of object recognition, but only over a limited portion of the field (see Treisman 1986; 1988; Wolfe 1992 for more detail).

To be useful, these parallel processors of basic features must be "spoken to" by other, higher processes. For example, consider a visual field full of spots of different colors. You can *see* all those colors all the time. If you want to determine whether this multicolored mass contains any green spots, however, you need to be able to *ask* the parallel color process to mark the green spots in some fashion. It is certainly possible to use parallel mechanisms to find one color in a heterogeneous array (Duncan 1989; Wolfe et al. 1990). For present purposes, it is interesting to compare this internal request from "you" to a feature processor with your overt speech about objects.

Our clearest data on this matter come from visual search for oriented lines. If one searches for a target of one orientation among homogeneous distractors of another orientation, search is very efficient with slopes of RT × set size functions near zero, reflecting the contribution of parallel orientation processing

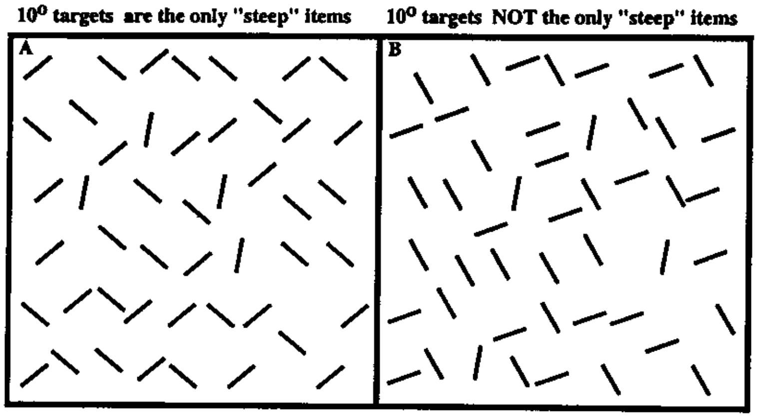



Figure 2 (Wolfe). So . . . how *do* we talk to ourselves about orientation? Search is quite easy if the targets possess a unique categorical attribute ("steepness" in 2a). Even if all geometrical relations are preserved, search is more difficult if the target is not categorically unique (2b). We can ask ourselves about "steep," "shallow," "left," and "right."

(Foster & Ward 1991). Search can become painfully inefficient when one searches for the same target among multiple distractor orientations or even among just two flanking orientations (Moraglia 1989; Wolfe et al. 1992). This is shown in Figure 1. In 1a–c, the target is tilted 20 deg to the right of vertical. When there is one distractor orientation, search is easy. When there are two or more, search becomes inefficient, even though the distractor orientations are quite different from the target. Our internal "language" apparently does not allow us to ask for "20 deg" as we would ask for "red" or "green." In 1d, the target is vertical but it is still not easy to find, suggesting that even "vertical" may not be in the internal vocabulary.

What *is* in the vocabulary? It would be unfortunate to have an orientation processor that only worked when faced with a unique item on a homogeneous background, as that situation is fairly rare in the real world. Fortunately, there is a vocabulary with which we can talk to ourselves about the orientation of lines. Like the "where" system descriptions proposed by L & J, that vocabulary appears to be extremely restricted. Specifically, it is restricted to *categorical* descriptions of the target. In orientation, it is not possible to talk to yourself about 20 deg lines but it is possible to ask the orientation processor for "steep," "shallow," "left," or "right." This is seen most convincingly in stimuli like those in Figure 2. In each panel, there are four targets. In 2a, the 10 deg targets are the only "steep" items among +50 and −50 deg shallow distractors. They are relatively easy to find. In 2b, the same 10 deg targets are the "steepest" items among −30 deg and +70 deg distractors. These are not easy to find because the −30 deg distractors share "steepness" and the +70 deg distractors share "rightness" with the targets. Note that in both 2a and 2b, the targets are 40 deg different from one type of distractor and 60 deg different from the other. Efficient search is possible only in 2a, however, where the target is categorically unique (Wolfe et al. 1992).

In size, we have found that the "vocabulary" is limited to *big* and *small*. It does not appear to be possible to search for the target of *medium* size (Wolfe & Bose, unpublished). In color, the vocabulary is probably restricted to categorical color names: Thus, a command for "olive drab" cannot restrict search to drab olives among olives of other shades of green and brown. Turning to spatial relations, it is much harder to find the "red thing near the vertical thing" than it is to find the "red vertical thing" (Grabowecky & Khurana 1990; Wolfe et al. 1990). It is very hard to find a "thing that has red and green parts" but comparatively easy to find a "red thing with a yellow part" (Wolfe & Friedman-Hill 1992).

As with L & J's "where" language, this internal language of preattentive commands appears to be profoundly limited. Nevertheless, it seems to be rich enough to get the work done in the real world. If we can use simple parallel processing to restrict our attention to "steep, brown things with green parts," we can afford to use capacity-limited "what" processes to determine whether the attended item is a redwood or a maple.

Categorical terms like "steep" and "shallow" in the preattentive vocabulary probably reflect underlying perceptual limitations and not perceptual learning. There is little evidence, for example, for an ability to learn orientation categories like "2 o'clock." It would be interesting to determine how much of the "where" aspect of spoken language is constrained by these properties of preattentive vision.

# Authors' Response

## Whence and whither in spatial language and spatial cognition?

Barbara Landau[a] and Ray Jackendoff[b]
[a]Department of Cognitive Science, University of California, Irvine, CA 92717 and [b]Linguistics Program, Brandeis University, Waltham, MA 02254
Electronic mail: [a]blandau@orion.uci.edu; [b]jackendoff@brandeis.bitnet

We are grateful to all the commentators for their consideration and gratified that so many of them found our argument at least partly persuasive. Considering their critiques, we find ourselves raising questions that are sometimes only hinted at in the commentaries themselves. We apologize in advance for schematizing commentators' arguments down to points and lines, a necessary strategy for reasons of everyone else's patience; and we hope allowance will be made for misunderstandings on our part. In the remarks to follow, we summarize the logic of our argument in such a way as to highlight our responses to the commentaries.

### R1. The codings relevant to our theory and the nature of their interfaces

In asking how language and spatial representation map onto each other, we first sketched a model of how we see spatial representations interfacing with other components of the representational system (Figure 1 in the target article). These are the particular codings relevant to our article, and their interfaces (or partial homologies) with other codings:

1. Linguistic representations: Most prominently for us, syntactic structures represent the grammatical structure of linguistic utterances. These interface with phonological structure for access to the sensorimotor periphery and with "propositional" semantics (as described, for example, by Jackendoff's 1987b conceptual structure) for access to spatial representation.

2. Spatial representation: a modality-independent coding in geometric form that contains the elements of spatial understanding. Some of the key elements include figure and reference objects, regions, places, paths, and trajectories. **Mainwaring** calls this system "sublinguistic," as though it were on a plane of understanding lower than language. We would prefer to think of it as a parallel system of equal status. In our target article, we assumed that spatial representation interfaces with the following:

codings proprietary to the visual, haptic, and auditory systems, so that one can use vision, touch, and audition to understand one's spatial environment;

codings involved in formulating action, so that one can use one's understanding of spatial layout to locate objects or to navigate;

codings proprietary to the linguistic system, in particular the level of meaning, so that we can talk about space.

Critical to this scheme is the notion of the interface between representational systems. A familiar example of an interface can be found within the linguistic system. There, information is represented in terms of (among other things) a phonological and a syntactic coding. These codings are partially homologous, in that they both involve a segmentation into words arranged in linear order. On the other hand, other aspects of the two codings diverge. Phonological coding divides words into syllables and syllables into individual speech sounds; these further divisions do not appear in syntax. Syntactic coding assigns to each word a part of speech (noun, verb, etc.) and groups words into larger syntactic constituents (noun phrase, verb phrase, etc.); these features are not reproduced in phonological coding.

Such partial homologies among distinct codings constitute the vocabulary they have in common. They are therefore the only means by which different codings can communicate with one another. For example, the phonological and syntactic codings of language are kept in synch by virtue of the correspondences in the linear order of words. (To put this in the terms of current neuroscience, the partial homologies are the informational links used in *binding* different codings together.) Because the homology is only partial, though, the interface behaves like a filter: Only certain aspects of each coding can filter through to the other.

The starting point of our target article is that there needs to be a partial homology between spatial representations and language that enables us to talk about what we see (or touch, or hear).

At the end of the paper, we explore the possibility that there are at least two distinct components of the spatial representation system: (1) object shape representation, which encodes the three-dimensional shapes of individual objects in the environment, and (2) object location representation, which encodes the spatial layout of these objects in the environment.

We propose that these two codings interface by individuating objects the same way (the same number of objects in both codings) and by schematizing objects in terms of one or two principal axes and a reference frame. Other than that, they diverge: (a) Object representation elaborates object shape in much greater detail and ignores location. (b) Object location representation (i) uses schematized objects to help define regions and trajectories in terms of which other schematized objects are located, ignoring much of the detailed coding of object shape that is represented in the object system. Object location representation can also (ii) encode location in relatively coarse terms, including qualitative distinctions in distance and direction.

Notice that it is impossible to view any of these systems as "filtered" or "unfiltered," in any absolute sense, as **Mainwaring** implies. Each is filtered relative to the other. Object shape representation encodes schematic and detailed shape but nothing about location; object location representation encodes schematic shape and location but not detailed shape. Language encodes aspects of shape but also abstract concepts such as quantification, totally absent in spatial representation. In fact, we are interested in just what these representations have in common so that an adequate interface can exist.

Spatial representation has separate interfaces with the motor system and the linguistic system. This means that different properties of spatial representation encoding could in principle appear in each of these other two. As we point out in the target article, there are great differences between the aspects of spatial representations that filter through into the motor system and those that filter into language. **Hoffman** points out the need to account for precision dribbling and shooting; **Bridgeman** describes elegant experimental evidence attesting to the fact that the motor system is encapsulated, at least from the processes involved in explicit perceptual judgments (see also **Ingle**).

Quite a number of commentators (**Hoffman, Tversky & Clark, Bridgeman**) object that we have neglected the precision required by the motor system. In section 3.1, however, we point to the precision of the motor system, and we conjecture that the interface between spatial representation and the linguistic system filters out metric precision. As Figure 1 of our article shows, such precision and filtering are a natural possibility of the model as presented and it becomes even more natural with possible additional structure as suggested to us by **Brown** (whose commentary we discuss later).

## R2. The logic of our argument

To bring linguistic evidence to bear on spatial representation, we made use of a basic premise: If linguistic judgments can be based on spatial properties of objects then the information involved in the linguistic judgment must be able to pass through the interface between spatial representation and language.

To pass through the interface, the information must be available in spatial representation. Thus, for example, if language can characterize a "hole" as wide or narrow, having sides, top, and bottom, spatial representation had better contain an entity with which the term "hole" can be associated, and this entity had better have properties relevant to determining width or picking out sides, top, and bottom.

This is a point on which we believe **Slobin** misses our argument. He writes that object names do not refer to shapes of objects; rather, they refer to the objects. We agree. The issue of shape enters in the following way: In agreeing or disagreeing with the statement, "This is an X," where "X" is an object name, one has to evaluate the referent of "this." If this evaluation is carried out visually, then, as we have pointed out, one of the crucial factors must be its shape. We conclude from this that names for physical objects must be mentally linked to, among other things, spatial encodings of their shapes. Therefore, spatial representation must be detailed enough to distinguish all the kinds of objects that can be distinguished by shape.

Should every distinction appearing in language be recoverable from the *visual image?* (Here we distinguish a visual image from the amodal *spatial representation* linked to it.) **Wagemans** points out that parts such as top and bottom of an object may not be recoverable from the image, and this prompts him to ask whether our "enriched" representations are perceptual or not. Actually, top and bottom (differentially marked ends) can often be recoverable as the opposite ends of the object's vertical axis. Exactly which end is the top and which is the bottom will also often correspond to information recoverable in the image (e.g., Fillmore [1975] suggests that the front of an animate object is determined by the positioning of its principal perceptual organs).

More broadly, however, we see spatial representation as a relatively central coding that draws on – but is not equivalent to – information from all the perceptual systems. Our point is that the distinctions uncovered by language must be represented within the spatial representational system. Some of these properties can be mapped from the visual image, for example, the object's principal axis. Others cannot be recovered from the visual image alone. For example, there is nothing in the image (that we can think of, anyway) that would correspond to the spatial notion of *region* (or, in fact, to the principal axis of a nonrotating sphere). But each of these – whether or not it is mappable from the image – must be encoded in spatial representation. It is an empirical problem to determine which elements of spatial representation can filter through interfaces all the way up from (or down to) the image, which elements can be linked to some intermediate form of representation (e.g., Marr's [1982] $2\frac{1}{2}$-D sketch), and so forth. In other words, the nature of the interface between spatial representation and strictly *visual* representation is an issue for further research in computational vision (and similarly for the other modalities that interact with spatial representation).

Following from our basic premise, we argue two points. First, the richness of spatial language points to a richer conception of spatial representation than is usually imagined by visual theorists. For example, terms like "hole" require us to include in spatial representation not just objects and their parts, but also "negative parts"; terms like "top," "side," and "front" require us to include schematizations of objects in terms of orienting axes; terms like "edge" require us to include schematizations of objects in terms of thickened bounded planes.

More central to our target article, the variety of ways in which language describes the spatial layout of stationary and moving objects requires us to include in spatial representation the abstract notions of region and trajectory, which are not directly present in perceptual input. The richness of linguistic description sets a lower bound on the richness of spatial representation: We must have at least those elements. There may, of course, be more.

The second aspect of our argument is that the primitive notions found in the spatial language for location have no evident explanation in terms of language itself. They do, however, have an obvious explanation in terms of geometrical notions that appear altogether natural in spatial representation. Such notions include, for example, the figure/ground opposition, the schematic axes of objects in different frames of reference (viewer-centered and object-centered), and the notion of relative distance from an object (interior vs. contact with surface vs. proximal vs. distal). Because these notions seem quite natural extensions of many current approaches to object and place encoding, we claim that the degrees of freedom found in language are homologous to parallel degrees of freedom in the description of spatial representation. That is, we are claiming that the partial homology between language and spatial representation includes at least these notions. No commentator seems to object seriously to this part of the argument.

The more controversial part of our argument was based on the extremely limited number of parameters involved in distinguishing preposition meanings in conceptual structure, resulting in such a small number of prepositions. Why should there be so few? We offered two possibilities. The Design of Language Hypothesis attributes the limitations of language to the interface: In present terms, spatial representation has a much richer selection of parameters, but only a limited number of them enter into the homology with language. The Design of Spatial Representation Hypothesis says that the interface is in fact pretty much as rich as it can be, and that the limitation actually lies in spatial representation itself. Note that with each of these hypotheses we explicitly reject the notion of isomorphism between language and spatial representation, despite **Friederici's** suggestion that we do claim such an isomorphism.

We are the first to admit that (as **Hoffman** points out) there is nothing in the linguistic evidence per se that can decide between these possibilities, or, as we suggested, some combination of the two. What intrigues us is that, with the exception of metrical precision, the location system in language has just about the right properties to interface with the "where" system described by neuroscience. We therefore conjecture (a weaker term than "claim") that what we find in the language of places has a fairly strong homology with the coding of objects and places situated in the "where" system of the brain. In particular, one kind of object description gets through the interface between spatial representation and language for naming (the "whats"), and another kind of object description does so for locating the "wheres."

We are pleased that quite a number of commentators saw our hypothesis as a good start, or better yet, as plausible, even providing converging evidence. **Bryant** agrees with our general approach to the mapping of space and language, emphasizing that the representations derived from larger linguistic units – discourse – also map onto spatial representations. **Wolfe** provides evidence for categorical coding, parallel to that in language, involved in preattentive processes, for example, distinctions between steep and shallow, or large and small. He suggests that the rough search accomplished with these processes can efficiently support the more effortful, capacity-limited search for more detail as is needed in the identification of particular object kinds. We regard such evidence as critical in illustrating the existence of categorical coding schemes derived from empirical studies quite different from our own. Similarly, **Brown** provides converging evidence for categorical coding, although she argues that it is a characteristic of the left hemisphere rather than language per se (more later). **Feldman** agrees that certain nonmetric regions will be more naturally represented than others, although he proposes a functional reason why (again, more later).

A number of commentators disagree with our methodology, our approach to the linguistic description of objects and places or our approach to spatial representation. In what follows, we will do our best to answer their objections.

## R3. Our interdisciplinary methodology

The basic stance of our target article is that evidence from language can provide insight into the structure of human spatial cognition that can extend our understanding as derived from standard psychological and neuropsychological techniques. Our assumption is that linguistic analyses should dovetail with and supplement these other approaches. As with any scientific problem, all available tools should be brought to bear on spatial cognition. The analysis of spatial language vis-à-vis nonlinguistic spatial representations seems a fruitful approach.

A number of commentators seem to have misunderstood our methodology. **Brown,** coming from the viewpoint of vision research, criticizes us for using tools that are specific to the left hemisphere, given that spatial cognition involves both hemispheres. This criterion would exclude the use of linguistic evidence altogether. We agree with Brown that right hemisphere evidence is equally important in providing an overall resolution of the problem, and in fact her comments about hemispheric specialization, if correct, add an important clarification to our position (see below). But at the same time, we think it is critical that the hemispheric differences she describes are found in the context of different linguistic instructions: Left hemisphere superiority is found with instructions using locations encoded by English prepositions such as "whether a dot is *above* or *below* a line," whereas right hemisphere superiority is found with locations encoded by phrases, such as "whether a dot is *within 3 millimeters of* a line." Analysis of the language itself suggests the same kind of differential encoding of space (categorical in the "left hemisphere" instructions, metric in the "right hemisphere" instructions) that Brown describes. Thus, evidence from language and visual cognition complement each other, as we believe they should.

Other commentators seem to misinterpret our approach from the other end, seeing it more as an isolated exercise in linguistic description. **Friederici** accuses us of circularity in using psychological evidence to help sharpen our hypothesis, apparently not understanding the notion of converging evidence. She also feels that our empirical data are not convincing, as they only show absence of shape representation when filtered by a spatial preposition task (Landau & Stecker 1990). But that is precisely the point: In our view, it is the preposition context that elicits the kind of representation that ignores detailed object shape.

**Ingle** tells us that we should compare object and spatial (locational) vision not in terms of their implicit complexity but in terms of which representations normally play a role in conscious deliberation. We see no reason for this restriction. Both aspects must be addressed: what information the mind processes, and how this information contributes to conscious experience.

## R4. Our approach to the linguistic representation of objects and places

**R4.1. Objects.** It is surprising that we did not have many complaints about our emphasis on the importance of shape in object naming. We say it is surprising because we have often encountered versions of this objection during colloquia based on this material. As an anonymous *BBS* referee put it,

> Space enters only very indirectly into naming. Terms like "dog" or "clothespin" seem to apply to kinds of objects, regardless of what it is that determines membership in a kind. . . . Spatial representations discussed here may enter directly into visual object recognition, but they seem to enter only weakly and derivatively into naming.

In just this way, **Slobin** objects to our emphasis on object shape as an important diagnostic for the variety of human object categories expressed by language. We reiterate that shape is involved in naming to the extent that one must establish by perceptual means the truth value of some statement, "This is an x." We believe that Slobin misrepresents our position when he objects that the word "frog" does not describe the shape of a frog, but rather references the object: "When I say 'frog,' I do not describe the shape of a frog, but evoke whatever representation of frog you have in your mind." We agree. And, although we showed in sections 1, 1.1, and 1.2 that shape recognition is crucial for object naming, we do not mean to exclude the other criteria Slobin mentions (see Note 3 in our target article for explicit acknowledgment of these).

Once again, our concern is the mappings between spatial cognition and language. We are therefore forced to deal with the fact that object shape is a key index of object kind. The importance of other factors should not diminish the significance of object shape. Likewise, the importance of object shape in naming object kinds does not automatically mean that shape should or should not count for other natural concepts. For example, at the moment, there is no independent theory of what makes crumpling perceptually different from crushing (to use **Slobin**'s example); but perhaps someday there will be, in which case we would want to draw on those representations in establishing links to those verbs. In contrast, there will never be a theory of what makes a wife perceptually different from a mistress (impossible, as Slobin observes), and we have no intention of claiming that object shape – or any perceptual property – should distinguish these.

**R4.2. Places.** There were far more objections to our linguistic treatment of location, in particular, to our use of English prepositions as the domain upon which to build our theory. It is curious that the commentators objected for very different (often incompatible) reasons. Some commentators (**Tversky & Clark, Slobin, Bennett**) objected that the description of location is actually infinitely detailed and that our restriction to prepositions (English yet!) had the effect of artificially restricting the range of meanings that could be expressed. A partially overlapping set of commentators (**Bridgeman, Bryant, Bennett, Feldman, Friederici, Heidorn & Hirtle**) objected that there are not more prepositions just because there are no more spatial relationships to be described! They say, in effect, "Of course there should not be many spatial relations in language. There are not many spatial relations, period. All you need is a frame of reference that establishes three spatial axes, and you have everything you need."

Considering the latter objection first, we think such limitations on possible spatial relations are actually not so self-evident. First, as the study of geometry shows, numerous spatial relationships are possible that simply do not get encoded by the basic vocabulary in English or other languages (All you have to do is look in a not-so-elementary geometry book to discover that there are a number of formally established geometries – e.g., Euclidean, similarity, projective, topological – that form a hierarchy, with Euclidean containing the finest-grained metric information.)

Although there are many well-formed geometric properties and relationships, however, not all of them are lexically encoded. For example, there is no word for "3 inches away from" despite the fact that this is a well-formed spatial relationship in metric geometry. On the other hand, there are words that represent nonmetric relationships, such as contact or attachment ("on"), containment ("in"), and so on. Our question is: Why should language express (as single morphemes) the particular relations it does, some of which (e.g., "along" and "throughout") are fairly eccentric, while not expressing others (e.g., our hypothetical relations "plin," "sprough," and "betwaft") that are prima facie no more complex?

Our conjecture is that the relations we find in language are built up from an exceedingly sparse primitive vocabulary that happens to allow the former relations but not the latter. The particulars of this primitive vocabulary are empirical issues, which is why we devoted more than half of our target article to running through the boring details of English prepositions. Precisely because we have this internal vocabulary, it *seems* self-evident that there are not many spatial relations "in the world." We are reminded of a passage from Fodor (1980, p. 333):

> *From in here* it looks as though we're fit to think whatever thoughts there are to think. . . . It *would*, of course, precisely because we *are* in here. But there is surely good reason to suppose that this is hubris bred of an epistemological illusion. No doubt spiders think that webs exhaust the options.

Other commentators (**Bennett, Heidorn & Hirtle, Slobin**) propose a somewhat different (though related) reason why there are so few spatial relations expressed in language, one that has a distinctly functional flavor. (We actually anticipated this objection in our discussion of the Design of Language Hypothesis.) The idea is that spatial representation of location is very rich, but that spatial language is only detailed enough to get the message across – the rest of the detail is taken care of by the visual system or by pragmatic understanding of the situation. This way, language can be processed on line more efficiently.

There are two problems here. First, although **Slobin** suggests that pragmatics derived from the scene can take care of disambiguating the spatial relationship, we believe he wildly overestimates the usefulness of such pragmatics, especially for child learners. He says, "The words 'frog' and 'jar' already tell you a good deal about the possible relations between these two objects." They do? The frog could be in the jar, but it could also be on the jar, next to the jar, behind, below, or above the jar. . . and these are only the lexically encoded concepts! What about 3″ from the left front of the jar, 2.5 mm. into the jar? Or it might have just its head and its left hind leg in the jar and the rest out. Although pragmatics of the sort Slobin invokes might help adults who already know how the language encodes spatial relationships, such information could not possibly allow the learner to narrow down the target meaning on the basis of observation alone. On the other hand, perhaps Slobin means to suggest that humans are predisposed to represent only certain spatial relationships by language, and so, given that predisposition, one can more readily narrow in on some target relationship. If so, we definitely agree.

The second problem concerns the very notion of relying on functional explanations of why we have so few prepositional meanings. We do acknowledge functional requirements as a possibility, but we are concerned that this kind of argument raises the danger of begging the question: How do you tell what counts as "enough detail," especially when it varies from language to language (as **Slobin** and **Tversky & Clark** point out)? As an example, **Bennett** suggests that "in front of" is natural, whereas "at the thick end of" is not, just because of the general usefulness of the former. But why should one be useful and the other not? What could "usefulness" possibly mean, in this context? Our own explanation of this particular case has nothing to do with usefulness; "at the thick end of" is unnatural as a lexicalized spatial relation because, like our hypothetical example "sprough," it refers to the shape of a geon.

We are interested in why these particular limitations exist and not others. The Design of Spatial Representation Hypothesis strikes us as an interesting direction to pursue. It may be wrong, but the potential payoff if it is right encourages us to stick our necks out.

Now back to the first objection: Why did we focus on English prepositions in the first place? Maybe it is really PPs (prepositional phrases) of potentially infinite complexity that describe location. First of all, can you imagine how long the target article would have been if we had described the semantics of English PPs?

More seriously, our basic argument was that it is possible to use evidence from language to provide insights into the structure of human spatial cognition. We therefore examined English prepositions not because we believe they are *isomorphic* to spatial representation, but because, after carefully filtering out the nonspatial senses (an onerous task in the linguistic literature that we have spared *BBS* readers), they provide a relatively pure expression of spatial relations. We argued at length that the meaning of a spatial preposition – its representation in terms of "propositional" structure – is a function that maps the reference object, canonically expressed by the object of the preposition, into a region. The verb of the sentence in turn locates the figure object in this region. (Incidentally, although **Friederici** supposes that "prepositions are considered to encode relations similar to formal logical relations," we know of no formal logical relations that behave like prepositions.)

**Bennett** and **Tversky & Clark** observe that the preposition alone cannot specify location; it is the preposition plus its object that specifies location. As just pointed out, we agree. The preposition expresses only a function. Our question is, What properties of the reference object can play a role in such functions, so that they serve to differentiate the resulting regions? Our answer is: very few.

More technically, we disagree with **Tversky & Clark**'s assertion that count nouns such as "the table" can represent places. Canonically, it is the prepositional phrase that represents the place (they call the PP a "locative predicate"; for detailed arguments against this common approach, see Jackendoff 1983, Chapter 4). We do agree with them that the meanings of many nouns and verbs include locative relations; however, formal analysis (Jackendoff 1983; 1990; Talmy 1980) discloses that such locative relations are of the same sort as we have described, namely, mapping objects into regions in terms of which other objects can be located. For example, the verbs "enter" and "exit" encode the two different directions *in* and *out*, each of which has the spatial properties we describe in our inventory of English prepositions.

Thus, when **Bennett** says it is a mistake to concentrate on prepositions and **Slobin** points out the unilluminating quality of the preposition systems in Turkish and Tzeltal, we think they miss the point. We agree that a full treatment of this topic would deal with the total variety of expressions for spatial location in all the languages of the world. We have just found English prepositions a convenient and telling entree into the subject. No doubt additional rigorous studies will provide information to sharpen or modify our hypotheses. We have certainly provided enough references to languages other than English and to parts of speech other than prepositions to see where further investigation might lead. And we welcome detailed theoretical descriptions of spatial meanings that differ from our own. For example, Tzeltal is often cited as a language characterized by geometrically rich spatial terms (which would not fit comfortably into our scheme). A careful reading of Levinson's (1992) analysis, however, shows that the system of terms fits remarkably well.

A separate issue is the important methodological question that **Wagemans** raises regarding our linguistic analysis. As we observed, languages often have idiosyncrasies of prepositional use that are not well predicted by our analysis. For example, in English one gets *in* cars but either *in* or *on* buses. He requests a principled way of determining which uses are regular and which idiomatic. Our reply here is of necessity a hedge. Language is a complex mixture of rule-regulated and idiomatic elements; much dispute among linguists concerns how this mixture balances out empirically, and how it should balance out in principle. For a simple example, there seems no principled explanation of why in American English one *makes* a decision but in British English one *takes* a decision. It is just a learned fact of language. In the area of prepositions alone, such works as Herskovits (1986) and Vandeloise (1986) are sizable studies concerned with this question. So, we suggest, there can at the moment be no principled answer to Wagemans's question. In our target article we attempted to extract what we judge to constitute the regularities; we acknowledge there may be disagreement. We think, though, that the general order of complexity in the meanings of spatial prepositions is not too much greater than our analysis suggests.

## R5. Our approach to spatial representation

A number of commentators wonder whether our characterization of spatial representation is altogether on the mark. **Deregowski** suggests that object representations are actually more "flattish" than three-dimensional. Unlikely, we think, but if so, we do not see what the consequences are for explaining how we talk about what we see. **Tarr** proposes that a 3-D representation can be dispensed with, and our analysis purged of "*ad hoc*" elements, by adopting a "multiple egocentric views" theory of object representation. **Feldman** proposes that location is represented in terms of potential causal relationships. **Corballis** suggests that spatial representation is an analog medium, more like Play-Doh than a Lego set.

A number of commentators worry about our hypotheses of how spatial representation and language connect. **Hoffman, Bridgeman,** and **Tversky & Clark** are concerned that we have ignored the precision metric coding evident in the motor system. They worry that if spatial representation were as coarse as our analysis of language indicates, there would be no way to account for the basketball stars of the world (not to mention the ability of 4-month-olds to "catch" moving targets – see our sect. 3.1). **Brown** suggests that the phenomena we describe can be explained by the properties of left versus right hemisphere processing.

As to the general componential framework for objects we have adopted, we agree that it is not the only possible way to characterize high-level vision. We submit, however, that any alternative theory must address the traditional problems of object constancy and spatial layout; in addition, it must now connect with the linguistic data of the sort we have presented. We are open to detailed and comprehensive alternatives.

In particular, we find **Tarr**'s view (on the face of it) insufficient: (1) Multiple egocentric views of single objects presuppose particular viewing distances and lighting conditions and must therefore be linked somehow to represent the same object over time and space. In any case, we do not get the punch line in his objection to our presentation. Would adopting his approach make any substantive difference to our theory? (2) Because haptic representations are clearly not based on the same surface characteristics as visual ones, Tarr's proposal provides no way to interface with the sense of touch, as our approach is designed to do. (Admittedly, none of the other specific visual approaches is currently designed to work for haptics.) (3) It is not clear how multiple egocentric views of individual objects can be integrated into a representation of spatial layout. (4) In particular, an observer anticipating his own path through a scene requires a representation in which the scene remains constant while the observer moves – clearly not an egocentric representation in Tarr's sense at all.

Finally, we do not find our proposed augmentations of the componential framework especially *ad hoc;* this is evidently a matter of taste. For example, without using something equivalent to our feature "hollow" versus "solid," how does **Tarr** represent the understanding of a closed box as hollow, and therefore potentially containing other objects? In short, although we find Tarr's approach potentially useful as a theory of the highest level of purely visual representations of shape (though not spatial layout), it can hardly be satisfactory as a comprehensive theory of spatial cognition.

**Feldman**'s causal approach to location has some of the same problems (although we find the basic thrust of his approach very interesting). We observed that some prepositional meanings incorporate forces (e.g., "run into the tree" and "lean against the wall"). But in general, a preposition like "in" expresses a spatial dependency, not a causal one. Food is not in the refrigerator for the same reasons that the comets are in the solar system or the fish are in the sea. The objects along the sight of a rifle are in different peril from that of houses along the beach or the pedestrians, trees, and tunnels along the road. What these cases share are their spatial relations; the various causal relations follow as inferences from knowing about the objects thus spatially related. And in any case, we wonder what independent means Feldman intends to use to discover which spatial relationships are causally related and which are not. It seems, from our vantage point, that his theory depends quite strongly on analyses such as ours to discover which of the locations are causally related. It seems to us that these are just the ones that are lexically encoded.

Insofar as we can understand **Corballis**'s nonsymbolic "unGADly" system, we do not see how it suffices for the tasks the spatial system must perform. Play-Doh cannot structure itself with precision. It requires an external agency with precise intentions to mold and interpret it. It has no internal structure over which similarity metrics can be defined; there is no way to establish a homology to other representations so that they can communicate with each other. In adopting something like the Marr (1982) theory of spatial representation, we have assumed that spatial representation contains some more or less analog degrees of freedom: the exact shape of generalized cones and the exact proportions, orientation, and curvature of axes. On the other hand, the degrees of freedom we have been most concerned with, such as the number of axes and the attachment of parts, are combinatorial in the usual sense. So in a way, we (like Marr) have weakened the strict analogue/digital distinction that **Corballis** and possibly **Mainwaring** assume.

As for the existence of metric precision in the motor system, we agree that this is a critical part of the puzzle. Spatial representation must clearly have adequate geometric power to account for metric capabilities that underlie motor tasks (while remaining "hidden," following **Ingle**), including object-focused tasks such as reaching and grasping a bottle versus a bowl and environment-focused tasks such as navigating through a town. We suggested that the Design of Language Hypothesis requires that all lexical items level out some degree of geometric detail and that the Design of Spatial Representation Hypothesis further levels out detail in the locational system. These two means of leveling result in an asymmetry in the detail with which objects are encoded when talking about "what" versus "where."

We acknowledge that a problem still remains: how to integrate metric properties with the coarse coding, depending on the task. For example, although we propose that objects are represented as points and lines in the "where" system, it is obvious that more detailed geometric information would be needed to grasp accurately a Ming vase versus a bowling ball. One possibility is that the "where" system is drawn on for the general trajectory of one's reach, whereas the "what" system is drawn on for the grasp. Another is that the "whats" and "wheres" are described in two different ways, both coarsely and in fine detail, and that these are drawn on differentially for different tasks.

A version of the latter hypothesis is offered by **Brown;** it is essentially a more elaborate version of the Design of Spatial Representation Hypothesis. The idea is that the linguistic system for location is closely homologous to spatial coding by the left hemisphere "where" system, which, like the linguistic coding, lacks metrical information. All metrical information is coded in the right hemisphere "where" system. That is, the loss of metrical precision and the introduction of discrete categories of location are properties of the interface between the left and right hemisphere spatial systems, not of the interface between the spatial system and language. **Tarr** suggests a similar segmentation of spatial representation, although without proposing brain localization. **Bridgeman** suggests a division of spatial cognition into "sensorimotor" and "cognitive," the former incorporating metrical information, the latter, not. If such further divisions of spatial cognition come to be supported independently, it will be a stunning result for us, as it will bring the linguistic facts even closer to those of spatial understanding.

We will end with a serious problem for our approach that has been raised by **Olson, Tarr,** and **Tversky & Clark,** as well as by many colleagues at colloquia where this material has been presented. The language used to express the relation of parts to objects is identical to that used for configurations of independent objects. To adapt one of Olson's examples, we speak both of a nose on one's face and a fly on one's face. If, as we have claimed, the "where" system is implicated in the latter case, then the linguistic evidence suggests it is implicated in the former as well. This in turn appears to undermine our claim that the "what" system encodes all the details of object shape (here the shape of a face), including their decomposition into parts. How does one reconcile our analysis with this evidence?

We see three possible solutions. One is just to abandon the connection to the "what" and "where" systems and fall back on the Design of Language Hypothesis, as many of the commentators have urged us to do. Even in the face of such a retreat, our evidence would still leave ample implications for the organization of spatial cognition. We are reluctant to give up so easily, however, for reasons already detailed in the present reply.

A second possibility is to take the bull by the horns and claim that the "where" system indeed encodes the relation of object parts to the body of an object. As we understand it, this is a problem being encountered by all who attempt to understand the cortical representation of objects (e.g., Kosslyn 1990). This encoding could be redundant with the encoding of the "what" system, or it might even supplant it. In the latter case, the details of the "what" system would be devoted primarily to a detailed description of the geons making up objects, with the "where" system putting the parts together. That is, the articulation of object shape might involve as elaborate an interweaving of "what" and "where" information as the description of spatial layout does.

A third possibility is that the "what" system operates more or less the way the componential approaches have envisioned, with objects articulated into connected geons. When one specifically wishes to mention or otherwise draw attention to an object part, however, it is necessary to regard it as a separate object with its own location. Under this condition, when the part is regarded as figural, its shape still falls under the "what" system, but its location falls under the "where" system; hence the usual prepositional expressions apply. We do not think evidence from language alone can distinguish among these three possibilities, or any others that might be developed. We leave the resolution to further research in other paradigms.

If nothing else, the linguistic evidence is instrumental in pointing out the distinctions among these hypotheses and demonstrating what is at stake. In this respect we stand firm in our original argument – that language can and should play a role alongside vision in our understanding of the organization of spatial cognition.

ACKNOWLEDGMENTS
Preparation of this commentary was supported by the granting agencies cited in our target article as well as by the University of Pennsylvania Institute for Research in Cognitive Science and Department of Psychology, who jointly sponsored a Visiting Professorship for Barbara Landau during 1992–93.

## References

**Letters *a* and *r* appearing before authors' initials refer to target article and response respectively.**


Anderson, S. & Keenan, E. (1985) Deixis. In: *Language typology and syntactic description*, ed. T. Shopen. Cambridge University Press (Cambridge). [aBL]

Angeli, S. J., Murray, E. A. & Mishkin, M. (1988) The hippocampus and place memory in rhesus monkeys. *Society of Neuroscience Abstracts* 14:232. [aBL]

Au, T. K. & Markman, E. M. (1987) Acquiring word meanings via linguistic contrast. *Cognitive Development* 2:217–36. [aBL]

Bavin, E. (1990) Locative terms and Warlpiri acquisition. *Journal of Child Language* 17:43–66. [aBL]

Beck, B. B. (1980) *Animal tool behavior: The use and manufacture of tools by animals*. Garland STPM Press. [MCC]

Bennett, D. (1975) *Spatial and temporal uses of English prepositions: An essay in stratificational semantics*. Longman Press. [aBL]

Biederman, I. (1987) Recognition-by-components: A theory of human image understanding. *Psychological Review* 94(a):115–47. [aBL, MCC, MJT]

Biederman, I. & Cooper, E. E. (1991) Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology* 23(3):393–419. [MJT]

Binford, O. B. (1971) Visual perception by computer. Presented at IEEE Systems, Science, and Cybernetics Conference, Miami, FL. [aBL]

Bornstein, M. (1985) Color-name versus shape-name learning in young children. *Journal of Child Language* 12:387–93. [aBL]

Bowerman, M. (1989) Learning a semantic system: What role do cognitive predispositions play? In: *The teachability of language*, ed. M. L. Rice & R. C. Schiefenbusch. Brooks. [aBL]

(1991) The origins of children's spatial semantic categories: Cognitive vs. linguistic determinants. In: *Rethinking linguistic relativity*, ed. J. J. Gumperz & S. C. Levinson. Cambridge University Press (Cambridge). [aBL]

Braunstein, M. L., Saidpour, A. & Hoffman, D. (1992) Interpolation in structure from motion. *Perception & Psychophysics* 51(2):105–17. [DDH]

Bridgeman, B. (1981) Cognitive factors in subjective stabilization of the visual world. *Acta Psychologica* 48:111–21. [BB]

(1986) Multiple sources of outflow in processing spatial information. *Acta Psychologica* 63:35–48. [BB]

(1991) Complementary cognitive and motor image processing. In: *Presbyopia research: From molecular biology to visual adaptation*, ed. G. Obrecht & L. Stark. Plenum Press. [BB]

Bridgeman, B., Kirch, M. & Sperling, A. (1981) Segregation of cognitive and motor aspects of visual function using induced motion. *Perception & Psychophysics* 29:336–42. [BB]

Bridgeman, B., Lewis, S., Heit, G. & Nagle, M. (1979) Relationship between cognitive and motor-oriented systems of visual position perception. *Journal of Experimental Psychology: Human Perception and Performance* 5:692–700. [BB]

Bridgeman, B. & Stark, L. (1979) Omnidirectional increase in threshold for image shifts during saccadic eye movements. *Perception & Psychophysics* 25:241–43. [BB]

Brugman, C. (1981) Story of *over*. Indiana University Linguistics Club. [aBL]

Bryant, D. J. & Tversky, B. (1991) Locating objects from memory or from sight. Paper presented at the 32nd annual meeting of the Psychonomic Society, San Francisco, CA. [DJB]

Bryant, D. J., Tversky, B. & Franklin, N. (1992) Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language* 31:74–98. [DJB]

Bulthoff, H. H. & Edelman, S. (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences USA* 89:60–64. [MJT]

Carey, S. (1985) *Conceptual change in childhood*. Bradfood Books/MIT Press. [aBL]

Chwistek, L. (1960) *Wielość rzeczywistości w sztuce*. Warsaw: Czytelnik. [JBD]

Cienki, A. (1988) *Spatial cognition and the semantics of prepositions in English, Polish, and Russian*. Ph.D. dissertation, Brown University. [aBL]

Clark, E. V. & Clark, H. H. (1979) When nouns surface as verbs. *Language* 55:767–811. [BT]

Clark, H. (1973) Space, time, semantics, and the child. In: *Cognitive development and the acquisition of language*, ed. T. Moore. Academic Press. [aBL]

Conley, G. (1985) *Theories of pictorial representation: Goodman's realism and the similarity theory*. Ph.D. dissertation, Department of Philosophy, University of Minnesota. [JBD]

Cooper, L. (1989) Mental models of the structure of visual objects. In: *Object perception: Structure and process*, ed. B. Shepp & S. Ballesteros. Erlbaum. [aBL]

(1992) Probing the nature of the mental representation of visual objects. Seminar on Cognitive Neuroscience, session IV: Memory. American Association for the Advancement of Science. [aBL]

Corballis, M. C. (1991) *The lopsided ape: Evolution of the generative mind*. Oxford University Press. [MCC]

(1992) On the evolution of language and generativity. *Cognition* 44:197–226. [MCC]

Coucelis, H., Golledge, R. G., Gale, N. & Tobler, W. (1987) Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology* 7:99–122. [aBL]

DeLoache, J. S. (1987) Rapid changes in the symbolic functioning of very young children. *Science* 238:1556–57. [aBL]

DeLoache, J. S., Strauss, M. & Maynard, J. (1979) Picture perception in infancy. *Infant Behavior and Development* 2:77–89. [aBL]

Deregowski, J. B. (1978) On examining Fortes' data: Some implications of drawings made by children who have never drawn before. *Perception* 7:479–84. [JBD]

(1980) *Illusions, patterns and pictures.* Academic Press. [JBD]

(1990a) Intercultural search for the origins of perspective. In: *Contemporary issues in cross-cultural psychology,* ed. N. Bleichrodt & P. J. D. Drenth. Swets & Zeitlinger. [JBD]

(1990b) On two distinct and quintessential kinds of pictorial representation. In: *Ecological perception research, visual communication and aesthetics,* ed. K. Landwehr. Springer-Verlag. [JBD]

(1990c) Traffic signs as embodiments of characteristic views. Zeichen (Theorie) in der Praxis. Paper presented at the Biennial Congress of the German Semiotic Association, Passau. [JBD]

Deregowski, J. B., Muldrow, E. S. & Muldrow, W. F. (1972) Pictorial recognition in a remote Ethiopian population. *Perception* 1:417–25. [JBD]

Downs, R. M. & Stea, D. (1973). *Image and environment.* Aldine. [aBL]

Duncan, J. (1989) Boundary conditions on parallel processing in human vision. *Perception* 18:457–69. [JMW]

Duncker, K. (1929) Uber induzierte Bewegung. *Psychologische Forschung* 12:180–259. (Translated and condensed in Ellis, W. [1950] *Source book of Gestalt psychology.* Humanities Press.) [aBL]

Dziurawiec, S. & Deregowski, J. B. (1992) Twisted perspective in young children's drawings. *British Journal of Developmental Psychology* 10:35–49. [JBD]

Easton, R. D. & Bentzen, B. L. (1987) Memory for verbally presented routes: A comparison of strategies used by blind and sighted people. *Journal of Visual Impairment and Blindness* 81:100–105. [DJB]

Engelson, S. P. & McDermott, D. V. (1992) Passive robot map building with exploration scripts. *YALEU/DCS/TR-898,* Department of Computer Science, Yale University. [MJT]

Farah, M. J. (1991) Patterns of co-occurrence among the associative agnosias. *Cognitive Neuropsychology* 8:1–19. [MCC]

Farah, M. J., Hammond, K., Levine, D. & Calvanio, R. (1988) Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology* 20:439–62. [aBL, BB]

Farah, M. J. & McClelland, J. L. (1991) A computational model of semantic memory impairment. *Journal of Experimental Psychology: General* 120:339–57. [MCC]

Feldman, J. (1991) Perceptual simplicity and modes of structural generation. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society,* Chicago, IL, August. [JF]

(1992) Constructing perceptual categories. *Proceedings of the I.E.E.E. Conference on Computer Vision and Pattern Recognition,* Champaign, IL, June. [JF]

Fillmore, C. (1975) Santa Cruz lectures on deixis, 1971. Indiana University Linguistics Club. [arBL]

Fodor, J. A. (1980) Reply to Putnam. In: *Language and learning: The debate between Jean Piaget and Noam Chomsky,* ed. M. Piatelli-Palmerini. Harvard University Press. [rBL]

(1983) *Modularity of mind.* MIT Press. [aBL]

Fortes, M. (1940) Children's drawings among the Tallensi. *Africa* 13:293–95. [JBD]

(1981) Tallensi children's drawings. In: *Universals of human thought,* ed. B. Lloyd & J. Gay. Cambridge University Press (Cambridge). [JBD]

Foster, D. H. & Ward, P. A. (1991) Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society of London B* 243:75–81. [JMW]

Franklin, N. & Tversky, B. (1990) Searching imagined environments. *Journal of Experimental Psychology: General* 119:63–76. [DJB]

Fussell, S. R. & Krauss, R. M. (1989) The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology* 25:203–19. [aBL]

Gallistel, C. R. (1990) *The organization of learning.* MIT Press. [aBL, DJB]

Gibson, E. (1969) *Principles of perceptual learning and development.* Appleton-Century-Crofts. [aBL]

Gibson, J. J. (1966) *The senses considered as perceptual systems.* Houghton Mifflin. [SDM]

(1979) *The ecological approach to visual perception.* Houghton Mifflin. [JW]

...berg, A. M., Meyer, M. & Lindem, K. (1987) Mental models contribute ...foregrounding during text comprehension. *Journal of Memory and Language* 26:69–83. [DJB]

...M. A. & Milner, A. D. (1992) Separate visual pathways for ...ption and action. *Trends in Neuroscience* 15(1):20–25. [DI]

...M. & Khurana, B. (1990) Features were meant to be integrated. ...*tive Ophthalmology and Vision Science* 31:105. [JMW]

Greenfield, P. M. (1991) Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences* 14:531–95. [MCC]

Greenfield, P. M., Reich, L. C. & Olver, R. R. (1987) On culture and equivalence: II. In: *Studies in cognitive growth,* ed. J. S. Bruner, R. R. Olver & P. M. Greenfield. Wiley. [JBD]

Greeno, J. G. (1992) The situation in cognitive theory. Paper presented to the American Psychological Society, San Diego, CA, June. [SDM]

Greeno, J. G. & Moore, J. L. (in press) Situativity and symbols: Response to Vera and Simon. *Cognitive Science.* [SDM]

Greeno, J. G., Smith, D. R. & Moore, J. L. (in press) Transfer of situated learning. In: *Transfer on trial,* ed. D. Detterman & R. Sternberg. Ablex. [SDM]

Grice, H. P. (1975) Logic and conversation. In: *Syntax and semantics,* vol. 3: *Speech acts,* ed. P. Cole & J. L. Morgan. Seminar Press. [PBH]

Gruber, J. (1976) *Lexical structures in syntax and semantics.* North-Holland. [aBL]

Hawkins, B. W. (1984) *The semantics of English spatial prepositions.* Ph.D. dissertation, Department of Linguistics, University of California at San Diego. [aBL]

Heibeck, T. H. & Markman, E. M. (1987) Word learning in children: An examination of fast mapping. *Child Development* 58:1021–34. [aBL]

Hein, A. & Jeannerod, M. (1983) *Spatially oriented behavior.* Springer-Verlag. [aBL]

Held, R. (1968) Dissociation of visual functions by deprivation and rearrangement. *Psychologische Forschung* 31:338–48. [DI]

Herskovits, A. (1986) *Language and spatial cognition: An interdisciplinary study of the prepositions in English.* Cambridge University Press (Cambridge). [arBL]

Hinton, G. E., McClelland, J. L. & Rumelhart, D. E. (1986) Distributed representations. In: *Parallel distributed processing: Explorations in the microstructure of cognition,* vol. 1, ed. D. E. Rumelhart, J. L. McClelland & the PDP Research Group. Bradford Books/MIT Press. [PBH]

Hintzman, D. L., O'Dell, C. S. & Arndt, D. R. (1981) Orientation in cognitive maps. *Cognitive Psychology* 13:149–206. [DJB]

Hirtle, S. & Jonides, J. (1985) Evidence of hierarchies in cognitive maps. *Memory and Cognition* 13:208–17. [aBL]

Hochberg, J. & Brooks, V. (1962) Pictorial recognition as an unlearned ability: A study of one child's performance. *American Journal of Psychology* 75:624–28. [aBL]

Hochberg, J. & Gellman, L. (1975) The effect of landmark features on mental rotation times. *Memory and Cognition* 5:23–26. [aBL]

Hoffman, D. & Richards, W. (1984) Parts of recognition. *Cognition* 18:65–96. [aBL]

Hooper, K. (1978) Perceptual aspects of architecture. In: *Handbook of perception,* vol. 10, ed. E. C. Carterette & M. P. Friedman. Academic Press. [aBL]

Huttenlocher, J. & Strauss, S. (1968) Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior* 7:300–304. [aBL]

Ingle, D. (1967) Two visual mechanisms underlying the behavior of fish. *Psychologische Forschung* 31:44–51. [DI]

Ingle, D., Schneider, G., Trevarthen, C. & Held, R. (1967) Locating and identifying: Two modes of visual processing: A symposium. *Psychologische Forschung* 31:42–43. [aBL, DI]

Jackendoff, R. (1973) The base rules for prepositional phrases. In: *A festschrift for Morris Halle,* ed. S. Anderson & P. Kiparsky. Holt, Rinehart & Winston. [aBL]

(1976) Toward an explanatory semantic representation. *Linguistic Inquiry* 7:89–150. [aBL]

(1983) *Semantics and cognition.* MIT Press. [arBL]

(1987a) On beyond zebra: The relation of linguistic and visual information. *Cognition* 26:89–114. [aBL, SDM]

(1987b) *Consciousness and the computational mind.* MIT Press. [arBL]

(1990) *Semantic structures.* MIT Press. [arBL]

(1992) *Languages of the mind.* Bradford Books/MIT Press. [rBL]

Jackendoff, R. & Landau, B. (1991) Spatial language and spatial cognition. In: *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman,* ed. D. J. Napoli & J. Kegl. Erlbaum. [aBL]

Jay, M. F. & Sparks, D. L. (1987) Sensorimotor integration in the primate superior colliculus, II: Coordinates of auditory signals. *Journal of Neurophysiology* 57:35–55. [MJT]

Jepson, A. & Richards, W. (1992) What makes a good feature? In: *Spatial vision in humans and robots,* ed. L. Harris & M. Jenkins. Cambridge University Press (Cambridge). [JF]

Pinker, S. (1989) *Learnability and cognition: The acquisition of argument structure.* MIT Press. [aBL]

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13(4):707–84. [aBL]

Pinxten, R., van Dooren, I. & Harvey, K. (1983) *The anthropology of space: Explorations into natural philosophy and semantics of the Navajo.* University of Pennsylvania Press. [aBL]

Retz-Schmidt, G. (1988) Various views on spatial prepositions. *AI Magazine* 9:95–105. [DJB]

Rieser, J. J. & Heiman, M. (1982) Spatial self-reference systems and shortest route behavior in toddlers. *Child Development* 53(2):524–33. [aBL]

Rosch, E. (1975) Cognitive reference points. *Cognitive Psychology* 7(4):532–47. [aBL]

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* 8:382–439. [aBL]

Rueckl, J., Cave, K. & Kosslyn, S. (1988) Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience* 1(2):171–86. [aBL, JF]

Sadalla, E., Burroughs, W. J. & Staplin, L. J. (1980) Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory* 6(5):516–28. [aBL]

Schneider, G. E. (1967) Contrasting visual functions of tectum and cortex in the golden hamster. *Psychologische Forschung* 31:52–62. [DI]

(1969) Two visual systems. *Science* 163:895–902. [aBL, BB]

Seibert, M. & Waxman, A. M. (1992) Adaptive 3-D object recognition from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2):107–24. [MJT]

Shepard, R. N. (1984) Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417–47. [SDM]

Shepard, R. & Cooper, L. (1982) *Mental images and their transformations.* MIT Press. [aBL]

Shepard, R. N. & Hurwitz, S. (1984) Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition* 18:161–93. [DJB]

Smith, L., Jones, S. & Landau, B. (1992) Count nouns, adjectives, and perceptual properties in novel word interpretations. *Developmental Psychology* 28(2):273–86. [aBL]

Stephens, A. & Coup, P. (1978) Distortions in judged spatial relations. *Cognitive Psychology* 10:422–37. [aBL]

Stiles-Davis, J., Kritchevsky, M. & Bellugi, U. eds. (1988) *Spatial cognition: Brain bases and development.* Erlbaum. [aBL]

Supalla, T. (1990) *Structure and acquisition of verbs in motion and location in A.S.L.* Bradford Books/MIT Press. [aBL]

Talmy, L. (1975) Semantics and syntax of motion. In: *Syntax and semantics*, vol. 4, ed. J. Kimball. Academic Press. [DIS]

(1978) The relation of grammar to cognition. In: *Proceedings of TINLAP-2: Theoretical issues in natural language processing*, ed. D. Waltz. University of Illinois Press. [aBL]

(1980) Lexicalization patterns: Semantic structure in lexical forms. In: *Language typology and syntactic description*, vol. 3, ed. T. Shopen. Cambridge University Press. [rBL]

(1983) How language structures space. In: *Spatial orientation: Theory, research, and application*, ed. H. Pick & L. Acredolo. Plenum Press. [aBL]

(1985) Lexicalization patterns: Semantic structure in lexical forms. In: *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon*, ed. T. Shopen. Cambridge University Press (Cambridge) [aBL]

Tarr, M. J. (1989) *Orientation dependence in three-dimensional object recognition.* Ph.D. dissertation, Department of Brain and Cognitive Sciences, MIT. [MJT]

Tarr, M. J. & Black, M. J. (1991) A computational and evolutionary perspective on the role of representation in vision. *YALEU/DCS/RR-899*, Department of Computer Science, Yale University. [MJT]

Tarr, M. J. & Pinker, S. (1989) Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology* 21(28):233–82. [aBL, MJT]

(1990) When does human object recognition use a viewer-centered reference frame? *Psychological Science* 1(42):253–56. [MJT]

Taylor, H. A. & Tversky, B. (1992) Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language* 31:261–92. [DJB]

Thompson, D. (1961) *On growth and form.* Cambridge University Press (Cambridge). [aBL]

Treisman, A. (1986) Properties, parts, and objects. In: *Handbook of human perception and performance*, ed. K. R. Boff, L. Kaufmann & J. P. Thomas. Wiley. [JMW]

(1988) Features and objects: The 14th Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology* 40A:201–37. [JMW]

Trevarthen, C. B. (1968) Two mechanisms of vision in primates. *Psychologische Forschung* 31:299–337. [DI]

Tversky, B. & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General* 113:169–93. [BT]

Ullman, S. & Basri, R. (1991) Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(10):992–1006. [MJT]

Ungerleider, L. G. & Mishkin, M. (1982) Two cortical visual systems. In: *Analysis of visual behavior*, ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield. MIT Press. [aBL, BB]

Vandeloise, C. (1986) *L'espace en français.* Editions du Seuil. (English translation: *Spatial prepositions.* University of Chicago Press, 1991.) [arBL]

Van Essen, D., Anderson, C. & Felleman, D. (1992) Information processing in the primate visual system: An integrated systems perspective. *Science* 255:419–23. [aBL]

von Hofsten, C. (1980) Predictive reaching for moving objects by human infants. *Journal of Educational and Child Psychology* 30:369–82. [aBL]

Warrington, E. K. & McCarthy, R. (1987) Categories of knowledge: Further fractionation and an attempted integration. *Brain* 110:1273–96. [MCC]

Warrington, E. K. & Shallice, T. (1984) Category-specific semantic impairments. *Brain* 107:829–54. [MCC]

White, R. (1989) Visual thinking in the ice age. *Scientific American* 261(1):92–99. [MJT]

Williams, P. (1991) Children's and adults' understanding of *across*. Undergraduate honors thesis, Department of Psychology, Columbia University. [aBL]

Wolfe, J. M. (1992) The parallel guidance of visual attention. *Current Directions in Psychological Science* 1(4):125–28. [JMW]

Wolfe, J. M. & Friedman-Hill, S. R. (1992) Part-whole relationships in visual search. *Investigative Ophthalmology and Visual Science* 33:1356. [JMW]

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I. & O'Connell, K. M. (1992) The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance* 18:34–49. [JMW]

Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R. & Cave, K. R. (1990) Limitations on the parallel guidance of visual search: Color × color and orientation × orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance* 16:879–92. [JMW]

Wong, E. & Mack, A. (1981) Saccadic programming and perceived location. *Acta Psychologica* 48:123–31. [BB]

Johansson, G. (1973) Visual perception of biological motion and a model for its analysis. *Perceptual Pschycophysics* 14:201–11. [DI]

Johnson-Laird, P. N. (1983) *Mental models: Towards a cognitive science of language, inference, and consciousness.* Harvard University Press. [DJB]

Johnston, J. (1984) Acquisition of locative meanings: *Behind* and *in front of. Journal of Child Language* 11:407–22. [aBL]

Jolicoeur, P. (1985) The time to name disoriented natural objects. *Memory & Cognition* 13:289–303. [MJT]

Jones, S., Smith, L. & Landau, B. (1991) Object properties and knowledge in early lexical learning. *Child Development* 62:499–516. [aBL]

Keil, F. (1989) *Concepts, kinds, and conceptual development.* MIT Press. [aBL]

Kendall, D. (1989) A survey of the statistical theory of shape. *Statistical Science* 4:87–120. [JF]

Koenderink, J. J. (1987) An internal representation for solid shape based on the topological properties of the apparent contour. In: *Image understanding 1985–86,* ed. W. Richards & S. Ullman. Ablex. [MJT]

Kosslyn, S. M. (1990) Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition* 34:203–77. [arBL]

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J. & Koenig, O. (1992) Categorical versus coordinate spatial representations: Computational analysis and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance* 18:562–77. [HDB]

Kosslyn, S. M. & Koenig, O. (1992) *Wet mind.* The Free Press. [HDB]

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J. & Gabrieli, J. D. E. (1989) Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance* 15:723–35. [HDB, MCC]

Kriegman, D. J. & Ponce, J. (1990) Computing exact aspect graphs of curved objects: Solids of revolution. *International Journal of Computer Vision* 5(21):119–35. [MJT]

Kuipers, B. (1978) Modeling spatial knowledge. *Cognitive Science* 2:129–53. [aBL]

Lakoff, G. (1987) *Women, fire, and dangerous things.* University of Chicago Press. [aBL]

Lakoff, G. & Turner, M. (1980) *Metaphors we live by.* University of Chicago Press. [aBL]

Landau, B. (1986) Early map use as an unlearned ability. *Cognition* 22:201–23. [aBL]

(1991) Spatial knowledge of objects in the young blind child. *Cognition* 38:145–78. [aBL]

(1993) Object, word and name in first language learning. In: *Special issue on the lexicon,* ed. L. R. Gleitman & B. Landau. Lingua (in press). [aBL]

Landau, B. & Gleitman, L. R. (1985) *Language and experience: Evidence from the blind child.* Harvard University Press. [aBL]

Landau, B., Jones, S. & Smith, L. (1992) Syntactic context and object properties in early lexical learning. *Journal of Memory and Language* (in press). [aBL]

Landau, B., Smith, L. & Jones, S. (1988) The importance of shape in early lexical learning. *Cognitive Development* 3:299–321. [aBL]

Landau, B., Spelke, E. & Gleitman, H. (1984) Spatial knowledge in a young blind child. *Cognition* 16:225–60. [aBL]

Landau, B. & Stecker, D. (1990) Objects and places: Syntactic and geometric representations in early lexical learning. *Cognitive Development* 5:287–312. [arBL, ADF]

Landau, B., Stecker, D. & Lederer, A. (forthcoming) Asymmetry in children's and adults' judgements of *near.* [aBL]

Lederman, S. & Klatzky, R. (1987) Hand movements: A window into haptic object recognition. *Cognitive Psychology* 19:342–68. [aBL]

Levelt, W. (1983) Some perceived limitations on talking about space. Unpublished manuscript, Max Planck Institute for Psycholinguistics, Nijmegen. [aBL]

(1984) Some perceptual limitations on talking about space. In: *Limits in perception,* ed. A. J. van Doorn, W. A. van de Grind & J. J. Koenderink. Utrecht: Coronet Books. [DJB]

Levine, D., Warach, J. & Farah, M. (1985) Two visual systems in mental imagery: Dissociation of "what" and "where" in imagery disorders due to bilateral posterior cerebral lesions. *Neurology* 35:1010–18. [aBL]

Levinson, S. C. (1991) *Relativity in spatial conception and description.* Working paper no. 1, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen. [DIS]

(1992) Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. Paper delivered to the Workshop on Spatial Description in Mayan Languages. Nijmegen. [rBL]

(in press) Relativity in spatial conception and description. In: *Rethinking linguistic relativity,* ed. J. J. Gumperz & S. C. Levinson. Cambridge University Press (Cambridge). [BT]

Leyton, M. (1989) Inferring causal history from shape. *Cognitive Science* 13:357–89. [aBL]

(1992) *Symmetry, causality, mind.* MIT Press. [aBL, JF]

Livingstone, M. & Hubel, D. (1989) Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science* 240:740–49. [aBL]

Logan, G. D. (1991) Linguistic and conceptual control of visual spatial attention. Paper presented at the 32nd annual meeting of the Psychonomic Society, San Francisco, CA. [DJB]

Lowe, D. (1985) *Perceptual organization and visual recognition.* Kluwer. [aBL]

Lynch, K. (1960). *The image of the city.* MIT Press. [aBL]

Marr, D. (1982) *Vision.* Freeman. [arBL, MCC, JBD, JF, SDM, JW]

Marr, D. & Nishihara, H. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B* 200:269–94. [aBL, MJT]

Marr, D. & Vaina, L. (1982) Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London* 214:501–24. [aBL]

Marsolek, C. J. (1992) Abstract-visual-form representations in the left cerebral hemisphere. (Submitted.) [HDB]

Marsolek, C. J., Kosslyn, S. M. & Squire, L. R. (1992) Form-specific visual priming in the right cerebral hemisphere. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:492–508. [HDB]

Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81–97 [MCC]

Miller, G. & Johnson-Laird, P. (1976) *Language and Perception.* Harvard University Press. [aBL]

Milner, A. D. & Goodale, M. A. (1992) Visual pathways to perception and action. In: *The visually responsive neuron,* ed. T. P. Hicks, S. Molotchinkoff & T. Ono. Elsevier. [DI]

Moore, G. T. (1976) Theory and research on the development of environmental knowing. In: *Environmental knowing,* ed. G. T. Moore & R. G. Golledge. Dowden, Hutchinson & Ross. [SDM]

Moraglia, G. (1989) Display organization and the detection of horizontal line segments. *Perception & Psychophysics* 45:265–72. [JMW]

Morrow, D. G., Greenspan, S. L. & Bower, G. H. (1987) Accessibility and situation models in narrative comprehension. *Journal of Memory and Language* 26:165–87. [DJB]

Murphy, G. & Medin, D. (1985) The role of theories in conceptual coherence. *Psychological Review* 92:289–316. [aBL]

Neisser, U. (1989) Direct perception and recognition as distinct perceptual systems. Paper presented to the Cognitive Science Society, Ann Arbor, MI, August. [SDM]

(1992) Distinct systems for "where" and "what": Reconciling the ecological and representational views of perception. Paper presented to the American Psychological Society, San Diego, CA, June. [SDM]

Newport, E. (1988) Constraints on language learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences* 10(1):147–61. [aBL]

Oakhill, J. V. & Johnson-Laird, P. N. (1984) Representation of spatial descriptions in working memory. *Current Psychological Research & Reviews* 3:52–62. [DJB]

O'Keefe, J. & Nadel, L. (1978) *the hippocampus as a cognitive map.* Oxford University Press (Oxford). [aBL]

Olson, D. & Bialystok, E. (1983) *Spatial cognition: The structure and development of the mental representation of spatial relations.* Erlbaum. [aBL, DRO]

Oppenheimer, F. (1934) Optische Versuche über Ruhe und Bewegung. *Psychologische Forschung* 20:1–46. [aBL]

Pager, H. (1972) *Ndedema.* International Book Services. [JBD]

Paillard, J. (1987) Cognitive versus sensorimotor encoding of spatial information. In: *Cognitive processes and spatial orientation in animal and man,* ed. P. Ellen & C. Thinus-Blanc. Martinus Nijhoff. [BB]

ed. (1991) *Brain and space.* Oxford University Press. [MJT]

Parker, D. M. & Deregowski, J. B. (1990) *Perception and artistic style.* North Holland. [JBD]

Parsons, L. H. (1987) Imagined spatial transformations of one's hands and feet. *Cognitive Psychology* 19(2):178–241. [aBL]

Pelisson, D., Prablanc, C., Goodale, M. & Jeannerod, M. (1986) Visual control of reaching movements without vision of the limb. II. Evidence for fast nonconscious process correcting the trajectory of the hand to the final position of a double-step stimulus. *Experimental Brain Research* 62:303–11. [BB]