Chapter 13

A Computational Analysis of the Apprehension of Spatial Relations

Gordon D. Logan and Daniel D. Sadler

13.1 Introduction

Spatial relations are important in many areas of cognitive science and cognitive neuroscience, including linguistics, philosophy, anthropology, and psychology. Each area has contributed substantially to our understanding of spatial relations over the last couple of decades, as is evident in the other chapters in this volume. The psychologists' contribution is a concern for *how* spatial relations are apprehended, a concern for the interaction of representations and processes underlying an individual's apprehension of spatial relations. This chapter presents a computational analysis of the representations and processes involved in apprehending spatial relations and interprets this analysis as a psychological theory of apprehension. The chapter begins with a theory and ends with data that test the assumptions of the theory and with some comments about generality.

13.2 Three Classes of Spatial Relations

A computational theory accounts for a phenomenon in terms of the representations and processes that underlie it, specifying how the processes operate on the representations to produce the observed behavior. Important clues to the nature of the representations and processes involved in the apprehension of spatial relations can be found in the linguistic and psycholinguistic literature that addresses the semantics of spatial relations (e.g., Clark 1973; Garnham 1989; Herskovits 1986; Jackendoff and Landau 1991; Levelt 1984; Miller and Johnson-Laird 1976; Talmy 1983; and Vandaloise 1991). That literature distinguishes between three classes of spatial relations, and the discriminanda that distinguish the classes suggest the requisite representations and processes.

13.2.1 Basic Relations

Garnham (1989) distinguished basic relations from deictic and intrinsic ones. Basic relations take one argument, expressing the position of one object with respect to the viewer (e.g., the viewer thinks, "This is here" and "That is there").[1] Basic relations are essentially the same as *spatial indices*, which are discussed in the literature on human and computer vision (e.g., Pylyshyn 1984, 1989; Ullman 1984). Spatial indices establish correspondence between perceptual objects and symbols, providing the viewer's cognitive system with a way to access perceptual information about an object. Spatial indices—basic relations—individuate objects without necessarily identifying, recognizing, or categorizing them. The conceptual part of a basic relation is a symbol or a token that stands for a perceptual object. It simply says, "Something is there," without saying what the "something" is. The token may be associated with an identity or a categorization, pending the results of further processing, but it need not be identified, recognized, or categorized in order to be associated with a perceptual object. The perceptual part of a basic relation is an object that occupies a specific point or region in perceptual space.

Basic relations represent space in that they associate a conceptual token with the object in a location in perceptual space. Conceptually, the representation of space is very crude—an object is "here" and "not there." Thus two objects that are indexed separately can either be in the same location or in different locations. If they are in different locations, their relative positions are not represented explicitly in the conceptual representation. Information about their relative locations may be available implicitly in perceptual space, but it is not made explicit in basic relations. Other relations and other computational machinery are necessary to make relative position explicit.

13.2.2 Deictic Relations

Although Garnham (1989) was the first to distinguish basic relations, most linguists and psycholinguists distinguish between deictic and intrinsic relations (e.g., Herskovits 1986; Jackendoff and Landau 1991; Levelt 1984; Miller and Johnson-Laird 1976; Talmy 1983; and Vandaloise 1991). Deictic relations take two or more objects as arguments, specifying the position of one object, the *located object*, in terms of the other(s), the *reference object(s)*. The position is specified with respect to the *reference frame* of the viewer, which is projected onto the reference object. Deictic relations specify the position of the located object with respect to the viewer if the viewer were to move to the position of the reference object. Thus "The ball is left of the tree" means that if the viewer were to walk to the tree, the ball would be on his or her left side.

Deictic relations are more complex computationally than basic relations because they relate objects to each other and not simply to the viewer. They represent the relative positions of objects explicitly. The arguments of deictic relations must be individuated but they need not be identified, recognized, or categorized. Individuation is necessary because the reference object is conceptually different from the located object (i.e., "*X* is above *Y*" and "*Y* is above *X*" mean different things), but the distinction between reference and located objects can be made by simply establishing tokens that represent perceptual objects, leaving identification, recognition, and categorization to subsequent processes.

13.2.3 Intrinsic Relations

Like deictic relations, intrinsic relations take two or more arguments and specify the position of a located object with respect to a reference object. They differ from deictic relations in that the position is specified with respect to a reference frame intrinsic to the reference object rather than the viewer's reference frame projected onto the reference object. Whereas deictic relations can apply to any reference object, intrinsic relations require reference objects that have intrinsic reference frames, that is, intrinsic tops and bottoms, fronts and backs, and left and right sides. Objects like people, houses, and cars can serve as reference objects for intrinsic relations because they have fronts, backs, tops, bottoms, and left and right sides. Objects like balls cannot serve as reference objects for intrinsic relations because they have no intrinsic tops, bottoms, and so on. Objects like trees have tops and bottoms but no fronts and backs or left and right sides, so they can support intrinsic *above* and *below* relations but not intrinsic *in front of* or *left of* relations; *in front of* and *left of* would have to be specified deictically. Objects like bullets and arrows have intrinsic fronts and backs but no intrinsic tops and bottoms or left and right sides. They can support intrinsic *in front of* and *behind* relations, but *above* and *left of* would have to be specified deictically.

Intrinsic relations are more complex computationally than deictic relations because they require the viewer to extract the reference frame from the reference object. An obvious way to extract the reference frame is to recognize the reference object or classify it as a member of some category and to impose the reference frame appropriate to that category. For example, seeing an ambiguous figure as a duck or a rabbit leads the viewer to assign *front* to different regions of the object (Peterson et al. 1992). However, it may be possible in some cases to assign an intrinsic reference frame without actually identifying the object. The main axis of the reference frame may be aligned with the object's axis of elongation (Biederman 1987; Marr and Nishihara 1978) or with the object's axis of symmetry (Biederman 1987; Palmer 1989).

13.2.4 Implications for Computation

The distinction between the three classes of spatial relations has at least two implications for a theory of the computation involved in apprehension. First, each class of relations describes the position of the located object in terms of a reference frame. The reference frame may coincide with the viewer's, as in basic relations, it may be projected onto the reference object, as in deictic relations, or it may be extracted from the asymmetries inherent in the reference object, as in intrinsic relations. In each case, the reference frame is a central part of the meaning of the spatial relation, and this suggests that reference frame computation is a central part of the process of apprehension.

Second, the distinction between reference objects and located objects suggests that the arguments of two- or three-place relations must be individuated somehow. "*X* is above *Y*" does not mean the same as "*Y* is above *X*." The process of spatial indexing —instantiating basic relations—is well suited for this purpose. Each object can be represented by a different token, and the tokens can be associated with the arguments that correspond to the located and reference object in the conceptual representation of the relation. The distinction between located and reference objects is also important in reference frame computation because the reference frame is projected onto or extracted from the reference object, not the located object. Spatial indexing is useful here as well. It is a central part of apprehension.

13.3 Spatial Templates as Regions of Acceptability

Reference frames and the distinction between located and reference objects suggest important parts of a computational theory of apprehension, but something is missing. They do not specify how one would decide whether a given spatial relation applied to a pair or triplet of objects. This issue has been discussed extensively in the linguistic and psycholinguistic literature. Various researchers have suggested computations involving geometric (Clark 1973; Miller and Johnson-Laird 1976), volumetric (Herskovits 1986; Talmy 1983), topological (Miller and Johnson-Laird 1976; Talmy 1983), and functional (Herskovits 1986; Vandaloise 1991) relations. We propose that people decide whether a relation applies by fitting a *spatial template* to the objects that represents *regions of acceptability* for the relation in question (see also Carlson-Radvansky and Irwin 1993; Hayward and Tarr 1995; Kosslyn et al. 1992; Logan 1994, 1995; Logan and Compton 1996).

A spatial template is a representation that is centered on the reference object and aligned with the reference frame imposed on or extracted from the reference object. It is a two- or three-dimensional field representing the degree to which objects appearing in each point in space are acceptable examples of the relation in question. The main idea is that pairs or triplets of objects vary in the degree to which they instantiate spatial relations. Roughly speaking, there are three main regions of acceptability: one reflecting good examples, one reflecting examples that are less than good but nevertheless acceptable, and one reflecting unacceptable examples. Good and acceptable regions are not distinct with a sharp border between them. Instead, they blend into one another gradually. With the relation *above*, for example, any object that is aligned with the upward projection of the up-down axis of the reference object is a good example. Any object above a horizontal plane aligned with the top of the reference object is an acceptable example, although not a good one (the closer it is to the upward projection of the up-down axis, the better). And any object below a horizontal plane aligned with the bottom of the reference object is a bad, unacceptable example.

We propose that people use spatial templates to determine whether a spatial relation applies to a pair of objects. If the located object falls in a good or an acceptable region when the template is centered on the reference object, then the relation can apply to the pair. If two relations can apply to the same pair of objects, the preferred relation is the one whose spatial template fits best. If both spatial relations fit reasonably, the viewer may assert both relations (e.g., "above and to the right"). Spatial templates provide information about goodness of fit. Exactly how information about goodness of fit is used depends on the viewer's goals and the viewer's task (see below).

13.4 Computational Theory of Apprehension

At this point the representations and processes necessary to apprehend spatial relations have been described in various ways, some in detail, some briefly, and some only implicitly. Now it is time to describe them explicitly and say how they work together.

13.4.1 Representations

The theory assumes that the apprehension of spatial relations depends on four different kinds of representations: a perceptual representation consisting of objects and surfaces, a conceptual representation consisting of spatial predicates, a reference frame, and a spatial template. It may be more accurate to say there are two kinds of representation, one perceptual and one conceptual, and two "intermediate" representations that map perception onto cognition and vice versa.

13.4.1.1 Perceptual Representation The perceptual representation is a two-, two-and-a-half-, or three-dimensional analog array of objects and surfaces. It is formed automatically by local parallel processes as an obligatory consequence of opening one's eyes (see, for example, Marr 1982; Pylyshyn 1984; and Ullman 1984). The representation contains information about the identities of the objects and the spatial relations between them, but that information is only implicit. Further computation is necessary to make it explicit. In other words, the representation contains the perceptual information required to identify the objects or to compute spatial relations between them, but that information does not result in an explicit identification of the object as an instance of a particular category or specific relation without further computation. That "further computation" is what the other representations and processes are required for.

The current version of the theory assumes that the perceptual representation is relatively low-level, and that need not be the case. We make that assumption because it is relatively clear how low-level representations can be constructed from light impinging on the retina (e.g., Biederman 1987; Marr 1982), and we want the theory to be tractable computationally. However, the spirit of the theory would not be very different if we assumed that the perceptual representation was much more abstract; for example, if we assumed that spatial information was represented amodally, combining visual, auditory, tactual, and imaginal information. The key idea is that the perceptual representation provides an analog array of objects that can be compared to a spatial template. In principle, the objects can be highly interpreted and abstracted from the sensory systems that gave rise to them.

13.4.1.2 Conceptual Representation The conceptual representation is a one-, two-, or three-place predicate that expresses a spatial relation. The conceptual representation identifies the relation (e.g., it distinguishes *above* from *below*); it individuates the arguments of the relation, distinguishing between the reference object and the located object; it identifies the relevant reference frame (depending on the nature of the reference object); and it identifies the relevant spatial template. The conceptual representation does not identify objects and relations directly in the perceptual representation; further processing and other representations are needed for that.

An important feature of the conceptual representation is that it is addressable by language. The mapping of conceptual representations onto language may be direct in some cases and indirect in others. In English, French, Dutch, and German, for example, many conceptual (spatial) relations are lexicalized as spatial prepositions; single words represent single relations. However, there is polysemy even in the class of spatial prepositions. Lakoff (1987), for example, distinguished several different senses of *over*. Moreover, some languages may use a single word to refer to different relations that are distinguished lexically in other languages. For example, English uses one word for three senses of *on* that are distinguished in Dutch (i.e., *om*, *op*, and *aan*; see Bowerman, chapter 10, this volume). Despite these complexities, we assume that

conceptual representations may be mapped onto language and vice versa. The mapping may not always be simple, but it is possible in principle (see also Jackendoff and Landau 1991; Landau and Jackendoff 1993).

13.4.1.3 Reference Frame The reference frame is a three-dimensional coordinate system that defines an origin, orientation, direction, and scale. It serves as a map between the conceptual representation and the perceptual representation, establishing correspondence between them. The distinction between reference and located objects gives a direction to the conceptual representation; the viewer's attention should move *from* the reference object *to* the located object (Logan 1995). The reference frame gives direction to perceptual space, defining *up*, *down*, *right*, *front*, and *back*. It orients the viewer in perceptual space.

We assume that reference frames are flexible representations. The different parameters can be set at will, depending on the viewer's intentions and the nature of the objects on which the reference frame is imposed. Many investigators distinguish different kinds of reference frames—viewer-based, object-based, environment-based, deictic, and intrinsic (Carlson-Radvansky and Irwin 1993, 1994; Levelt 1984; Marr 1982; Marr and Nishihara 1978). We assume that the same representation underlies all of these different reference frames (i.e., a three-dimensional, four-parameter coordinate system). The differences between them lie in the parameter settings. Viewer-based and object-based reference frames (also known as "deictic" and "intrinsic" reference frames) differ in origin (the viewer vs. the object), orientation (major axis of viewer vs. major axis of object), direction (viewer's head up vs. object's "head" up), and scale (viewer's vs. object's).

13.4.1.4 Spatial Template As we just said, the spatial template is a representation of the regions of acceptability associated with a given relation. When the spatial template is centered on the reference object and aligned with its reference frame, it specifies the goodness with which located objects in different positions exemplify the associated relation.

We assume that different relations have different spatial templates associated with them and that similar relations have similar templates. More specifically, we assume that spatial templates are associated with conceptual representations of spatial relations. Consequently, they are addressable by language, but the addressing is mediated by linguistic access to the conceptual representation. We assume there are spatial templates for lexicalized conceptual representations, but in cases of polysemy where there is more than one conceptual representation associated with a given word (e.g., *over*; Lakoff 1987), there is a different spatial template for each conceptual representation. Moreover, we assume that spatial templates can be combined to represent compound relations (e.g., "above right") and decomposed to represent finer distinctions (e.g., "directly above").

13.4.2 Processes

The theory assumes that the apprehension of spatial relations depends on four different kinds of processes: spatial indexing, reference frame adjustment, spatial template alignment, and computing goodness of fit. The first two establish correspondence between perceptual and conceptual representations; the last two establish the relevance or the validity of the relation in question.

13.4.2.1 Spatial Indexing Spatial indexing is required to bind the arguments of the relation in the conceptual representation to objects in the perceptual representation. Spatial indexing amounts to establishing correspondence between a symbol and a percept. A perceptual object is "marked" in the perceptual representation (Ullman 1984), and a symbol or a token corresponding to it is set up in the conceptual representation (Pylyshyn 1984, 1989). The correspondence between them allows conceptual processes to access the perceptual representation of the object so that perceptual information about other aspects of the object can be evaluated (e.g., its identity). Essentially, the viewer asserts two or three basic relations, one for the located object and one or two for the reference objects.

13.4.2.2 Reference Frame Adjustment The relevant reference frame must be imposed on or extracted from the reference object. The processes involved translate the origin of the reference frame, rotate its axes to the relevant orientation, choose a direction, and choose a scale. Not all of these adjustments are required for every relation. *Near* requires setting the origin and the scale, whereas *above* requires setting origin, orientation, and direction.

Different processes may be involved in setting the different parameters. The origin may be set by spatial indexing (Ullman 1984) or by a process analogous to mental curve tracing (Jolicoeur, Ullman, and MacKay 1986, 1991). Orientation may be set by a process analogous to mental rotation (Cooper and Shepard 1973; Corballis 1988). Different reference frames or different parameter settings may compete with each other, and the adjustment process must resolve the competition (Carlson-Radvansky and Irwin 1994).

13.4.2.3 Spatial Template Alignment The spatial template must be imposed on the reference object and aligned with the reference frame. In deictic relations, the spatial template is aligned with the viewer's reference frame projected onto the reference object. In intrinsic relations, it is aligned with the intrinsic reference frame extracted from the object.

13.4.2.4 Computing Goodness of Fit Once the relevant spatial template is aligned with the reference object, goodness of fit can be computed. The position occupied by the located object is compared with the template to determine whether it falls in a good, acceptable, or bad region. We assume that the comparison is done in parallel over the whole visual (or imaginal) field. Spatial templates can be represented computationally as a matrix of weights, and the activation value of each object in the visual-imaginal field can be multiplied by the weights in its region to assess goodness of fit. Weights in the good region can be set to 1.0; weights in the bad region can be set to 0.0, and weights in acceptable but not good regions can be set to values between 0.0 and 1.0. With these assumptions, the better the example, the less the activation changes when the spatial template is applied. The activation of good examples will not change at all; the activation of bad examples will vanish (to 0.0); and the activation of acceptable examples will be somewhat diminished.

Alternatively, weights for bad regions could be set to 1.0, weights for acceptable regions could be greater than 1.0, and weights for the good region could be well above 1.0. With these assumptions, the better the example, the greater the change in activation when the spatial template is applied. The activation of bad examples will not change; the weights of acceptable but not good examples will change a little; and the weights of good examples will change substantially. In either case, the acceptability of candidate objects can be assessed and rank-ordered. Other processes and other considerations can choose among the candidates.

13.4.3 Programs and Routines

Spatial relations are apprehended for different reasons in different contexts. Sometimes apprehension itself is the main purpose, as when we want to determine which horse is ahead of which at the finish line. Other times, apprehension is subordinate to other goals, as when we want to look behind the horse that finished first to see who finished second. A computational analysis of apprehension should account for this flexibility. To this end, we interpret the representations and processes described above as elements that can be arranged in different ways and executed in different orders to fulfill different purposes, like the data structures and the instruction set in a programming language. Ordered combinations of representations and processes are interpreted as *programs* or *routines* (cf. Ullman 1984). In this section, we consider three routines that serve different purposes.

13.4.3.1 Relation Judgments Apprehension is the main purpose of relation judgments. A viewer who is asked, "Where is Gordon?" or "Where is Gordon with respect to Jane?" is expected to report the relation between Gordon and a reference object. In the first case, the reference object is not given. The viewer must (1) find the located object (Gordon); (2) find a suitable reference object (i.e., one the questioner knows about or can find easily); (3) impose a reference frame on the reference object; (4) choose a relation whose region of acceptability best represents the position of the located object; and (5) produce an answer (e.g., "Gordon is in front of the statue"). In the second case, the reference object is given (i.e., Jane). The viewer must (1) find the reference object; (2) impose a reference frame on it; (3) find the located object (i.e., Gordon); (4) choose a relation whose region of acceptability best represents the position of the located object; and (5) produce an answer (e.g., "on her left side.").

We assume that viewers find located objects by spatially indexing objects in the perceptual representation and comparing them to a description of the specified located object (e.g., "Does that look like Gordon?"). When reference objects are specified in advance, we assume they are found in the same manner. If they are not specified in advance, as in the first case, then the most prominent objects are considered as reasonable candidates for reference objects (Clark and Chase 1974; Talmy 1983). The relation itself is chosen by iterating through a set of candidate relations—imposing the associated spatial templates on the reference object, aligning them with the reference frame, and computing goodness of fit—until one with the best fit or one with an acceptable fit is found.

Relation judgments have been studied often in the psychological literature. Subjects are told in advance what the arguments of the relation will be, but they are not told the relation between them. Their task is to find the arguments, figure out the relation between them, and report it. Thus Logan and Zbrodoff (1979) had subjects report whether a word appeared above or below the fixation point; Logan (1980) had subjects decide whether an asterisk appeared above or below a word. A common focus in relation judgments is Stroop-like interference from irrelevant spatial information (e.g., the identity of the word in the first case; the position occupied by the word-asterisk pair in the second).

13.4.3.2 Cuing Tasks In cuing tasks, apprehension is used in the service of another goal. A viewer who is asked, "Who is beside Mary?" must (1) find the reference object (i.e., Mary); (2) impose reference frame on it; (3) align the relevant spatial template with the reference frame (i.e., the one for *beside*); (4) choose as the located object the perceptual object that is the best example (or the first acceptable example) of the relation; and (5) produce an answer (e.g., "Paul").

Cuing tasks have been studied extensively in the psychological literature. Experiments on visual spatial attention require subjects to report a target that stands in a prespecified relation to a cue (e.g., Eriksen and St. James 1986). The cue is the reference object and the target is the located object. Usually, the focus is on factors other than the apprehension of spatial relations; nevertheless, apprehension is a major computational requirement in these tasks (see, for example, Logan 1995).

13.4.3.3 Verification Tasks Verification tasks present the viewer with a completely specified relation (e.g., "Is Daisy sitting next to Stella?") and ask whether it applies to a given scene or a given display. The focus may be on one or the other of the arguments, as in "Is that *Daisy* sitting next to Stella?"; or it may be on the relation itself, as in "Is Daisy sitting *next to* Stella?" If the focus is on the arguments, verification could be done as a cuing task. The viewer could (1) find the reference object (e.g., Stella); (2) impose a reference frame on it; (3) align the relevant spatial template with the reference frame (the one for *next to*); (3) choose a located object that occupies a good or acceptable region; (4) compare that object with the one specified in the question (i.e., Is it Daisy?); and (5) report "yes" if it matches or "no" if it does not. Alternatively, if the focus is on the relation, verification could be done as a judgment task. The viewer could (1) find the located object (Daisy); (2) find the reference object (Stella); (3) impose a reference frame on it; (4) iterate through spatial templates until the best fit is found or until an acceptable fit is found; (5) compare the relation associated with that template with the one asserted in the question; and (6) report "yes" if it matches and "no" if it does not.

Verification tasks are common in the psychological literature. A host of experiments in the 1970s studied comparisons between sentences and pictures, and spatial relations figured largely in that work (e.g., Clark, Carpenter, and Just 1973). Subjects were given sentences that described spatial layouts and then pictures that depicted them. The task was to decide whether the sentence described the picture.

13.5 Evidence for the Theory

The theory claims that apprehension of spatial relations requires spatial indexing, reference frame computation, and assessment of goodness of fit of spatial templates. The psychological literature contains evidence for the first two claims, and that evidence will be reviewed briefly below. The third claim has not yet been tested. The remainder of this chapter presents four pieces of evidence that bear on its validity.

13.5.1 Apprehension Requires Spatial Indexing

Logan (1994) found evidence that apprehension of spatial relations requires spatial indexing in visual search tasks. On each trial, subjects were presented with a sentence that described the relation between a dash and a plus (e.g., "dash right of plus"), followed by a display of dash-plus pairs. Half of the time, one of the pairs matched the description in the sentence (e.g., one dash was *right of* one plus), and half of the time, no pair matched the description. All pairs except the target were arranged in the opposite spatial relation (e.g., all the other dashes were *left of* the corresponding pluses). The experiments examined the relations *above*, *below*, *left of*, and *right of*.

In one experiment, the number of dash-plus pairs was varied, and reaction time increased linearly with the number of pairs. The slope was very steep (85 ms/item when the target was present; 118 ms/item when it was absent), which suggests that the pairs were examined one at a time until a target was found (i.e., the pairs were spatially indexed element by element until a target was found). A subsequent experiment replicated these results over twelve sessions of practice (6,144 trials), suggesting that subjects could not learn to compute spatial relations without spatial indexing.

In a third experiment, the number of pairs was fixed and attention was directed to one pair in the display by coloring it differently from the rest. When the differently colored pair was the target, performance was facilitated; subjects were faster and more accurate. When the differently colored pair was not the target, performance was impaired; subjects were slower and less accurate. This suggests that apprehension of spatial relations requires the kind of attentional process that is directed by cues like discrepant colors (i.e., spatial indexing).

13.5.2 Apprehension Requires Reference Frame Computation

Logan (1995) found evidence that apprehension of spatial relations requires reference frame computation in experiments in which attention was directed from a cue to a target. The relation between the cue and the target was varied within and between experiments. Overall, six relations were investigated: *above*, *below*, *front*, *back*, *left of*, and *right of*. The operation of a reference frame was inferred from differences in reaction time with different relations: *above* and *below* were faster than *front* and *back*, and *front* and *back* were faster than *left of* and *right of*. Clark (1973) predicted these differences from an analysis of the environmental support for each relation, and Tversky and colleagues confirmed Clark's predictions in tasks that required searching imagined environments (Bryant, Tversky, and Franklin 1992; Franklin and Tversky 1990). According to Clark's (1973) analysis, *above* and *below* are easy because they are consistent with gravity, consistent over translations and rotations produced by locomotion, and supported by bodily asymmetries (heads are different from feet). *Front* and *back* are harder because they are supported by bodily asymmetries but not by gravity and they change with locomotion through the environment. *Left* and *right* are hardest of all because they are not supported by gravity or bodily asymmetries and they change with locomotion; they are often defined with reference to other axes. Our theory would account for these differences in terms of the difficulty of aligning reference frames and computing direction.

In Logan's (1995) experiments, subjects reported targets that were defined by their spatial relation to a cue. Some experiments studied deictic relations, using an asterisk as a cue and asking subjects to project their own reference frames onto the asterisk. Subjects saw a display describing a spatial relation (*above*, *below*, *left*, or *right*) and then a picture containing several objects surrounding an asterisk cue. Their task was to report the object that stood in the relation to the asterisk cue that we specified in the first display. Subjects were faster to access objects *above* and *below* the cue than to access objects *right* and *left* of it, consistent with Clark's (1973) hypothesis and with our assumption that orienting reference frames and deciding direction take time.

Other experiments studied intrinsic relations, using a picture of a human head as a cue and asking subjects to extract the intrinsic axes of the head. Again, the first display contained a relation (*above*, *below*, *front*, *back*, *left*, or *right*) and the second contained a display in which objects surrounded a picture of the head. Subjects were faster with *above* and *below* than with *front* and *back*, and faster with *front* and *back* than with *left* and *right*.

In some experiments, the same object could be accessed via different relations. Access to the object was easy when the relation was *above* or *below* and hard when it was *left* or *right*. The cue was presented in different positions, and the regions that were easy and hard to access moved around the display with the cue. This suggests that the reference frame can be translated across space.

In other experiments, the orientation of the reference frame was varied. With deictic cues, subjects were told to imagine that the left side, the right side, or the bottom of the display was the top, and the advantage of *above* and *below* over the other relations rotated with the imagined top. With intrinsic cues, the orientation of the head cue was varied, and the advantage of *above* and *below* over the other relations rotated with the orientation of the head. These data suggest that the reference frame can be rotated at will.

13.6 Evidence for Spatial Templates

The theory assumes that spatial relations are apprehended by computing the goodness of fit between the position of the located object and a spatial template representing the relation that is centered on and aligned with the reference object. The idea that spatial templates are involved in apprehension is new and there is not much evidence for it (but see Hayward and Tarr 1995). Sections 13.7–13.10 present four experiments that test different aspects of the idea. The first experiment assesses the parts of space that correspond to the regions of greatest acceptability, using a production task. The second assesses parts of space corresponding to good, acceptable, and bad regions, using a task in which subjects rate how well sentences describe pictures. The third assesses the importance of spatial templates in thinking about spatial relations, using a task in which subjects rate the similarities of words that describe (lexicalized) spatial relations and comparing the multidimensional similarity space underlying those ratings with one constructed from the ratings of pictures in the second experiment. The final experiment tests the idea that spatial templates are applied in parallel, using a reaction time task in which subjects verify spatial relations between objects.

13.7 Experiment 1: Production Task

The first experiment attempted to capture the regions of space corresponding to the best examples of twelve spatial relations: *above*, *below*, *left of*, *right of*, *over*, *under*, *next to*, *away from*, *near to*, *far from*, *on*, and *in*. Subjects were presented with twelve frames, with a box drawn in the center of each one; above each frame was an instruction to draw an *X* in one of the twelve relations to the box (e.g., "Draw an *X* above the box"). We assumed they would draw each *X* in the region corresponding to the best example of each relation, though we did not require them to. There were 68 subjects, who were volunteers from an introductory psychology class. The frames were drawn on three sheets of paper, four frames per sheet, and three different orders of sheets were presented.[2] Each frame was 5.9 cm square and the central box was 8.5 mm square.

The data were collated by making transparencies of each of the twelve frames. For each relation, we superimposed the transparency on each subject's drawing and drew a dot on the transparency (with a felt pen) at the point corresponding to the center of the *X* that the subject drew, accumulating dots across subjects. The data for *above*, *below*, *over*, *under*, *left of*, and *right of* are presented in figure 13.1, the data for *next to*, *away from*, *near*, *far from*, *in*, and *on* are presented in figure 13.2.

The relations in figure 13.1 differ primarily in the orientation and direction of the reference frame. The patterns in each panel are similar to each other, except for rotation. The main exception is *over*, where some subjects drew *X*s that were superimposed on the box, apparently interpreting *over* as covering (which is a legitimate interpretation; see Lakoff 1987). Note that distance did not matter much. Some *X*s were placed close to the box but others were placed quite far away, near the edge of the frame. In each case, the *X*s appeared roughly centered on the axis of the reference frame extended outward from the box.

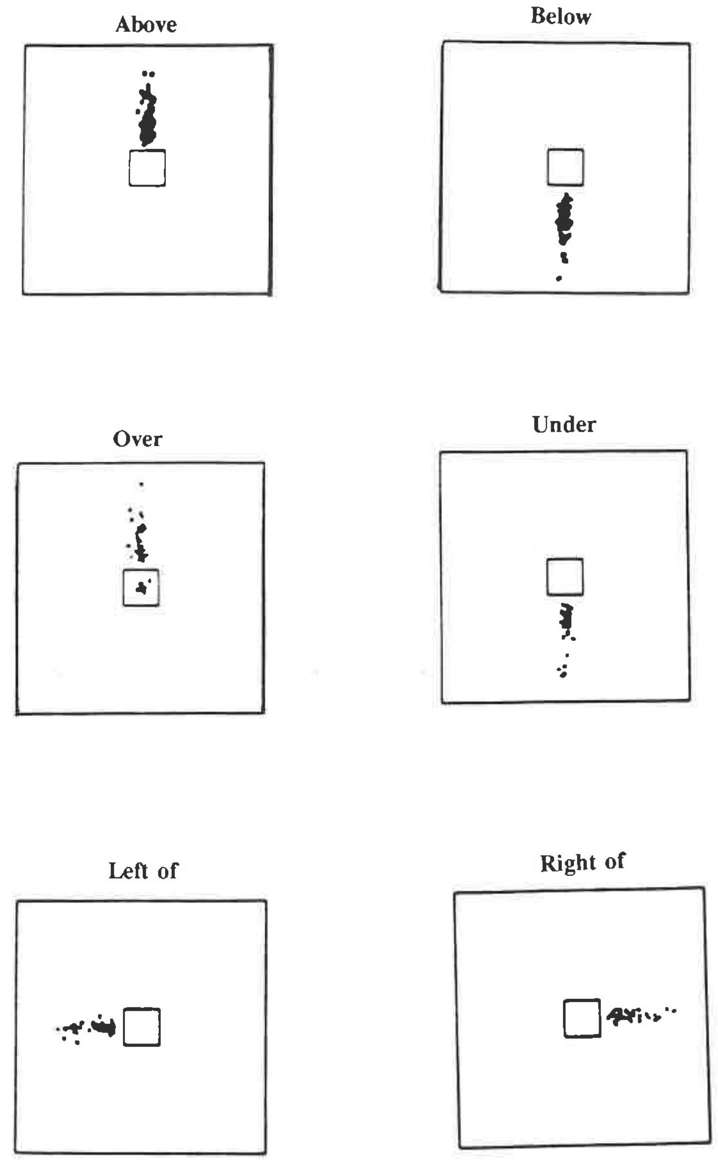

Figure 13.1
Data for *above*, *below*, *over*, *under*, *left of*, and *right of* from the production task in experiment 1. Each point represents the center of an *X* drawn by a different subject to stand in the relation to the central box that is specified above each frame.

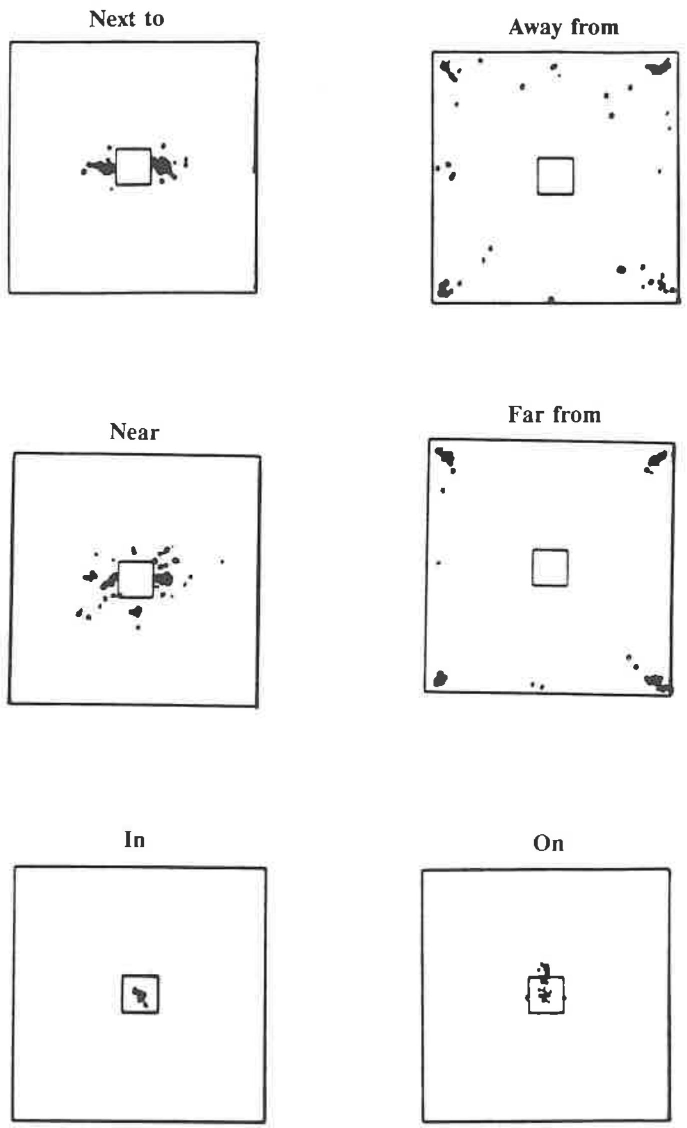

Figure 13.2
Data for *next to*, *away from*, *near*, *far from*, *in*, and *on* from the production task in experiment 1. Each point represents the center of an X drawn by a different subject to stand in the relation to the central box that is specified above each frame.

The relations in the top four panels of figure 13.2 depend primarily on the scale of the reference frame and not on orientation or direction. Xs exemplifying *next to* and *near* were placed close to the box, whereas Xs exemplifying *away from* and *far from* were placed some distance from it, close to the corners (especially *for far form*). One unexpected result was that *next to* was interpreted as horizontal proximity. No subject drew an X above or below the box for *next to*, though many did so for *near*. This unanticipated result appears again in the next experiment.

The bottom two panels of figure 13.2 represent *in* and *on*. All subjects drew their Xs so that their centers were within the boundaries of the box for *in*, but not all subjects did so for *on*. Some drew the X as if it were on top of the box, and one drew the X centered on each side of the box. All of these are legitimate interpretations of the relations.

13.8 Experiment 2: Goodness Rating Task

The second experiment attempted to capture the regions corresponding to good, acceptable, and bad examples of ten of the relations used in experiment 1: *above*, *below*, *left of*, *right of*, *over*, *under*, *next to*, *away from*, *near to*, and *far from*. Subjects were shown sentences, followed by pictures on computer monitors, and were asked to rate how well the sentence described the picture on a scale from 1 (bad) to 9 (good). Each sentence was of the form "The X is [relation] the O" and each picture contained an O in the center of a 7×7 grid and an X in one of the 48 surrounding positions. The grid, which was not visible to the subjects, was 8.8 cm wide and 9.3 cm high on the computer screen. Viewed at a distance of 60 cm, this corresponded to 8.3 degrees $\times$ 8.8 degrees of visual angle. Each of the 48 positions was tested for each relation so that we could get ratings from good, acceptable, and bad regions. There were 480 trials altogether (48 positions $\times$ 10 relations). Subjects reported their rating by pressing one of the numeric keys in the row above the standard QWERTY keyboard. There were thirty-two subjects, volunteers from an introductory psychology class. The data were collated by averaging ratings across subjects. The average ratings are plotted in figures 13.3 and 13.4 and presented in table 13.1. Subjects were very consistent; the mean standard error of the averages in figures 13.3 and 13.4 is 0.271.

Figure 13.3 presents the average ratings for *above*, *below*, *over*, *under*, *left of*, and *right of* drawn as three-dimensional graphs. Screen positions are represented in the up-down axis and the left-right axis. The up-down axis goes from upper left to lower right; the left-right axis goes from lower left to upper right. Ratings are represented in the third dimension, which is essentially vertical on the page. The central position, which was occupied by the O, is blank.

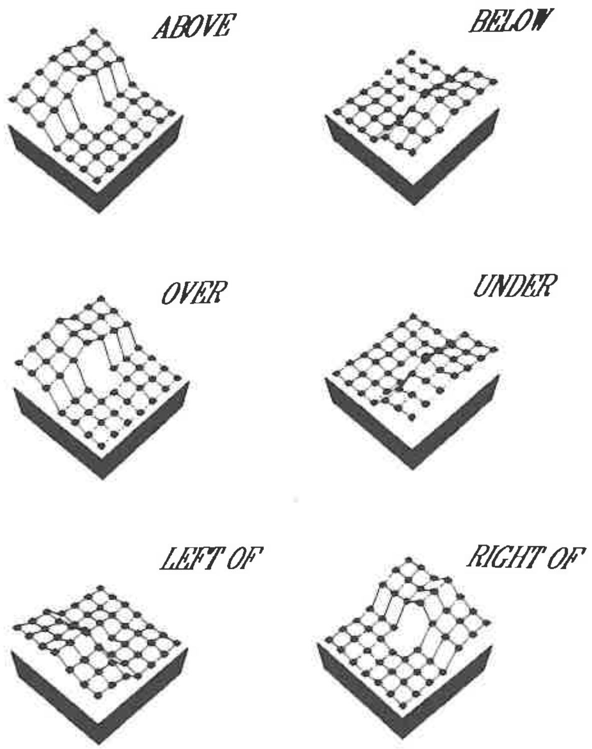

Figure 13.3
Average ratings for *above*, *below*, *over*, *under*, *left of*, and *right of* from the goodness rating task in experiment 2. Each point represents the average goodness on a scale from 1 (bad) to 9 (good) with which an *X* presented in the position of the point exemplifies the relation to an *O* presented in the central position.

As with the production task the patterns in the different panels appear to be the same except for changes in orientation and direction. The highest ratings—near 9—were given to the three points directly *above*, *below*, *over*, *under*, *left of*, or *right of* the central position, which correspond to the "best" regions that we saw in experiment 1. Note that distance did not matter much in the "best" regions; ratings were close to 9 whether the *X* was near to the *O* or far from it. Intermediate ratings were given to the 18 positions on either side of the three best positions, and the lowest ratings (near 1) were given to the remaining 27 points. There was a sharp boundary between bad and acceptable regions. The boundary between acceptable and good regions was less marked. The acceptable regions themselves were not uniform. With *above*, for example, ratings in the first position higher than the *O* tended to decrease

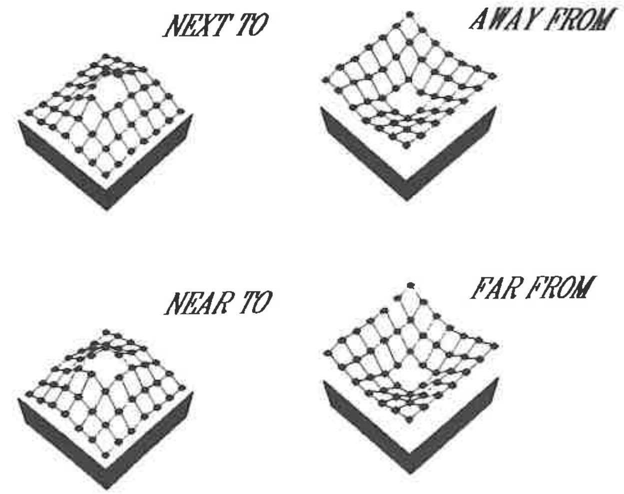

Figure 13.4
Average ratings for *next to*, *away from*, *near to*, and *far from* from the goodness rating task in experiment 2. Each point represents the average goodness on a scale from 1 (bad) to 9 (good) with which an *X* presented in the position of the point exemplifies the relation to an *O* presented in the central position.

as the position of the *X* extended farther to the left and the right, whereas ratings for the highest positions were not affected much by distance from the center, as if the region of intermediate fit were slightly U-shaped. The mean ratings for the first position higher than the *O* were 5.63, 6.41, 7.09, 8.53, 7.35, 6.74, and 5.53 from left to right. The mean ratings for positions directly above the *O* were 8.53, 8.55, and 8.61 from bottom to top. The same trends can be seen with the other relations.

The average ratings for *next to*, *away from*, *near to*, and *far from* are presented in figure 13.4 using the same three-dimensional format as figure 13.3. For *next to* and *near to*, ratings were highest in positions adjacent to the central position (occupied by the *O*) and they diminished gradually as distance increased. Consistent with experiment 1, there was a tendency to interpret *next to* horizontally; positions to the left and right of the central position were rated higher than positions the same distance away but above and below the central position. The mean ratings for the positions immediately left and right of the *O* were 8.17 and 8.39, respectively, whereas the mean ratings for the positions immediately above and below the *O* were 6.07 and 6.19, respectively.

Away from and *far from* were "mirror images" of *next to* and *near to*. Ratings were lowest in positions immediately adjacent to the central position and rose gradually as

Table 13.1
Mean Goodness Ratings for Each Relation in Experiment 2 as a Function of the Position Occupied by the *X*

Above						
7.00	7.66	8.10	8.61	8.19	7.32	7.66
6.69	6.56	7.66	8.55	7.13	7.16	6.88
5.63	6.41	7.09	8.53	7.35	6.74	5.53
1.94	2.16	1.88		1.97	1.88	2.00
1.94	1.78	1.66	1.13	1.63	2.41	1.66
1.81	1.94	1.42	1.03	1.50	1.84	1.58
1.44	1.38	1.34	1.19	1.34	2.08	1.44
Below						
1.50	1.66	1.29	1.03	1.33	1.75	1.59
1.71	2.09	1.40	1.31	1.44	1.66	1.45
1.94	2.09	1.65	1.72	1.88	2.39	2.00
2.16	2.29	2.03		2.41	1.94	2.00
5.66	6.31	6.94	8.16	6.94	6.00	5.81
6.00	7.09	7.74	8.71	7.78	7.10	6.88
7.42	5.0	6.88	8.40	7.72	7.71	7.53
Over						
8.84	7.65	8.10	8.90	7.59	7.38	7.10
6.75	6.94	7.19	8.29	7.45	7.32	8.41
5.69	5.97	7.07	8.42	7.19	6.38	5.58
1.91	2.19	2.09		2.13	1.94	2.25
2.28	1.91	1.71	1.28	1.97	2.09	2.00
1.69	2.00	1.28	1.45	2.19	1.69	1.66
1.52	1.59	1.52	1.20	1.28	1.66	1.66
Under						
1.81	1.94	1.38	1.39	1.59	1.72	1.47
1.83	1.53	2.03	1.41	1.44	1.63	1.84
1.77	1.78	1.63	1.44	1.59	1.68	2.19
2.06	2.22	1.91		2.25	2.39	2.00
5.71	5.66	6.75	8.23	6.84	5.88	5.84
6.59	7.00	7.59	7.45	7.38	6.50	6.10
7.22	7.55	7.90	8.72	7.78	7.74	7.03
Left						
6.56	5.65	5.28	2.56	2.13	1.88	1.66
7.00	6.06	5.39	2.25	2.16	1.53	1.80
7.13	6.52	6.34	2.31	2.47	1.94	2.10
8.35	7.83	8.06		1.10	1.59	1.94
6.84	6.39	6.65	2.16	2.03	2.10	1.41
6.03	6.23	5.63	2.48	1.90	2.22	1.59
6.16	5.77	4.94	1.90	1.94	1.94	2.03

Table 13.1 (continued)

Right						
1.72	1.97	1.66	2.22	5.50	6.45	6.59
1.90	2.00	2.00	2.28	5.78	6.52	7.06
1.33	1.68	2.13	2.39	6.39	6.84	7.03
1.09	1.35	1.38		8.35	8.52	8.71
1.69	1.74	2.25	2.09	6.03	6.81	7.50
1.66	1.94	1.81	2.03	5.59	6.72	6.63
1.50	1.72	1.94	1.87	5.47	6.13	6.44
Next to						
2.65	2.06	2.10	2.03	2.29	1.94	1.70
2.84	3.32	3.31	3.91	3.35	3.34	2.94
4.06	4.75	5.90	6.07	6.57	4.72	3.87
4.52	6.00	8.17		8.39	6.69	4.88
3.56	4.59	6.59	6.19	5.91	5.38	4.13
2.94	3.58	3.66	4.06	4.00	3.32	3.06
2.37	2.06	2.53	2.31	1.81	2.00	1.69
Away from						
7.38	7.94	7.45	7.74	7.72	8.10	8.44
7.41	6.84	5.74	5.16	5.69	6.72	7.22
5.90	4.75	2.94	2.91	2.78	5.13	6.47
5.35	4.38	2.13		1.88	4.58	6.25
6.32	4.81	3.09	2.50	3.44	5.41	6.45
7.28	6.09	5.34	4.97	5.41	5.75	7.66
8.10	7.50	7.58	7.63	7.44	7.83	8.26
Near to						
1.74	1.90	2.84	3.16	2.34	1.81	2.13
2.61	3.84	4.66	4.97	4.90	3.56	3.26
4.06	5.56	7.55	7.97	7.29	4.80	3.91
4.47	5.91	8.52		7.90	6.13	4.63
3.47	4.81	6.94	7.56	7.31	5.59	3.63
3.25	4.03	4.50	4.78	4.41	3.47	3.10
1.84	2.23	2.03	3.06	2.53	2.13	2.00
Far from						
7.48	7.94	7.56	7.42	7.38	7.88	8.48
6.56	5.78	5.41	5.41	5.19	5.38	7.03
5.69	4.03	2.28	1.78	2.84	4.13	6.06
5.59	3.44	1.87		1.66	4.22	5.71
6.90	4.56	2.28	1.81	2.31	4.09	6.13
7.09	6.03	4.88	5.19	5.16	6.00	7.42
7.68	7.77	7.58	7.13	7.47	7.78	8.41

distance increased. The corner positions, which were the most distant, got the highest ratings. As with figure 13.3, the ratings in figure 13.4 appear to capture the regions of best fit that were found in experiment 1. The parts of space that received the highest ratings were the parts of space in which subjects tended to draw their *X*s.

The data in figures 13.3 and 13.4 capture our idea of spatial templates quite graphically. One can imagine centering the shape in each panel on a reference object, rotating it into alignment with a reference frame, and using it to determine whether a located object falls in a good, acceptable, or bad position.

13.9 Experiment 3: Similarity Rating Task

The data in figures 13.1–13.4 suggest a pattern of similarities among the relations. Templates corresponding to *above*, *below*, *over*, *under*, *left of*, and *right of* have similar shapes but differ from each other in orientation and direction. Templates corresponding to *next to*, *away from*, *near to*, and *far from* have different shapes from *above*, *below*, and so on, but are similar to each other except that *next to* and *near to* are reflections of *away from* and *far from*. The purpose of the third experiment was to capture these similarities in a task that did not involve external, visible relations.

Subjects were presented with all possible pairs of words describing the twelve relations, *above*, *below*, *left of*, *right of*, *over*, *under*, *next to*, *away from*, *near to*, *far from*, *in*, and *on*, and they were asked to rate their similarity on a scale of 1 (dissimilar) to 10 (similar). The words were printed in pairs with a blank beside them, in which subjects were to write their rating. The 66 pairs were presented in two single-spaced columns on a single sheet of paper. There were four groups of subjects (26, 28, 19, and 28 in each group) who received the pairs in different orders. The subjects were 101 volunteers from an introductory psychology class.

The ratings for each word pair were averaged across subjects, and the averages were subjected to a multidimensional scaling analysis, using KYST (Kruskal, Young, and Seery 1977). We tried one-, two-, and three-dimensional solutions and found that stress (a measure of goodness of fit, analogous to $1 - r^2$) was minimized with a three-dimensional fit. The stress values were .383, .191, and .077 for the one-, two-, and three-dimensional solutions, respectively. The similarity space for the three-dimensional solution is depicted in figures 13.5, 13.6, and 13.7.

Figure 13.5 shows the plot of dimension 1 against dimension 2, which appears to be a plot of an *above-below* dimension against a *near-far* dimension. *Above* and *over* appear in the bottom right, and *below* and *under* appear in the top left. *Away from* and *far* appear in the bottom left, and *next to*, *near*, *in*, and *on* appear in the top right. *Left* and *right* appear in the middle, reflecting their projection on the *above-below* × *near-far* plane.

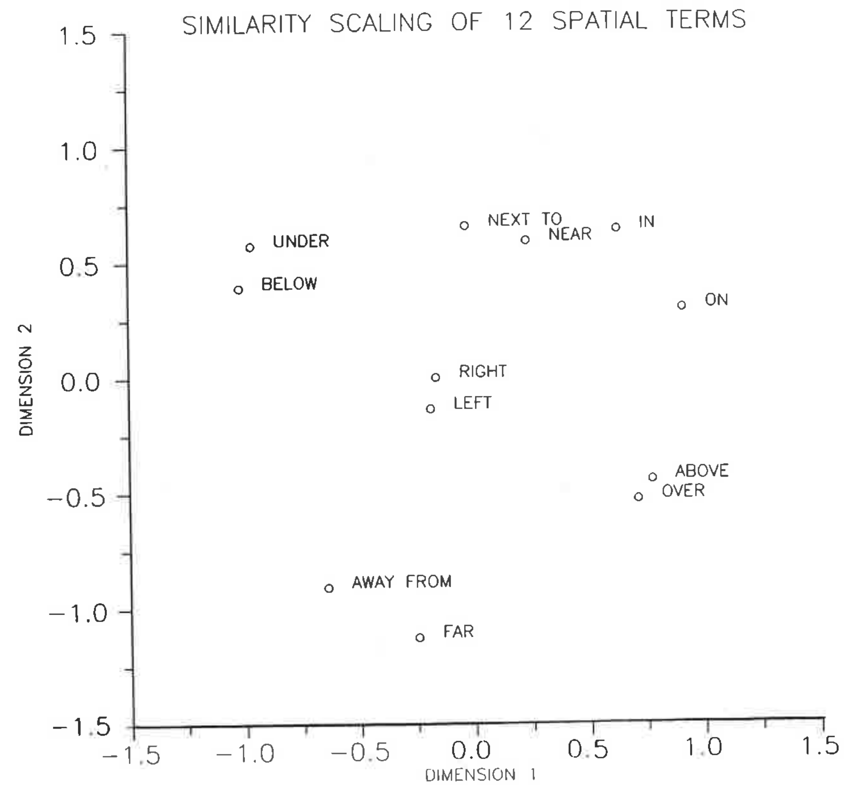

Figure 13.5
Dimension 1 × dimension 2 plotted from a similarity space constructed from a multidimensional scaling of similarity ratings of twelve spatial terms in experiment 3 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* × *near-far*.

Figure 13.6 shows the plot of dimension 1 against dimension 3, which appears to be a plot of an *above-below* dimension against a *left-right* dimension. *Above* and *over* appear on the left side, and *below* and *under* appear on the right. *Left* appears on the top, and *right* appears on the bottom. The other relations are scattered over the middle of the plot, reflecting the projection of the *near-far* axis on the *above-below* × *left-right* plane.

Figure 13.7 shows the plot of dimension 2 against dimension 3. This appears to be a plot of *near-far* against *left-right*. *In*, *on*, *next to*, and *near* appear on the top, whereas *far* and *away from* appear on the bottom. *Right* appears on the left side, while *left* appears on the right. *Above*, *over*, *below*, and *under* are scattered over the plane, reflecting the projection of the *above-below* axis on the *near-far* × *left-right* plane.

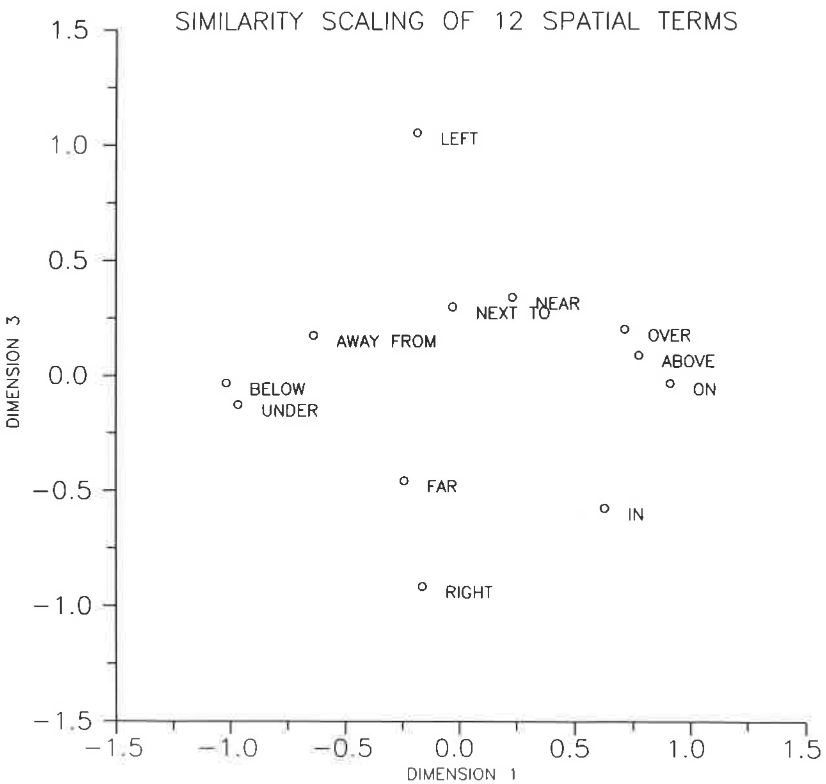

Figure 13.6
Dimension 1 × dimension 3 plotted from a similarity space constructed from a multidimensional scaling of similarity ratings of twelve spatial terms in experiment 3 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* × *left-right*.

The similarity structure in these plots resembles that seen in figures 13.1–13.4. The templates for *above* and *over* have similar shapes, opposite to those for *below* and *under*. The templates for *left* and *right* are opposite to each other and orthogonal to *above* and *below*. The templates for *far* and *away from* are similar to each other and opposite to *near* and *next to*, and all of their shapes are different from those of *above*, *below*, *left*, *right*, and so on.

In order to formalize these intuitions, we calculated similarity scores from the spatial templates in figures 13.3 and 13.4 and subjected them to multidimensional scaling, using KYST. The procedure involved several steps. We treated the forty-eight ratings for each relation as a vector and assessed similarity between relations by computing the dot product of the corresponding vectors. That is, we multiplied the

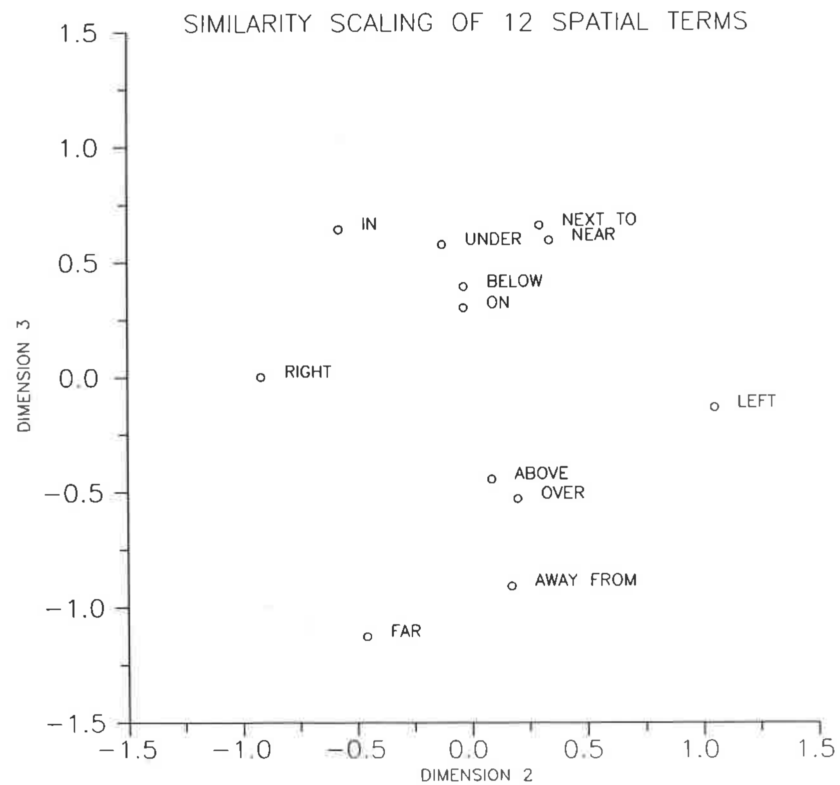

Figure 13.7
Dimension 2 × dimension 3 plotted from a similarity space constructed from a multidimensional scaling of similarity ratings of twelve spatial terms in experiment 3 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *near-far* × *left-right*.

ratings in corresponding cells and added them up to produce a similarity score analogous to a correlation coefficient. Before computing the dot product, we normalized the vectors, setting the sum of their squared values to the same value for each relation. There were forty-five dot products, reflecting all possible pairs of the ten relations examined in experiment 2. These forty-five dot products were treated as similarity ratings and ran through the KYST program. As before, we tried one-, two-, and three-dimensional solutions and found stress minimized with a three-dimensional solution. The stress values were .315, .139, and .009 for one, two, and three dimensions, respectively. The three-dimensional similarity space is plotted in figures 13.8, 13.9, and 13.10.

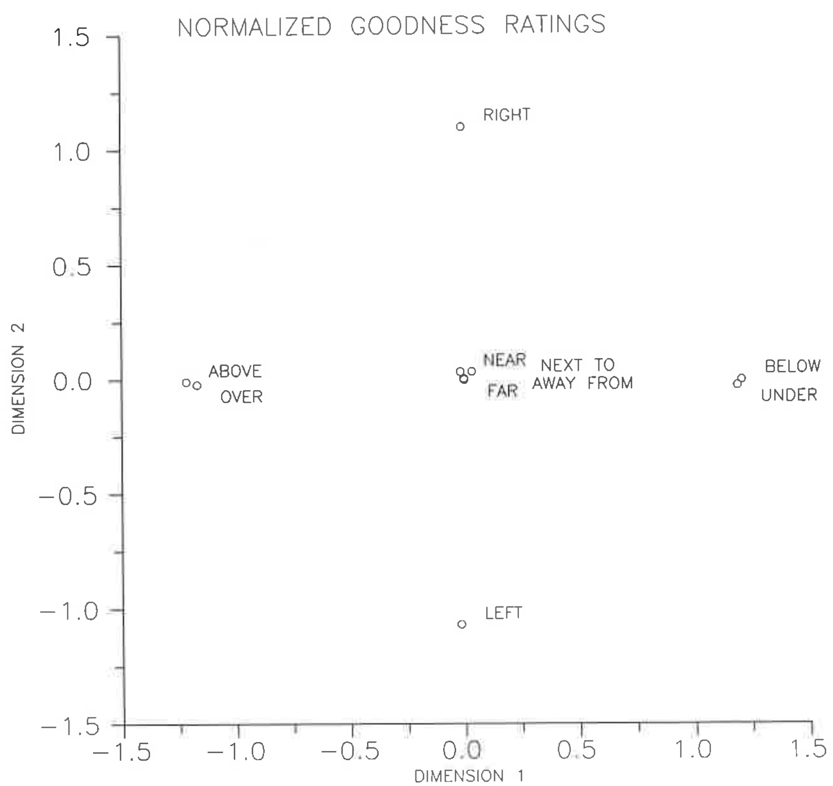

Figure 13.8
Dimension 1 × dimension 2 plotted from a similarity space constructed from a multidimensional scaling of dot products from goodness ratings of ten spatial terms from experiment 2 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* × *left-right*.

The dimensional structure that emerged from the scaling analysis of the goodness ratings was very similar to the one that emerged from the similarity ratings. The structure had three dimensions and the three dimensions could be interpreted similarly. Figure 13.8 contains the plot of dimension 1 against dimension 2, which is easily interpretable as a plot of the *above-below* axis against the *left-right* axis. Figure 13.9 contains the plot of dimension 1 against dimension 3, which appears to be a plot of the *above-below* axis against the *near-far* axis. Figure 13.10 contains the plot of dimension 2 against dimension 3, which appears to be a plot of the *left-right* axis against the *near-far* axis. We assessed the similarity of the fits quantitatively by calculating the correlation between the interpoint distances in the two solutions. Each

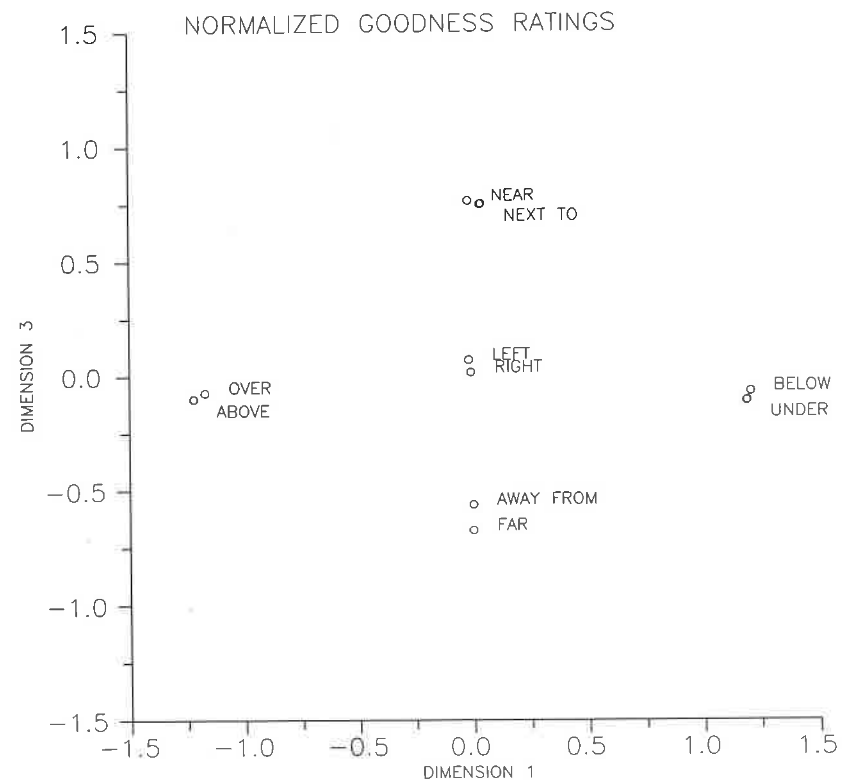

Figure 13.9
Dimension 1 × dimension 3 plotted from a similarity space constructed from a multidimensional scaling of dot products from goodness ratings of ten spatial terms from experiment 2 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *above-below* × *near-far*.

solution gives the distance between each pair of relations in multidimensional space. If the solutions are similar, then the distances between the same pairs of relations in the two spaces should be similar. The correlation was .858, indicating good agreement.

The similarity of the scaling solutions and the high correlation between distances suggests that the ratings of pictures in experiment 2 and the ratings of words in the present experiment were based on common, underlying knowledge structures. We would like to conclude that subjects used spatial templates to perform both tasks. Thus they rated pictures by aligning spatial templates with the reference object and computing the goodness of fit for the located object, and they rated words by

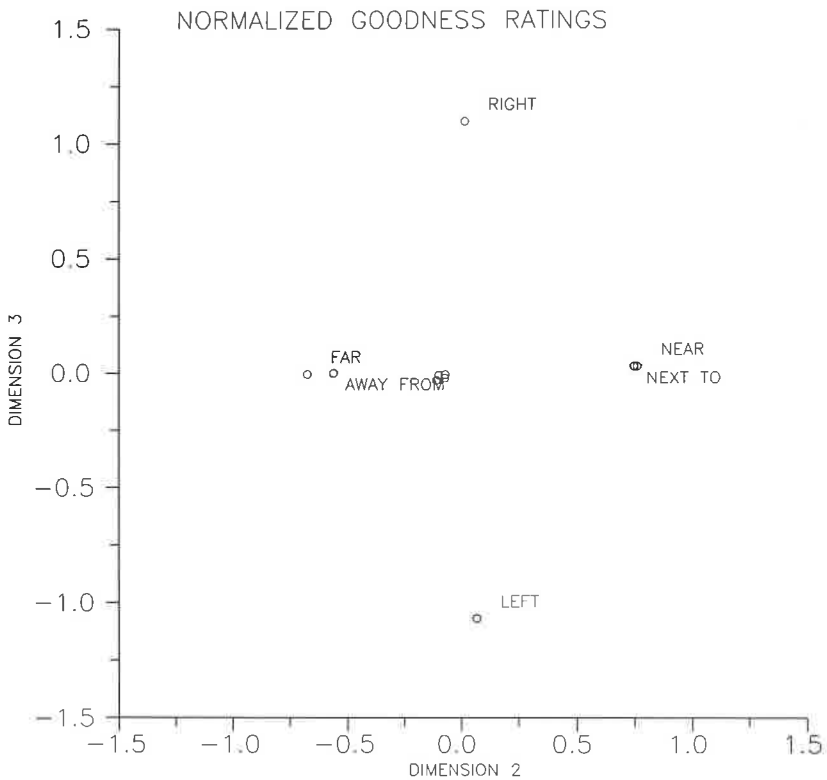

Figure 13.10
Dimension 2 × dimension 3 plotted from a similarity space constructed from a multidimensional scaling of dot products from goodness ratings of ten spatial terms from experiment 2 (the numbers on the axes are arbitrary measures of distance). The dimensions appear to be *left-right* × *near-far*.

comparing the spatial templates associated with them. This conclusion is speculative, however. Although there is some evidence that subjects may compare images when given words (Shepard and Chipman 1970), other representations and processes could produce the same outcomes. The data are consistent with our conclusion, but they do not rule out competing interpretations.

13.10 Experiment 4: Relation Judgment Task

The results of experiments 1–3 are consistent with the hypothesis that spatial templates were applied in parallel to the whole perceptual representation, but they do not support that hypothesis uniquely. The same results could have been produced by

applying *serial visual routines* instead of spatial templates. Serial visual routines are processes that operate sequentially on perceptual representations to compute a number of things, including spatial relations (Ullman 1984). For example, *above* could be produced by centering a "mental cursor" on the reference object and moving upward along the up-down axis of the reference frame until the located object was found (Jolicoeur, Ullman, and MacKay 1986, 1991). If the located object was not directly above the reference object, the cursor could move from one side to the other covering the region above the top of the reference object until the located object was found. From this perspective, the spatial templates evidenced in experiments 1 and 2 may reflect preferred trajectories for serial visual routines rather than explicit representations used to compute spatial relations directly (i.e., by multiplying activation values as described earlier). The purpose of the fourth experiment was to contrast spatial templates with serial visual routines in the apprehension of spatial relations (see also Logan and Compton 1996; Sergent 1991).

The main point of contrast between spatial templates and serial visual routines is the effect of distance in judging spatial relations. Spatial templates are applied in parallel to the whole visual field, so distance between located and reference objects does not matter. The time taken to apply a spatial template should not depend on distance. By contrast, serial visual routines operate sequentially, examining the visual field bit by bit, so distance between located and reference objects should make a difference. The time taken to apply a serial visual routine should increase monotonically with distance.

Note, the evidence in experiments 1 and 2 that distance has no effect on the goodness of examples of *above*, *below*, *over*, *under*, *left of*, and *right of* does not bear on this issue because time was neither stressed nor measured. Subjects could have taken more time to rate greater distances even though they gave the same rating. The rating could have depended on the relation between the located object and the reference frame centered on the reference object, not on the time taken to compute the relation.

Experiment 4 had subjects perform a verification task in which the distance between reference and located objects was varied systematically (cf. Clark, Carpenter, and Just 1973). The range of distances used in this experiment (1–6 degrees of visual angle) was well within the range that shows monotonic increases, in reaction time in other tasks, such as mental curve tracing (Jolicoeur, Ullman, and MacKay 1986, 1991); if serial visual routines had been used to compute spatial relations in the present experiments, reaction time should therefore have increased with distance.

The experiments focused on the relations *above* and *below*. Each trial began with a fixation point exposed for 500 ms in the center of a computer screen. It was extinguished and replaced with a sentence expressing the relation between a dash and a plus (i.e., "Dash above plus?"; "Dash below plus?", "Plus above dash?", or "Plus below dash?") that was exposed for 1,000 ms. After the sentence was extinguished, the fixation point appeared for another 500 ms. Then a picture of a dash above or below a plus was exposed for 200 ms, too briefly to allow eye movements. Half of the time, the relation between the dash and plus matched the sentence, and half of the time, the opposite relation held. Subjects were told to respond "true" to the former case and "false" to the latter. After the 200-ms exposure of the picture, the screen went blank until the subject responded. After the response, the screen remained blank for a 1,500 ms intertrial interval. There were 384 trials in all.

The main manipulation was the distance between the dash and the plus. There were four different distances. In one version of the experiment, the dash and plus were separated by 1, 2, 3, or 4 screen lines (corresponding to .74, 1.48, 2.22, and 2.96 degrees of visual angle when viewed from a distance of 60 cm). In another version, distances were doubled. The dash and the plus were separated by 2, 4, 6, or 8 screen lines (1.48, 2.96, 4.44, or 5.92 degrees of visual angle). Stimuli separated by the different distances appeared in several different locations on the screen. In the version in which distances were 1–4 screen lines, stimuli with a distance of 1 appeared in positions 1 and 2, 2 and 3, 3 and 4, and 4 and 5; stimuli with a distance of 2 appeared in positions 1 and 3, 2 and 4, and 3 and 5; stimuli with a distance of 3 appeared in positions 1 and 4, and 3 and 5; and stimuli with a distance of 4 appeared in positions 1 and 5. The same scheme was used in the version in which distances were 2–8 screen lines, except that positions 1–5 were two lines apart. Distances, relations (*above* vs. *below*), and true and false trials occurred in random order. A different random order was constructed for each subject. The subjects were 48 volunteers from an introductory psychology class. Twenty-four served in each version of the experiment.

Mean reaction times were computed for "true" and "false" responses as a function of distance. The means across subjects are plotted in figures 13.11 and 13.12. Figure 13.11 plots reaction time as a function of absolute distance, expressed in degrees of visual angle. It shows that reaction time was longer for "false" responses than for "true" responses in both versions of the experiment, $F(1,44) = 78.97$, $p < .01$, mean square error (MSE) = 102,274.38. Reaction time was longer in the version with the greater distances, but the difference was not significant, $F(1,44) < 1.0$. The most important result for our present purposes is the effect of distance. Serial visual routines predict a monotonic increase in reaction time as distance increases, whereas spatial templates predict no effect. Analysis of variance showed a significant main effect of distance, $F(3,132) = 4.33$, $p < .01$, $MSE = 57{,}930.55$, and the linear trend was significant, $F(1,132) = 4.77$, $p < .01$, indicating a tendency for reaction time to *decrease* as distance increased. The observed pattern is clearly inconsistent with serial

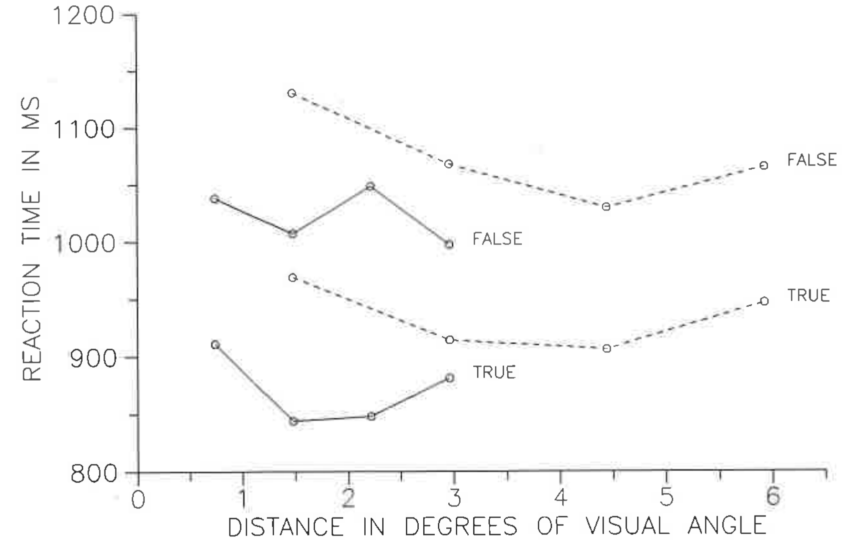

Figure 13.11
Reaction time as a function of absolute distance between reference and located objects from two versions of experiment 4 in which subjects judged *above* and *below*. "True" versus "false" response and long (dotted lines) versus short (solid lines) distances are the parameters.

visual routines. In both versions of the experiment, reaction time was longest for the shortest and longest distances and fastest for the intermediate distances.

The pattern of reaction times is not exactly what one would expect from the spatial template hypothesis, which predicted no effect of distance. However, the pattern may be consistent with theory of apprehension in which spatial templates play a part, if the slower reaction times at the longest and shortest distances can be explained. We suggest that the pattern reflects a process of reference frame adjustment. Subjects may have set the scale of their reference frames to the average distances they experienced—distances of 2 and 3 in one version and distances of 4 and 6 in the other. They may have adjusted them if the distance were longer or shorter than the average—distances 4 and 1 in one version and 8 and 2 in the other. This would produce the observed pattern of results. The effect can be seen more clearly in figure 13.12, which plots reaction time as a function of ordinal distance rather than relative distance. The patterns from the two versions of the experiment align nicely in figure 13.12. Of course, this explanation is post hoc, and must be taken with a grain or two of salt (however, no distance effects were found by Logan and Compton 1996 and by Sergent 1991).

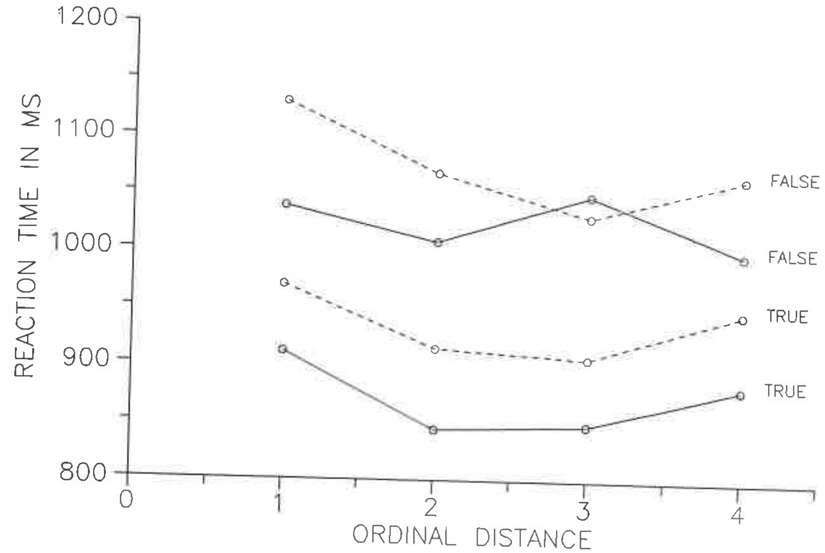

Figure 13.12
Reaction time as a function of ordinal distance between reference and located objects from two versions of experiment 4 in which subjects judged *above* and *below*. "True" versus "false" response and long (dotted lines) versus short (solid lines) distances are the parameters.

13.11 Conclusions

The data from experiments 1–4 support the idea that spatial templates underlie the apprehension of spatial relations. Experiments 1 and 2 showed that the space around a reference object is divided into regions that represent good, acceptable, and bad examples of a given relation (see also Hayward and Tarr 1995). Experiment 3 showed that similarities in the meanings of spatial terms can be accounted for in terms of similarities in the spatial templates that correspond to them. And Experiment 4 showed that distance between reference and located objects has little effect on the time required to apprehend relations, as if spatial templates were applied in to the whole visual field in simultaneously (see also Logan and Compton 1996; Sergent 1991). Together with the other data (Logan 1994, 1995), the experiments support the computational analysis of apprehension presented earlier in the chapter and argue for its viability as a psychological theory of apprehension in humans.

Several parts of the theory were taken from existing analyses of spatial relations. Reference frames and spatial indices play important roles in linguistic and psycholinguistic analyses (see Carlson-Radvansky and Irwin 1993, 1994; Clark 1973; Garnham 1989; Herskovits 1986; Jackendoff and Landau 1991; Landau and Jackendoff

1993; Levelt 1984; Logan 1995; Miller and Johnson-Laird 1976; and Talmy 1983). The novel contribution is the idea that goodness of fit is computed with spatial templates. We suggested this idea because it is computationally simple and easy to implement in software or "wetware." It would be interesting to contrast spatial templates with other ways to compute goodness of fit in future research (e.g., geometric, volumetric, topological, or functional relations).

The theory was developed to account for the apprehension of spatial prepositions in English. As is readily apparent in the other chapters in this volume, different languages express spatial relations in different ways, so it is important to consider how the theory might generalize to other languages. What is general across languages and what is specific to English? We suspect that the theory could be adapted to most languages. Most languages express relations between objects in terms of reference frames applied to reference objects. We suspect that reference frame computation and spatial indexing (which is required to distinguish reference objects from located objects) may be common to all languages. The spatial templates applied to the reference objects may vary between languages. We suspect that spatial templates are shaped by the linguistic environment to capture the distinctions that are important in particular languages. The perceptual representation must be common to all languages because it is precognitive and thus prelinguistic. The conceptual representations clearly vary between languages. We suggest that the conceptual representations may be distinguished from each other in terms of the spatial templates with which they are associated.

The spatial templates measured in this chapter are crude approximations to the templates that people might actually use (if they use them at all). The measurements were coarse (e.g., experiment 2 used a 7 × 7 grid) and the reference and located objects were simple (boxes, *O*s and *X*s). We suspect that the results would generalize to finer measurements and more sophisticated objects. Indeed, Hayward and Tarr (1995) and Carlson-Radvansky and Irwin (1993) found similar results with several different reference and located objects. Certainly, the methods could be adapted to more precise measurements, different classes of objects, and even different spatial relations. Thus we do not view the experiments as the final answer, but rather, as a promising beginning to an exciting area of inquiry.

The measurements in the present experiments may not have captured all of the differences between the relations we contrasted. Experiment 1, for example, found evidence of two different senses of *over* (*above* and *covering*), whereas experiment 2 found evidence of only one of them (*above*). The displays in experiment 2 could not have picked up the second meaning because the located and reference objects were always separated. However, it should be possible to pick up the contrast with displays in which located and reference objects overlap. Subjects should rate overlapping displays as good examples of *over* but bad examples of *above*. Thus the limitations of the present experiments lie in the specific procedures we used rather than in the general methodology. With appropriately designed displays, rating procedures should be able to capture subtle differences between relations.

Spatial templates may not capture the meanings of all spatial relations. *On*, for example, implies contact and support (Bowerman, chapter 10, this volume), neither of which can be described sufficiently in terms of occupancy of regions of space. The reference object and the located object must occupy the same region of space, but contact and support imply more than that. Contact may be assessed by examining junctions between the contours of the objects using something like templates (Biederman 1987), but support cannot be perceived so easily. *In*, as another example, implies containment (Herskovits 1986) and that is a functional relationship that cannot be described easily in terms of regions of space. Flowers *in* a vase occupy a different region of space than water *in* a vase.

Despite these limitations, spatial templates are clearly useful in describing the meanings of many spatial relations. Moreover, they are tractable computationally, and the computational analysis is readily interpretable as a psychological theory of how people actually apprehend spatial relations. The data in the present experiments and others (Carlson-Radvansky and Irwin 1993; Hayward and Tarr 1995; Logan 1994, 1995; Logan and Compton 1996) are consistent with the psychological theory, suggesting it has some validity. Competitive theories, based on assessment of geometric, topological, and functional relations, have not yet reached this stage of development.

Acknowledgments

This research was supported in part by National Science Foundation grant BNS 91-09856 to Gordon Logan. We are grateful to Jane Zbrodoff for valuable discussion. We would like to thank Paul Bloom and Mary Peterson for helpful comments on the manuscript.

Notes

1. "This is here" and "That is there" are often interpreted as deictic relations in linguistic analyses (e.g., Levelt 1984). However, in those analyses, the expressions are interpreted as sentences that one person utters to another. The listener must interpret what the speaker says in terms of two-argument relation between two external objects—the speaker as a reference object and "this" or "that" as a located object. Moreover, the listener must interpret what the speaker says in terms of the speaker's frame of reference, with "here" meaning *near* and "there" meaning *far*. Basic relations are intrapersonal rather than interpersonal. There is only one argument ("this" or "that") and there is no external frame of reference (i.e., the viewer's own frame of reference suffices). The viewer is telling himself or herself that an object exists in a location. We expressed the result of that process as a sentence to communicate the idea to the reader, but the viewer need not do so. The viewer's representation is conceptual rather than linguistic.

2. One sheet contained *under*, *near*, *in*, and *away from* in the top left, top right, bottom left, and bottom right positions, respectively. Another contained *above*, *on*, *right of*, and *next to*. The third contained *left of*, *over*, *below*, and *far from*. Roughly equal numbers of subjects received the three different orders of sheets (25, 20, and 23, respectively).

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

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

Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition*, *46*, 223–244.

Carlson-Radvansky, L. A., and Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, *33*, 646–671.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 27–63. New York: Academic Press.

Clark, H. H., Carpenter, P. A., and Just, M. A. (1973). On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing*, 311–381. New York: Academic Press.

Clark, H. H., and Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory and Cognition*, *2*, 101–111.

Cooper, L. A., and Shepard, R. (1973). The time required to prepare for a rotated stimulus. *Memory and Cognition*, *1*, 246–250.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review*, *95*, 115–123.

Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, *40*, 225–240.

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

Garnham, A. (1989). A unified theory of the meaning of some spatial relational terms. *Cognition*, *31*, 45–60.

Hayward, W. G., and Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition*, *55*, 39–84.

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

Jackendoff, R., and Landau, B. (1991). Spatial language and spatial cognition. In D. J. Napoli and J. A. Kegl (Eds.), *Bridges between psychology and linguists: A Swarthmore festschrift for Lila Gleitman*, 145–169. Hillsdale, NJ: Erlbaum.

Jolicoeur, P., Ullman, S., and MacKay, L. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory and Cognition, 14*, 129–140.

Jolicoeur, P., Ullman, S., and MacKay, L. (1991). Visual curve tracing properties. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 997–1022.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., and Koenig, O. (1992). Categorical versus coordinate spatial relations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 562–577.

Kruskal, J. B., Young, F. W., and Seery, J. B. (1977). How to use KYST-2: A very flexible program to do multidimensional scaling and unfolding. Unpublished manuscript. Bell Laboratories, Murray Hill, NJ.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Landau, B., Jackendoff, R. (1993). "What" and "where" in spatial cognition and spatial language. *Brain and Behavioral Sciences, 16*, 217–238.

Levelt, W. J. M. (1984). Some perceptual limitations in talking about space. In A. J. van Doorn, W. A. de Grind, and J. J. Koenderink (Eds.), *Limits on perception*, 323–358. Utrecht: VNU Science Press.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523–553.

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1015–1036.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*, 103–174.

Logan, G. D., and Compton, B. J. (1996). Distance and distraction effects in the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 159–172.

Logan, G. D., and Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting trials in a Stroop-like task. *Memory and Cognition, 7*, 166–174.

Marr, D. (1982). *Vision*. New York: Freeman.

Marr, D., and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Philosophical Transactions of the Royal Society, London, 200*, 269–294.

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

Palmer, S. E. (1989). Reference frames in the perception of shape and orientation. In B. E. Shepp and S. Ballesteros (Eds.), *Object perception: Structure and process*, 121–163. Hillsdale, NJ: Erlbaum.

Peterson, M. A., Kihlstrom, J. F., Rose, P. M., and Glisky, M. L. (1992). Mental images can be ambiguous: Reconstruals and reference frame reversals. *Memory and Cognition, 20*, 107–123.

Pylyshyn, Z. (1984). *Computation and cognition*. Cambridge, MA: Harvard University Press.

Pylyshyn, Z. (1989). The role of location indices in spatial perception: A sketch of the FINST spatial index model. *Cognition, 32*, 65–97.

Sergent, J. (1991). Judgments of relative position and distance on represention of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 762–780.

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

Talmy, L. (1983). How language structures space. In H. L. Pick and L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application*, 225–282. New York: Plenum Press.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97–159.

Vandaloise, C. (1991). *Spatial preposition: A case study from French.* Chicago: University of Chicago Press.